

# A Program Assessment Guide:

Best Practices for Designing  
Effective Assessment Plans



2019  
Adam Andrews  
Academic Affairs

# A Program Assessment Guide:

Best Practices for Designing Effective  
Assessment Plans

## Contents

Purpose of this Guide. . . . .	4
Principles of Good Practice for Assessing Student Learning . . . . .	5
Part 1: Assessment as a Scholarship of Learning . . . . .	8
Part 2: Defining Goals and Outcomes . . . . .	10
Goals. . . . .	10
Outcomes. . . . .	13
Alignment and Curriculum Mapping. . . . .	22
Part 3: Designing an Assessment Plan . . . . .	28
Different Kinds of Assessment. . . . .	29
Kinds of Evidence. . . . .	30
Using Rubrics . . . . .	33
Planning for Sustainable Assessment. . . . .	36
The Assessment Plan Essentials . . . . .	39
Assessment Beyond the Essentials. . . . .	43
Finalizing the Assessment Plan . . . . .	46
Part 4: Analyzing & Sharing Results . . . . .	48
Analyzing Assessment Data. . . . .	48
Sharing Data. . . . .	54
Part 5: Using Results with an Action Plan . . . . .	56
Part 6: Reporting Assessment Work . . . . .	60
Resources . . . . .	61
Bibliography . . . . .	63
Glossary . . . . .	67

*This page intentionally left blank.*

# Purpose of this Guide

This guide is meant to help UWM faculty take control of the program assessment process by actively using assessment to create improved outcomes for students. It draws on established assessment research and best practices to assist undergraduate and graduate program coordinators, and other faculty involved with assessment, with thinking about program assessment. Having a good program level assessment plan that produces usable data is necessary for the accreditation process, but it is also a crucial mechanism for understanding the needs of our students and making informed decisions about how to improve student learning in relation to the learning and experiences that faculty want them to have.

# Principles of Good Practice for Assessing Student Learning

In 1992, the American Association for Higher Education released their “Principles of Good Practice,” written by a group of prominent assessment scholars to help educational institutions embrace assessment practices that could lead to improved student experiences and outcomes. Since then, they have served as a touchstone for assessment practitioners nationwide. The language of these principles is broad enough to anchor assessment practices at the course, program, and institutional levels by identifying the core values and practices common to all.

Their principles are as follows:

- 1. The assessment of student learning begins with educational values.** Assessment is not an end in itself but a vehicle for educational improvement. Its effective practice, then, begins with and enacts a vision of the kinds of learning we most value for students and strive to help them achieve. Educational values should drive not only what we choose to assess but also how we do so. Where questions about educational mission and values are skipped over, assessment threatens to be an exercise in measuring what's easy, rather than a process of improving what we really care about.
- 2. Assessment is most effective when it reflects an understanding of learning as multidimensional, integrated, and revealed in performance over time.** Learning is a complex process. It entails not only what students know but what they can do with what they know; it involves not only knowledge and abilities but values, attitudes, and habits of mind that affect both academic success and performance beyond the classroom. Assessment should reflect these understandings by employing a diverse array of methods including those that call for actual performance, using them over time so as to reveal change, growth, and increasing degrees of integration. Such an approach aims for a more complete and accurate picture of learning, and therefore firmer bases for improving our students' educational experience.
- 3. Assessment works best when the programs it seeks to improve have clear, explicitly stated purposes.** Assessment is a goal-oriented process. It entails comparing educational performance with educational purposes and expectations--these derived from the institution's mission, from faculty intentions in program and course design, and from knowledge of students' own goals. Where program purposes lack specificity or agreement, assessment as a process pushes a campus toward clarity about where to aim and what standards to apply; assessment also prompts attention to where and how



program goals will be taught and learned. Clear, shared, implementable goals are the cornerstone for assessment that is focused and useful.

**4. Assessment requires attention to outcomes but also and equally to the experiences that lead to those outcomes.** Information about outcomes is of high importance; where students “end up” matters greatly. But to improve outcomes, we need to know about student experience along the way--about the curricula, teaching, and kind of student effort that lead to particular outcomes. Assessment can help understand which students learn best under what conditions; with such knowledge comes the capacity to improve the whole of their learning.

**5. Assessment works best when it is ongoing, not episodic.** Assessment is a process whose power is cumulative. Though isolated, “one-shot” assessment can be better than none, improvement is best fostered when assessment entails a linked series of activities undertaken over time. This may mean tracking the progress of individual students, or of cohorts of students; it may mean collecting the same examples of student performance or using the same instrument semester after semester. The point is to monitor progress toward intended goals in a spirit of continuous improvement. Along the way, the assessment process itself should be evaluated and refined in light of emerging insights.

**6. Assessment fosters wider improvement when representatives from across the educational community are involved.** Student learning is a campus-wide responsibility, and assessment is a way of enacting that responsibility. Thus, while assessment efforts may start small, the aim over time is to involve people from across the educational community. Faculty play an especially important role, but assessment’s questions can’t be fully addressed without participation by student-affairs educators, librarians, administrators, and students. Assessment may also involve individuals from beyond the campus (alumni/ae, trustees, employers) whose experience can enrich the sense of appropriate aims and standards for learning. Thus, understood, assessment is not a task for small groups of experts but a collaborative activity; its aim is wider, better-informed attention to student learning by all parties with a stake in its improvement.

**7. Assessment makes a difference when it begins with issues of use and illuminates questions that people really care about.** Assessment recognizes the value of information in the process of improvement. But to be useful, information must be connected to issues or questions that people really care about. This implies assessment approaches that produce evidence that relevant parties will find credible, suggestive, and applicable to decisions that need to be made. It means thinking in advance about how the information will be used, and by whom. The point of assessment is not to

gather data and return “results”; it is a process that starts with the questions of decision-makers, that involves them in the gathering and interpreting of data, and that informs and helps guide continuous improvement.

**8. Assessment is most likely to lead to improvement when it is part of a larger set of conditions that promote change.** Assessment alone changes little. Its greatest contribution comes on campuses where the quality of teaching and learning is visibly valued and practiced. On such campuses, the push to improve educational performance is a visible and primary goal of leadership; improving the quality of undergraduate education is central to the institution's planning, budgeting, and personnel decisions. On such campuses, information about learning outcomes is seen as an integral part of decision making, and avidly sought.

**9. Through assessment, educators meet responsibilities to students and to the public.** There is compelling public stake in education. As educators, we have a responsibility to the publics that support or depend on us to provide information about the ways in which our students meet goals and expectations. But that responsibility goes beyond the reporting of such information; our deeper obligation--to ourselves, our students, and society--is to improve. Those to whom educators are accountable have a corresponding obligation to support such attempts at improvement.

#### **The Authors**

Alexander W. Astin, University of California at Los Angeles; Trudy W. Banta, Indiana University-Purdue University at Indianapolis; K. Patricia Cross, University of California, Berkeley; Elaine El-Khawas, American Council on Education; Peter T. Ewell, National Center for Higher Education Management Systems; Pat Hutchings, American Association for Higher Education; Theodore J. Marchese, American Association for Higher Education; Kay M. McClenney, Education Commission of the States; Marcia Mentkowski, Alverno College; Margaret A. Miller, State Council of Higher Education for Virginia; E. Thomas Moran, State University of New York, Plattsburgh; Barbara D. Wright, University of Connecticut.

This document was developed under the auspices of the AAHE Assessment Forum, a project of the American Association for Higher Education, with support from the Fund for the Improvement of Postsecondary Education. It builds on earlier efforts, by campuses and other groups, to articulate guidelines for assessment's practice; its intent is to synthesize important work already done and to invite further statements about the responsible and effective conduct of assessment.

Development of this document was sponsored by the American Association for Higher Education (AAHE) and supported by the Fund for the Improvement of Postsecondary Education (FIPSE); publication and dissemination was supported by the Exxon Education Foundation. Copies may be made without restriction.

For more information, visit <http://www.learningoutcomesassessment.org/PrinciplesofAssessment.html>  
The National Institute of Learning Outcomes Assessment (NILOA)

# Part I: Assessment as a Scholarship of Learning

Assessment is research into student learning and experience in relation to institutional and program goals. In doing assessment, faculty are tasked with approaching student learning with the same rigor, and demand for evidence, with which they approach research at an R1 institution. It asks world-renowned scholar-teachers to engage not only in research in their disciplines, but also in their teaching practice as well.

## KEY IDEAS:

- ASSESSMENT IS A FORM OF RESEARCH
- ASKING BETTER QUESTIONS LEADS TO MORE USEFUL ASSESSMENT DATA

Good assessment begins with asking questions to frame a process that can draw conclusions based on evidence. It is, at its best, inquiry into student learning that can help programs improve (Litterst & Thompkins,

2000; Banta, 2002). While much emphasis has been placed on assessment as a form of accountability, assessment's role in improving student learning and its intersections with the scholarship of teaching and learning is where its real value for programs and faculty can be found (Ewell, 2009; Hutchings, Huber, & Ciccone, 2011). In reality, most accrediting agencies have embraced assessment with this more nuanced view in mind (Ewell, 2009, p. 12-13). The difference in these two paradigms (accountability vs. improvement) is deeply connected to the questions being asked.

**The first, most basic, assessment question is “What, or to what degree, have our students learned?”** This question generates evidence of learning as a necessary and useful first step toward improvement. This question can help programs determine if students are graduating with discipline-appropriate abilities to apply, synthesize, and evaluate information, or if they truly have the skills the program wants students to acquire. In practice, however, when this question is only asked from an accountability or compliance point of view, it can lead faculty to gather assessment data, report it, and then forget about it. The value in assessment—even for accreditors—is in how faculty *use* the data. And to ensure that it is useful, faculty may need to ask some additional or more complex questions.

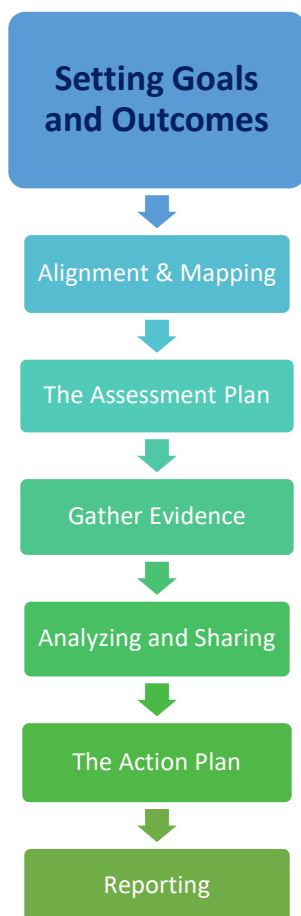
**A second, more complex, assessment question asks, “What factors are affecting our students’ learning?”** Variations of this question can guide assessment plans that examine how a new curriculum is affecting student test scores, or if a curriculum is working equally well for the different student populations in the program. Such assessment plans could ask if a TA training program is working to produce professionals that can teach effectively in their discipline, or if



students who took a prerequisite course are actually doing better than students who tested out of it. An assessment plan focused on comparative questions *also* generates evidence of learning since we can't determine what might be affecting learning unless we are also measuring it. If we ask the second question, we usually also answer the first one as well.

Faculty may informally ask questions like these about their programs all the time, and are interested in the answers. When faculty take control of the assessment process and formulate valuable questions to frame their assessment plan, the results benefit both their programs and students, and have more in common with scholarship and research than they do compliance.

## Part 2: Defining Goals and Outcomes



Before developing meaningful program assessment plan, departments and program faculty need to agree on **goals and outcomes** for the program and develop a strong **curriculum plan**. While all programs have a curriculum plan of some kind, they sometimes lack the clarity necessary to formulate a usable assessment plan. The first step is to articulate what students should be learning. **Goals** are higher level, aspirational statements, while **outcomes** (sometimes called objectives) are specific and measurable. Program Goals are helpful for developing an assessment plan that can assess program effectiveness, and once articulated, they make writing good outcomes easier.

Goals and outcomes can be articulated at the **institutional level, the program level, and the course level**. Institutional level goals and outcomes at UWM are generally managed by Academic Affairs and the Office of Assessment and Institutional Research (OAIR). Program goals and outcomes are determined by individual departments and programs, and course level goals and outcomes are determined by individual faculty or groups working together in a program. Ideally, institutional, program, and course goals and outcomes should align with one another (more on that below).

First, it's helpful to distinguish between goals and outcomes. There is considerable variation in how these terms are used across institutions. Some institutions use the term "goals," others use "outcomes," some use "objectives," and in some cases they use these terms interchangeably to refer to statements being articulated at all three levels (institutional, program, and course).

However, there is some utility (and more flexibility for faculty) in differentiating goals from outcomes. **UWM distinguishes between goals and outcomes in order to keep documentation and reporting consistent.** The primary distinction is the level of generality: Goals are general, outcomes are specific and measurable.

### Goals

Department *and program goals* can often be found on their website or in their mission statements. That is, most departments already have some version of a goal statement. They tend to be very general statements of intention that provide a framework for understanding

the learning and experiences students should encounter in the program. **Goals are about what a department or program wants for its students:** what kinds of **knowledge, skills, or attributes** they want students to develop, what kinds of **conditions** they want to provide to students to facilitate that learning, and what kinds of overall results the program strives for. Goals can provide a map for thinking about learning outcomes and assessing student learning, but they also provide a map for evaluating departmental or program effectiveness. Below is an example of program-level learning goals for a fictional Folklore Studies program at UWM.

**Mission Statement:**

The Folklore program at UWM studies and celebrates the diversity of informal and expressive cultural traditions that infuse the fabric of daily life. Students in the folklore program build fundamental skills such as communication, analysis, and critical thinking, while developing a broad base of knowledge about human cultures. They learn to use folkloristic perspectives to understand our world, engage meaningfully with civic life, and appreciate cultural diversity. Students cultivate habits of mind such as curiosity and resilience through ethnographic encounters with the world around them. Students with a degree in Folklore studies work in museums, archives, the media, public sector folklore, fields like arts administration, or go on to graduate study in the humanities or social sciences. The Folklore program provides students with an engaging course of study that will help them become capable global citizens, well-rounded individuals, and scholars of the vibrant cultural life of their communities.

**Program Goals:**

Students in the Folklore Studies program will:

- **(Folklore)** Study informal, traditional, and expressive aspects of human culture, such as storytelling, mythmaking, ritual, folk art, dance, folk music, memes, and urban legends.
- **(Theory)** Learn about and apply folkloristic theoretical perspectives to understand cultural, historical, social, and psychological dimensions of human activity.
- **(Fieldwork)** Use ethnographic fieldwork methods to study how people invent, transform, and derive meaning from tradition.
- **(Ethics)** Understand and appropriately navigate the ethical dimensions of ethnography and fieldwork.
- **(Ethnography)** Practice producing and presenting ethnography in oral, print, film, and digital mediums.
- **(Community)** Understand relationships between identity, community, and expressive traditions.
- **(Culture)** Cultivate a critical understanding of their own and other cultures.

In the example above, the Mission Statement provides a framework for understanding the purpose of the learning goals below it. It also points to some **strategic program goals**, such as program targets for job placement or graduate school acceptance rates, that can also be part of an assessment plan.

The **verbs** for *goals* point in a direction but offer little ability to determine when we've arrived. They are general and aspirational, which allows them to be flexibly adapted into multiple outcomes for different courses or tracks, but this flexibility also means **they cannot be used effectively for assessment by themselves**. The verb "to understand," for example, is acceptable for a goal statement, but is not useful for an outcomes statement because it doesn't clarify how "understanding" is measured or demonstrated.

#### **Example Verbs for Goals:**

Carry out	Gain experience
Cultivate	Learn
Develop	Practice
Engage	Study
Examine	Understand
Explore	Use

Goal statements allow for flexibility and help create coherence across the program. In being more general or broadly worded, they allow for programs to aim high in creating expectations that may be hard to measure, such as "preparing students to become ethical decision makers and good citizens" (Walvoord, 2010). Such goals *can* potentially be assessed using indirect methods such as surveys, focus groups, interviews, or other qualitative methods. But even if they are not, they can serve a useful function in guiding curriculum decisions or establishing the ethos of a program.

**Strategic Program Goals:** Strategic goals are those aimed at maintaining or improving the program's ability to support student learning and fulfill its mission. In addition to student learning, strategic goals can address issues like retention, enrollment, job placement, equity, student satisfaction, community engagement, or any other concerns tied to the program mission. Strategic goals and outcomes are included in each program's assessment plan *in addition* to learning goals and outcomes.

**Course Goals:** At the course level, as a general rule, *outcomes* are preferable to goals. Regardless of the terminology used (whether we refer to them as course goals or course outcomes), if they don't have an appropriate level of specificity, they can make it difficult to tell how the work of the course and student assessment actually relates to the objectives of the

course (and how those relate to program goals and outcomes). While faculty may wish to articulate goals for a course, it's important that the course's *outcomes* clarify how those goals will be achieved.

### Writing Goals for Graduate Programs

Establishing goals for graduate programs follows much the same process as it does for undergraduate programs. To formulate a list of graduate learning goals, first ask "What are characteristics of an ideal doctoral graduate in your field?" Graduate programs may want to establish goals aimed at things like:

- Depth and breadth of disciplinary knowledge
- Scholarly and research skills
- Oral and written communication skills
- Ability to teach effectively in the discipline
- Preparation for the breadth of available career opportunities
- Professional socialization, behaviors, and habits of mind
- Professional skills such as working collaboratively, grant writing, & publication

Adapted from Golde, C. M., Jones, L., Bueschel, A. C., & Walker, G. E. (2006). The challenges of doctoral program assessment. *The assessment of doctoral education: Emerging criteria and new models for improving outcomes*. Stylus Publishing, Sterling, VA, 53-82.

## Outcomes

Student Learning Outcomes (SLOs) are **specific** statements of what a student will know or do, written to specify how the student will demonstrate that knowledge or ability. They are measurable, specific, and provide the evidence we need to determine if program goals are being met. Programs and courses both need outcomes. At the course level, outcomes help students know what they are expected to do and how they will be assessed. At the program level, outcomes help programs measure and understand student learning. **Outcomes exist at the point of assessment**, operationalizing broad goals into specific and attainable objectives. Outcomes are the necessary link between learning goals and the instruments or methods we use to determine if they have been met. We ask programs to specify *both* the goal and the outcome that particular assessment measures are associated with.



### Program Learning Outcomes (PLOs)

To understand the role of Program Learning Outcomes, it is important to distinguish between assessing students and assessing a program: Student assessment (which depends on *course* outcomes) asks if a student has demonstrated a skill or learning sufficiently to pass or earn a particular grade. Program assessment uses indicators (which may include student assessment results) to evaluate the health or effectiveness of the *program*. **Program outcomes establish what will be measured, demonstrated, or observed in order to evaluate the effectiveness of the program.**

To know what to assess, programs need clearly articulated outcomes. Program Learning Outcomes articulate, as clearly as possible, what **skills, knowledge, or proficiencies students graduating from the program should have.**

Determining program outcomes doesn't have to be an arduous process. Programs often report on evidence of learning outcomes that have been established at the course level (for example, in capstone courses), in relation to their associated program goals. Course outcomes for advanced, capstone, and other senior level courses are a good place to start.

However, **programs can also establish outcomes at the program level that are assessed outside of particular courses** (for example, setting a benchmark for student performance on a national certification or licensure exam). Any outcome (whether it originates from the course level or program level) that is to be included in an assessment plan as evidence of learning should be formulated as a Program Learning Outcome (PLO).

#### **Distinguishing the Uses of Outcomes: Student Assessment vs. Program Assessment**

A student passes a capstone course because of excellent performance on four out of five required outcomes. However, their performance on the fifth outcome was weak. When student assessment data is aggregated by the program, they discover that nearly half of the students in the course showed weak performance on the fifth outcome: there is a curricular issue with how that outcome is being taught or assessed that the program needs to address.

While starting with already existing course outcomes can help with writing strong PLOs, when adapting them as program outcomes faculty will want to distinguish between performance standards appropriate for particular courses (or points in a student's course of study) and performance standards that are appropriate for a graduating senior (in the case of capstone courses, these may be the same). For example, a formative assessment strategy might gather data about student learning at the end of the first year, allowing faculty to judge how their

introductory courses are working and whether or not interventions or changes are needed to keep a cohort on track to meet the program outcomes before graduating. However, the performance standards for the end of the first year won't be the same as those for a student about to graduate.

It's important that faculty have this conversation and craft program level outcomes that align with their program goals, including making any necessary adjustments at the course level (see below, alignment). Programs need to report on outcomes, and which goals they are associated with, as part of the accreditation process. **While programs are encouraged to articulate, assess, and report on non-academic outcomes that faculty care about (such as student satisfaction, equity, jobs placement, or other strategic goals), the accreditation process is most concerned with program outcomes focused on student learning.**

### Writing Outcomes

Program learning outcomes should spell out **what students graduating from the program should know or be able to do**. The assessment plan should also specify the **conditions** under which their knowledge or skills should be demonstrated (the associated measure), and a **target** for acceptable program performance (e.g., 80% of students will score a 3 or higher).

The example below presents program outcomes and their associated goals:

#### **Program Learning Outcomes**

Students graduating from the Folklore Program will be able to:

PLO1: Identify, describe, and explain several common forms of folklore and folklife.  
[Folklore]

PLO2: Accurately explain and discuss the uses and limits of folkloristic approaches, including structuralism, functionalism, structural functionalism, and performance theory. [Theory]

PLO3: Accurately describe the ideas and impact of historical figures that influenced the field of Folklore Studies. [Theory]

PLO4: Plan and carry out an ethnographic research project, including choosing an appropriate research subject, engaging in participant observation, interviewing informants, and keeping field notes. [Fieldwork]

PLO5: Explicitly and responsibly reflect on the ethical dimensions of the student's ethnographic research projects and ethnographic writing in field notes and a reflective essay. [Ethics]

PLO6: Compose and present effective ethnographic texts in print, film, or digital formats. [Ethnography]

PLO7: Effectively apply folkloristic theories and perspectives to create arguments about the meaning and significance of particular cultural traditions. [Theory]

PLO8: Explain and explore specific examples of the role of folklore and tradition in shaping identity, belief, and community. [Community]

PLO9: Contextualize and explain the social significances of specific traditional, informal, or expressive cultural activities from several different cultures. [Culture]

While avoiding ambiguity is ideal, in some cases, faculty may deliberately embrace it: if there is more than one “effective” way to compose ethnography, it may not be useful to standardize the definition across multiple courses. However, each course must then define “effective” for its students, and faculty would need to plan a strategy for assessing this outcome that accounts for the variations.

Student Learning Outcomes (whether at the course or program level) should be:

- Brief, succinct
- Use a verb that is observable, demonstrable, or measurable
- Reflect actual knowledge or skills faculty want students to achieve
- Use the appropriate cognitive level (Bloom's Taxonomy)
- Be discipline-specific, even when incorporating skills that are part of the Gen Ed curriculum (for example, outcomes dealing with critical thinking should reflect the nature of critical thinking in the discipline)

Example of Student Learning Outcomes for a fictional **graduate** program:

**Program Learning Outcomes: Ph.D. in Folklore**

Students in the Ph.D. program will:

PLO1: Demonstrate understanding of the history, theory, concepts, and debates in at least two major areas of folkloristic research by summarizing, synthesizing, and critiquing relevant literature. [Disciplinary Knowledge, Critical Thinking]

PLO2: Develop and complete an original research project that pursues a significant research question in the field of Folklore Studies. [Research]

PLO3: Demonstrate the ability to ethically conduct ethnographic research, including methods, analysis, and writing. [Methods]

PLO4: Demonstrate oral and written communication skills suitable for professional conference presentation, academic publication (print or digital), applications for grants, awards, and fellowships, and other forms of professional discourse. [Communication]

PLO5: Demonstrate the necessary knowledge and skills for teaching undergraduate courses in the discipline, including the assessment of student learning, by successfully developing and teaching at least one course under the guidance of a faculty advisor. [Pedagogy]

PLO6: Demonstrate professionalization through familiarity with major professional institutions and organizations, publications, conferences, and compliance with professional expectations for ethics, collegiality, and service by attending and presenting at national conferences, engaging in collaborative projects, writing grant proposals, and other professional activity. [Professionalism]

In the above example, most of these graduate level PLOs are implicitly tied to specific artifacts or events where they will be assessed, such as the qualifying exam, the proposal, or the oral defense.

The verbs for outcomes should be observable, demonstrable, or measurable, and should reflect the desired complexity of the activity. It has become standard practice to use Bloom's Taxonomy (specifically, the **cognitive taxonomy**) to write outcomes statements. However, some assessment professionals have made the case that not everything academic programs value neatly fall into the cognitive taxonomy, particularly soft skills, habits of mind, or even

physical abilities (e.g., performing arts programs). Bloom's team also created **affective** and **psychomotor taxonomies** that some departments or programs may find are appropriate for crafting outcomes statements in their disciplines, and Fink also distinguishes useful cognitive and non-cognitive learning experiences (Fink, 2013). There are alternatives to using Bloom's, which will be shared below.

There are some verbs that generally should not be used in outcomes statements, even if they are permissible in goal statements. For example, "to understand" is vague and, by itself, unmeasurable. Similarly, the concept of "critical thinking" should not be used in outcomes statements, because each discipline conceptualizes critical thinking differently--instead outcomes statements should be framed around the discrete activities that are considered to be critical thinking or that demonstrate understanding or ability. The more amenable an outcome statement is to being directly and transparently measured, observed, or demonstrated, the better. Achieving that is a combination of avoiding vaguely or broadly defined nouns (like "critical thinking") and using the right verbs.

Bloom's Revised Taxonomy provides us with verbs and cognitive levels:

Level	Remember	Understand	Apply	Analyze	Evaluate	Create
<b>Definition</b>	Exhibit memory of previously learned material by recalling facts, terms, basic concepts, and answers.	Demonstrate understanding of facts and ideas by organizing, comparing, interpreting, giving descriptions, and stating main ideas.	Solve problems to new situations by applying acquired knowledge, facts, techniques and rules in a different way.	Examine and break information into parts by identifying motives or causes. Make inferences and find evidence to support generalizations.	Present and defend opinions by making judgments about information, validity of ideas, or quality of work based on a set of criteria.	Compile information together in a different way by combining elements in a new pattern or proposing new solutions.
<b>Example Verbs</b>	<i>define describe duplicate enumerate examine identify label learn list locate match memorize name</i>	<i>ask associate cite classify compare contrast convert describe differentiate discover discuss distinguish estimate</i>	<i>apply calculate carry out classify complete compute demonstrate dramatize employ execute experiment generalize illustrate</i>	<i>analyze break down categorize classify compare connect conclude contrast correlate deconstruct deduce detect diagram</i>	<i>appraise apprise argue assess compare conclude consider contrast convince criticize critique debate decide</i>	<i>adapt arrange assemble build collect collaborate combine compile compose constitute construct create design</i>



	<i>observe</i> <i>omit</i> <i>quote</i> <i>read</i> <i>recall</i> <i>recite</i> <i>recognize</i> <i>record</i> <i>repeat</i> <i>reproduce</i> <i>retell</i> <i>select</i> <i>state</i> <i>tabulate</i> <i>tell</i> <i>visualize</i>	<i>explain</i> <i>express</i> <i>extend</i> <i>generalize</i> <i>give examples</i> <i>group</i> <i>identify</i> <i>illustrate</i> <i>indicate</i> <i>infer</i> <i>interpret</i> <i>judge</i> <i>observe</i> <i>order</i> <i>paraphrase</i> <i>predict</i> <i>reflect</i> <i>relate</i> <i>report</i> <i>represent</i> <i>research</i> <i>restate</i> <i>review</i> <i>rewrite</i> <i>select</i> <i>show</i> <i>summarize</i> <i>trace</i> <i>transform</i> <i>translate</i>	<i>implement</i> <i>infer</i> <i>interpret</i> <i>manipulate</i> <i>modify</i> <i>operate</i> <i>organize</i> <i>outline</i> <i>predict</i> <i>solve</i> <i>transfer</i> <i>translate</i> <i>use</i>	<i>differentiate</i> <i>discriminate</i> <i>distinguish</i> <i>divide</i> <i>examine</i> <i>experiment</i> <i>explain</i> <i>identify</i> <i>infer</i> <i>integrate</i> <i>inventory</i> <i>order</i> <i>organize</i> <i>relate</i> <i>separate</i>	<i>defend</i> <i>determine</i> <i>discriminate</i> <i>evaluate</i> <i>grade</i> <i>judge</i> <i>justify</i> <i>measure</i> <i>prioritize</i> <i>rank</i> <i>rate</i> <i>recommend</i> <i>review</i> <i>score</i> <i>select</i> <i>support</i> <i>test</i> <i>validate</i>	<i>develop</i> <i>devise</i> <i>formulate</i> <i>generate</i> <i>hypothesize</i> <i>integrate</i> <i>invent</i> <i>make</i> <i>manage</i> <i>modify</i> <i>organize</i> <i>perform</i> <i>plan</i> <i>prepare</i> <i>produce</i> <i>propose</i> <i>rearrange</i> <i>reconstruct</i> <i>reorganize</i> <i>revise</i> <i>rewrite</i> <i>solve</i> <i>specify</i> <i>structure</i> <i>synthesize</i> <i>test</i> <i>validate</i> <i>write</i>
Level	<b>Remember</b>	<b>Understand</b>	<b>Apply</b>	<b>Analyze</b>	<b>Evaluate</b>	<b>Create</b>

Adapted from Anderson, L. W., & Krathwohl, D. R. (2001). A taxonomy for learning, teaching, and assessing, Abridged Edition. Boston, MA: Allyn and Bacon.

Clifford Adelman, in an occasional paper for the National Institute of Learning Outcomes Assessment (NILOA), suggests an alternate set of verb groups related to the various *activities* faculty want their students to engage in:

Verb Groups:	Example Verbs:
Verbs for student acquisition and preparation of tools, materials, texts	<i>access, acquire, collect, accumulate, extract, gather, locate, obtain, retrieve</i>

Verbs indicating what students do to certify information, materials, texts, etc.	<i>cite, document, record, reference, source</i>
Verbs indicating the modes of student characterization of the objects of knowledge or observation	<i>categorize, classify, define, describe, determine, frame, identify, prioritize, specify</i>
Verbs describing what students do in processing data and information	<i>calculate, determine, estimate, manipulate, measure, solve, test, relate</i>
Verbs describing the ways in which students format or present data, information, materials	<i>arrange, assemble, collate, organize, sort</i>
Verbs describing what students do in explaining a position, creation, set of observations, or a text	<i>articulate, clarify, explicate, illustrate, interpret, outline, translate, elaborate, elucidate</i>
Verbs falling under the cognitive activities of analysis	<i>compare, contrast, differentiate, distinguish, formulate, map, match, equate</i>
Verbs describing what students do when they “inquire”	<i>examine, experiment, explore, hypothesize, investigate, research, test</i>
Verbs describing what students do when they combine ideas, materials, observations	<i>assimilate, consolidate, merge, connect, integrate, link, synthesize, summarize</i>
Verbs that describe what students do in various forms of “making”	<i>build, compose, construct, craft, create, design, develop, generate, model, shape, simulate, write</i>
Verbs that describe the various ways students use the materials of learning	<i>apply, carry out, conduct, demonstrate, employ, implement, perform, produce, use</i>
Verbs that describe executive functions students perform	<i>operate, administer, control, coordinate, engage, lead, maintain, manage, navigate, optimize, plan</i>
Verbs that describe forms of deliberative activity	<i>argue, challenge, debate, defend, justify, resolve, dispute, advocate, persuade, critique</i>
Verbs that indicate how students value objects, experiences, texts, productions, etc.	<i>audit, appraise, assess, evaluate, judge, rank</i>
Verbs that reference the types of communication in which we ask students to engage	<i>report, edit, encode/decode, map, display, draw/diagram, chart (v),</i>
Verbs, related to student communication in groups	<i>collaborate, contribute, negotiate, respond</i>
Verbs that describe what students do in rethinking or reconstructing	<i>accommodate, adapt, adjust, improve, modify, refine, reflect, review, revise</i>

Adapted from Adelman, C. (2015). To Imagine a Verb: The Language and Syntax of Learning Outcomes Statements. Occasional Paper# 24. National Institute for Learning Outcomes Assessment.

Other alternatives to using Bloom's Cognitive Taxonomy for writing Student Learning Outcomes:

Biggs, J. B., & Collis, K. F. (2014). Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome). Academic Press.

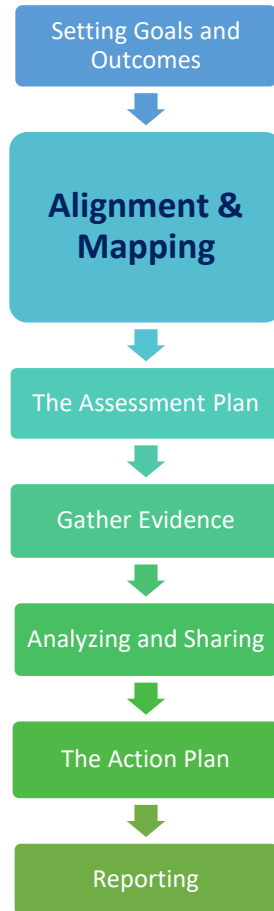
Fink, L. D. (2013). Creating significant learning experiences: An integrated approach to designing college courses. John Wiley & Sons.

Harrow, A. J. (1972). A taxonomy of the psychomotor domain: A guide for developing behavioral objectives. Addison-Wesley Longman Ltd.

Krathwohl, D.R., Bloom, B.S. and Masia, B. B. (1964). Taxonomy of educational objectives, Book II. Affective domain. New York, NY. David McKay Company, Inc.

Wiggins, G. P., & McTighe, J. (2005). Understanding by design. Ascd.

# Alignment and Curriculum Mapping



Alignment and curriculum mapping are two major processes for making sure that goals and outcomes correspond to one another in ways that make sense, and that a program is systematically delivering content and opportunities to students that fulfill their stated goals and outcomes. Both processes are tools for helping faculty to be *intentional* about how their courses and programs fit together, and how those fit into the institution to create a cohesive student experience.

## Alignment

**Alignment** is the process of making sure that course outcomes, program outcomes, and institutional goals correspond with one another. It also can involve making sure that the goals and outcomes that have been articulated correspond with what faculty actually want students to know or do (are they aligned with intention?). The first example of alignment above is the correspondence of program goals with program outcomes--the outcomes are written to operationalize the goals in a way that can be measured, observed, or demonstrated. The link between goals and outcomes should make sense to the faculty that write them--do the outcomes validly offer a way to determine if the goal has been met?

Alignment between the institution, program, and course is usually where we get into trouble, especially if faculty have not explicitly thought about the relationships between course content, programmatic intention, and the university's mission. **Alignment is the process of making those relationships explicit.** It is primarily a bottom up activity, in which courses align with the program and programs align with the institution. It's a fundamental step in ensuring that institutions, programs, and courses are actually teaching what they say their students should be learning.

## UW System Shared Learning Goals

Programs at UWM should consider (and make explicit) how their program goals and outcomes correspond to the [UW System Shared Learning Goals](#). For example, some of the program goals above map to "Knowledge of Human Cultures and the Natural World," to "Intercultural Knowledge and Competence," and to the "ethical reasoning" criteria included in the UW System Shared Learning Goals, although they are all framed in discipline-specific ways.

## Institutional Goals ↔ Program Goals and Outcomes ↔ Course Outcomes

This chart summarizes how the CLOs for fictional Course A align with program goals and outcomes and with institutional goals:

Institutional Goals: (UW Shared System Goals)	Goal A: Knowledge of Human Cultures and the Natural World	Goal B: Critical and Creative Thinking	Goal C: Effective Communication	Goal D: Intercultural Knowledge		Goal D: Individual, Social, and Environmental Responsibility	
	↕   ↕	↕   ↕	↕   ↕	↕   ↕	↕   ↕	↕   ↕	
Program Goals and Outcomes:	Folklore PLO 1	Theory PLO 2, PLO3, PLO 7	Ethnography PLO 6	Community PLO 8	Culture PLO 9	Ethics PLO 5	Fieldwork PLO 4
	↕   ↕	↕   ↕	↕   ↕	↕   ↕	↕   ↕	↕   ↕	↕   ↕
Course Learning Outcomes for Course A:	CLO 1	CLO 3 CLO 7	CLO 4, CLO 7	CLO 2 CLO 8	CLO 8	CLO 9	CLO 6
	PLO= Program Learning Outcomes		CLO= Course Learning Outcomes			Unaligned CLOs: CLO 5	

Engaging in alignment efforts will often reveal issues that require further thought or refinement. It can reveal problems with how goals are constructed or show that the program isn't actually addressing a goal through related PLOs and CLOs.

### Do all course outcomes need to align with program outcomes (PLOs)?

The idea is not that every program outcome, or every course outcome, has to correspond to a higher-level goal. PLO 4 is discipline-specific, and so doesn't need to align with institutional goals. This course seems to have one outcome (CLO 5) that doesn't correspond directly to any program outcomes, although it may correspond to another institutional initiative. Courses may establish goals and outcomes for themselves that go beyond the program's minimum requirements--while that course will strive for internal alignment between its own goals and outcomes, they don't *all* need to map onto program outcomes. But certainly, *some* of the course's outcomes should map to program outcomes. The same is true of a program's outcomes in relation to the institution.

### Who is responsible for alignment?

Aligning course outcomes with program goals and outcomes is primarily the responsibility of individual instructors or groups of faculty in a program, although the program as a whole has an



interest in ensuring that such alignment is happening. Courses with multiple sections or with differing faculty, instructors or TAs across terms should consider adopting standard or core learning outcomes that all versions of the course can address. While perfect conformity isn't desirable, a program also can't be *intentional*

about its course of study without some consistency in what students achieve in introductory courses. These can be challenging conversations for faculty to have, but ultimately, they are productive and necessary conversations to have.

**Articulating and aligning the program's outcomes with institutional goals and initiatives is the responsibility of the program as a whole.** That may mean, where such articulation or alignment doesn't exist, that program chairs and coordinators will need to bring the issue to the attention of faculty and facilitate a process for addressing it.

#### Concentrations or Tracks within a Major

Many departments have "tracks" within them representing different concentrations within their major(s). If one set of program learning outcomes is used (across concentrations or tracks), then courses in each track, through their CLOs, may interpret and fulfill program learning outcomes in their own way. However, when tracks within a program are substantially different, they may need to articulate their *own* program learning outcomes (in addition to any core PLOs essential to the program that are shared across all tracks in the program) to clarify how the broader program *goals* are interpreted and expressed within the track.

Using Goals and Outcomes to Distinguish Tracks:		
Goal 1	PLO 1	Assessed in a course required for all students in the major, or using a similar method in courses for different tracks
Goal 2	PLO2a (Track A)	PLO2a interprets Goal 2 in the context of Track A, and assesses it using a method and assignment appropriate to Track A.
	PLO2b (Track B)	PLO2b interprets Goal 2 in the context of Track B, and assesses it using a method and assignment appropriate to Track B.

Courses within each track would focus on aligning with both shared PLOs, and with the PLOs specific to their concentration.

#### Helping Faculty, TAs, and Instructional Staff with Alignment:

One way to ensure alignment is happening is to facilitate the process. Invite instructors to bring their course learning outcomes to a workshop where they can get feedback from fellow faculty on how their outcomes align with program outcomes.

### Alignment within Courses

While the focus of this guide is assessment and alignment for programs, **alignment issues may also exist between course assignments and assessment methods and course outcomes.**

Aligning course outcomes to the program only helps if the actual content, assignments, and assessments being used in the course *also* align with those outcomes.

### Curriculum Mapping

Curriculum mapping is the second crucial process for ensuring an intentional curriculum plan. It helps programs to systematically deliver content and opportunities to students that fulfill their stated goals and outcomes. Curriculum mapping is a formalized process of making explicit where in the course of study each program goal or outcome is being addressed and to what degree or level. For example, it can clarify in which course students will be first introduced to a concept or skill from a program goal, which courses will develop their understanding or facility with that skill, and in which course they will demonstrate mastery appropriate for a graduating senior. It can help programs identify where in the curriculum program outcomes may need reinforcing or more practice, identify outcomes needing higher level of achievement, or where individual courses may be able to shift assignments or assessments in order to achieve stronger alignment and improvement in student learning.

It doesn't have to be overly complicated to work. A simple chart listing program outcomes and courses is a good place to start. Only courses (or course groups) that are *required* should be listed in the curriculum map. Because the program outcomes are what *all* students graduating from the program should achieve, it follows that the outcomes need to be mapped onto courses that all students take. Purely elective courses do not need to be included in a curriculum map.

In the case of course groups or clusters, where students only need to take one or two courses from a designated group of courses, it's important that each course in the cluster is serving its intended role. Because the program doesn't know which combination of courses a student will take, all of the courses in the group must have a core set of *shared learning outcomes* aligned with the program outcomes that fulfill their role in the curriculum plan. Those courses might also have other outcomes aimed at particular tracks or concentrations, or might incorporate additional program outcomes.

### A Simple Curriculum Map:

	Course A	Course B	Course C	Courses D, E, F	Course G	Courses H, I, J, K, L	Course M	Capstone Course N
PLO 1	I, R	R	R	R	M	R	M	M
PLO 2	I	R	R	R	M			M
PLO 3	I	I	R				M	
PLO 4	I	I		R	R		R	M
PLO 5	I	R			R			M
PLO 6	I, R	I, R			R	R	M	R, M
PLO 7	I	R	R		R		M	M
PLO 8	I		R	R	M	R, M		
PLO = Program Learning Outcomes		I= Introduce R= Reinforce, developing competency M= Mastery, advanced competency						

Curriculum maps can be created for tracks, minors, certificates, or any program of study. They help programs to be more intentional about where and how their program objectives are being met. But they also point to opportunities for both formative and summative assessment. A program assessment plan will use the curriculum map to identify the key opportunities for assessing the learning that's been happening. While the Capstone course is an obvious opportunity for assessing higher level mastery (i.e., doing summative assessment), that may not be the only point that the program wants to gather data from. Some programs will want to also gather formative assessment data at the end of their first or second year courses, both to track how well those courses are doing at helping students develop expected competencies, and to identify shortfalls that the program may want to address while they still have the opportunity with that cohort of students.

### Analyzing Your Curriculum Map

Some things to consider:

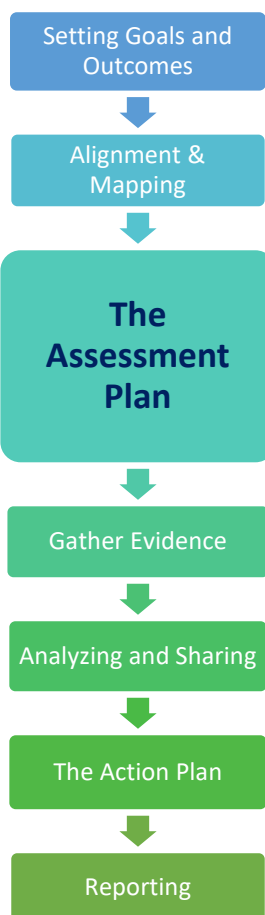
- Are students actually taking courses that *introduce* concepts and skills before they take intermediate and advanced courses?
- Do the verbs in course outcomes at each point in the curriculum map appropriately and clearly articulate the complexity of the learning or skills expected *at that point* in the course of study?
- Does the *assessment* happening at each point (in each course) reflect the desired level of mastery? Is it geared too high or too low?

- Are students being given adequate opportunity in the curriculum plan to develop appropriate mastery for each of the program outcomes?

(Mis) Applying Bloom's to a Curriculum Map

The CLOs for each course should enact the program's learning outcomes at a level appropriate to the course. However, Bloom's taxonomy does not imply that introductory courses should only focus on knowledge or understanding. Designing lower level courses to avoid higher cognitive levels can result in missed opportunities. While knowledge is a foundation for more complex activities, introductory students are also capable of performing learning tasks that ask for analysis, evaluation, or creating at a level *appropriate* to them. Doing so is likely to help students retain more and be more engaged with the work of the course.

## Part 3: Designing an Assessment Plan



Designing an assessment plan includes developing a mindset and making decisions about:

- What question(s) your assessment plan is asking
- The types of assessment your program needs to do
- What evidence you will use
- How to make your plan sustainable
- Setting benchmarks, targets
- Going beyond the essentials of assessment

### Assessment Plans as Programmatic Research

When designing an assessment plan, it is helpful to think about it as a form of research. While few assessment plans rise to the level of experimental research, they are **nonetheless amenable to systematic approaches**. A strict experimental design (i.e., with control groups) isn't necessary (and could be problematic or even unethical), because the audience and purpose of assessment research is ultimately different than for other kinds of research.

The purpose of assessment is local; rather than seeking to draw broad conclusions about teaching X in general, assessment is about understanding and improving our teaching and curricular effectiveness

here at UWM. Assessment is a kind of evaluation research that sometimes incorporates elements of “action research” (e.g., Kemmis and McTaggart, 2000), in which the focus is directly on improving instructional practice or curriculum design rather than making broad generalizations (Suskie, 2018). While different in intention, local assessment still requires a systematic approach with attention to validity and reliability, collecting and analyzing evidence, and to understanding the limits of the conclusions drawn. Where possible, it still aims to illuminate issues about student learning that concern faculty and uncover convincing evidence of causes.

### Asking Questions

In Part 1, assessment was framed as a kind of research that asks questions. Like any research, if we don't know what we are trying to find out, it's hard to determine what kind of data we need or how to get it. This part of the process is where it most frequently breaks down: If faculty don't ask questions that they want to know the answer to, then they are likely to view the assessment process as a waste of time motivated by outside forces (i.e., accreditation), rather



than by their own investment in teaching or in their program. Faculty already care about their students and what's happening in their programs, and they often already know about problems in their programs (although they might not know what the cause is or how to address it). In this sense, assessment can be part of a *problem-based* practice of assessment, in which the assessment process itself is used to study these problems, investigate causes, and formulate solutions (Maki, 2012, Chapter 4). In order to do that, programs need to be clear about what they are asking when they design an assessment process.

Example Assessment Questions:

- Which students did better on PLO2, and why?
- How does our program compare with similar programs at other institutions?
- What effect did the recent curriculum change have on student achievement for PLO7?
- Which knowledge and skills are students not successfully transferring from Course X into Course Y, and why?
- Did students taking the prerequisite course first actually do better in Course X on PLOs 4 and 5 than those who tested directly into it?

To answer any of these questions, we will also end up gathering data addressed to the most fundamental assessment questions:

- Did students learn or become proficient in X?
- *To what degree* did students learn or become proficient in X?

## Different Kinds of Assessment

Program assessment involves multiple kinds of assessment, including the assessment of student learning and other forms of program evaluation that are both part of an ongoing process of continuous improvement. The distinction between these kinds of assessment and evaluation work is important in understanding why some kinds of evidence, while nonetheless important for program evaluation, are by themselves insufficient to satisfy accreditation requirements for the assessment of *student learning* (i.e., outcomes assessment). The kinds of evidence gathered by a program assessment plan, however, are often addressed to both outcomes assessment and program evaluation, and in reality these two processes overlap significantly. Outcomes assessment is a central tool of program evaluation.

The chart below distinguishes between outcomes assessment and program evaluation:

Program Assessment	
Outcomes Assessment	Program Evaluation
<ul style="list-style-type: none"> <li>• Assessment of student learning goals and outcomes</li> <li>• Required for accreditation</li> <li>• Focus on student learning: How well are students learning what is intended?</li> <li>• Continuous improvement of student learning</li> <li>• <b>Requires direct evidence</b> of student learning for <i>specific</i> outcomes</li> <li>• Uses indirect evidence of student learning to provide context and help interpret direct evidence</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Incorporates outcomes assessment results</b>, and assessment of other, non-academic, program goals and outcomes</li> <li>• Required for program review</li> <li>• Focus on the program: goals, curriculum, policies, professional development, big picture outcomes (job placement, student satisfaction, etc.)</li> <li>• Continuous improvement of the program</li> <li>• Uses other <b>indirect</b> and <b>supporting</b> evidence to evaluate strategic goals and program efficacy, and to interpret the results of outcomes assessment</li> </ul>

## Kinds of Evidence

Once outcomes and assessment questions have been established, faculty will need to decide what evidence they need to gather to answer those questions. As a baseline, **all programs should regularly collect at least minimal direct evidence of learning such as the % of students achieving acceptable proficiency for each program outcome.** However, programs should ideally collect *both* direct and indirect evidence, and use a variety of different assessment methods, to assemble the most complete and nuanced picture of their program.

### Types of Evidence

**Direct evidence** measures student learning by examining student work or performance. They are usually quantitative measures, offering insight into what and to what extent students have learned through evaluating exams, papers, performances, observations, or other artifacts of student work. Direct evidence is compelling in that it offers the ability to make judgements about the relative degree of learning or mastery that students have achieved. Direct evidence's power, however, relies on *validity* and *reliability*. It's important that such measures *narrowly* measure the specific outcome they are tied to and produce accurate data consistently. For this reason, grades and GPA data are not a form of direct evidence.

**Indirect evidence** suggests that learning has taken place (but without the capacity to quantify *how much* learning) and can often provide important insight about or context for interpreting direct evidence. Indirect evidence can be either quantitative or qualitative. It may also be the only kind of evidence available for aspirational program goals aimed at developing particular dispositions, habits of mind, or attitudes in its students. As such, most assessment plans should try to include some kinds of indirect evidence in conjunction with direct evidence.

**Supporting Evidence** is other data that doesn't necessarily illuminate student learning of program outcomes, but which can provide necessary context or help programs design their assessment plan to address questions such evidence has raised. Such evidence can come from analyzing program documents, program and institutional data (pass rates, retention), or other institutional records.

The chart below details examples of assessment evidence and their most common uses:

Examples of Assessment Evidence and Uses					
Data	Type	Program Assessment Use(s)	Assessment Tool	Who or What is Analyzed?	What Can be Assessed?
Achievement Tests	Direct	Outcomes Assessment	Test score analysis Content analysis Scoring rubrics Pre/Post Tests	competitions embedded questions on exams locally developed exams oral thesis defenses oral exams, recitals standardized tests	mastery and knowledge of principles, skills value-added
Student Academic Work	Direct	Outcomes Assessment	Content analysis (e.g., examining a random sample of student writing from within a program) Scoring rubrics Embedded assessment (using existing course assessments as program evidence)	capstone course products homework, papers <a href="#">signature assignments</a> portfolios presentations performances, publications research reports term papers, theses	mastery and knowledge of principles, skills values processes value-added

Self-reports	Indirect	Outcomes Assessment, Program Evaluation	Student self-assessments Focus groups/interviews Phone surveys/interviews Alumni surveys Employer surveys Reflective essays Surveys (local or standardized) SEIs (Student Evaluations of Instruction) Exit interviews	alumni, employers enrolled students faculty graduating students entering students supervisors parents staff	Perceptions about: campus climate perceived learning evaluate processes value-added educational outcomes attitudes values
Observations	Indirect	Outcomes Assessment, Program Evaluation	Case studies Observations	campus events (sports, theater) classes club meetings faculty offices fieldwork sites student services offices	attitudes campus climate interactions processes services student involvement student learning
Campus Records & Other Documents	Supporting	Program Evaluation	Course and program outcomes Syllabi Analysis of forms GPA data Grade breakdowns (% of As, Bs, etc.) Course pass rates Graduation and retention data Job placement data Graduate school admittance rates # or % of course repeats Enrollment data	administrative units departments, programs student services offices course syllabi student transcripts/records	accuracy cohesion/consistency efficiency structure for promoting objectives processes equity non-academic program outcomes/strategic goals Program efficacy

Includes material adapted from California State University, Bakersfield, *PACT Outcomes Assessment Handbook* (1999) and UMass Amherst, *Program-Based Review and Assessment* (2001)

*Finding a Balance:* Some programs may want to collect a lot of quantitative (direct) data, using only some qualitative or indirect evidence as a means to contextualize or interpret that data. Other programs may elect to collect only the most basic quantitative data (e.g., the percent of students meeting a benchmark compared to a target %), and instead rely more heavily on qualitative data such as surveys of students or faculty, exit interviews, or focus groups. This decision will reflect not only disciplinary norms, but also program size, available resources, time, and expertise.

## Using Rubrics

Many, or even most assessment plans will involve the use of rubrics as a means to collect direct and quantitative evidence of learning. Rubrics can be used for summative or formative assessment, and for either student or program assessment. If a rubric is well-designed and well-aligned with both course and program outcomes, it can sometimes be used simultaneously for *both*

### Grades as Supporting Evidence

Grades are rarely designed to reflect *specific* learning outcomes, but rather combinations of outcomes and student behaviors. As evidence of overall student performance, they do their job, but **as evidence of the effectiveness of a program, they are messy and complicated indicators at best.**

Assessment is more effective when evidence specific to each outcome is separately gathered and reported. For this reason, **GPA data or course grades by themselves are generally not helpful evidence in program assessment.** Grades are, however, a useful place to start. Programs should regularly gather evidence about pass rates or grade breakdowns (what % of students earned As, Bs, Cs, and so on) and use them to look for red flags. The grading process itself is also an ideal opportunity to gather assessment data, since grading usually also involves some estimation of proficiency (Walvoord, 2012, p. 575-6). Assignment grades can work as assessment data *if they are attuned only to one learning outcome.*

student assessment and program assessment (a type of built-in assessment). However, given the differing purposes of student and program assessment, creating different rubrics for each is sometimes preferable. Using separate rubrics allows them to be more closely tailored to their intended use, i.e., an instructor can create a rubric closely aligned with their specific assignments and course outcomes to assess students, while using a more general program rubric to assess student learning for the program. Using a single rubric for both requires more careful attention to rubric design and the role the rubric plays in determining individual grades but can save time and work in conducting program assessment.

There are several different kinds of rubrics that programs can choose from that each have different strengths and weaknesses. The chart below details rubric types.

Kinds of Rubrics			
Type:	Description	Pros/Cons	Uses
Checklist	pass/fail; element is present and satisfactory or it is not	Ease of use. Measures minimal competence, or adherence to a procedure. Does not gather evidence of the degree or learning or of advanced competence. Does not provide feedback for students.	Limited uses for student assessment and program assessment.

Minimal Rubrics	Outcomes or traits are rated on a scale (e.g., 0 1 2 3) corresponding to levels (e.g., does not meet, minimally meets, meets, exceeds)	Provides evidence of levels of learning or mastery. Rubric doesn't explain what each rating means; users "norm" to agree on how to use the scale. Does not result in usable feedback for students.	Program assessment.
Analytic Rubric	Lists outcomes or traits, and clear levels of performance for each.	Provides evidence of learning and levels of learning. Provides students with feedback for each outcome or trait.	Student and program assessment. ("Gold" standard for rubrics.)
Holistic rubric	Defines levels of performance for all outcomes or traits at once, often with a narrative description of each performance level.	Holistic determinations often require "judgement calls" because students rarely meet all criteria for a level equally. Does not provide usable data on individual outcomes.	Student assessment.
Developmental Rubric	Defines levels of achievement or mastery for each learning outcome or trait tracked across a course of study (i.e., performance expectations for first year, second year, third year, and program graduates).	Track learning over time using established benchmarks or levels of mastery. More general than most analytic rubrics, these are less useful for summative student assessment, but work well for program assessment.	Program assessment, formative student assessment. <a href="#">VALUE rubrics</a> are an example of these.
Structured Observation Guide	Rubric lists outcomes or traits with a brief description of the highest level of performance for each outcome, and then space for the reviewer to comment on the student's performance.	Qualitative only. Provides useful feedback to students.	Student assessment. For program assessment, a meta-analysis of instructor comments is needed.

Rubric types and descriptions adapted from:

Suskie, Linda. (2012) "Rubric Development." In Secolsky, C., & Denison, D. B. (Eds.). *Handbook on measurement, assessment, and evaluation in higher education* (pp545-7). Routledge.

### Validity and Reliability

Rubrics need to be carefully constructed to ensure *validity* (the rubric validly measures the outcomes it purports to measure) and *reliability* (the rubric produces consistent scores over time and across different users). Validity can be established by comparing the ratings or scores a rubric produces with another known instrument that measures the same outcome, by testing the rubric over time, and gathering feedback from raters or faculty about how well they believe the rubric captures their view of student work after using it. Rubrics may need to be revised several times over a course of months or years to improve validity.

The ability of different rubric users to produce the same or consistent scores when evaluating the same artifact is referred to as *interrater reliability* (Maki, 2012). If faculty disagree about what a rubric means or how to apply it, scores become

increasingly less reliable and meaningful. A rigorous process of norming or calibration can help ensure that faculty using a rubric are consistent interpreting and applying the performance levels.

**All rubrics, when used by more than one rater, need some kind of norming or calibration activity.**

Even highly detailed rubrics can produce varying interpretations across different faculty and need to be discussed in relation to a range of examples.

*Norming/Calibration:* In the most common calibration process raters are asked to independently score a sample of student work that represents a range of performance across levels. Raters then meet to discuss their scores and resolve any differences. They repeat this process until all of the raters are producing consistent scores (Maki, 2012, “Strategies to Ensure Interrater Reliability”). Another strategy is to ask raters to read a sample of student work with a wide range of performance, assigning letters or numbers to each sample. Raters use the rubric to score each sample and discuss their findings. The raters then agree on at least two samples that exemplify each performance level for every outcome being measured. These are referred to as “anchors.” The anchors then serve as a point of reference for using the rubric and can be used to train other raters to use the rubric (Oakleaf, 2009).

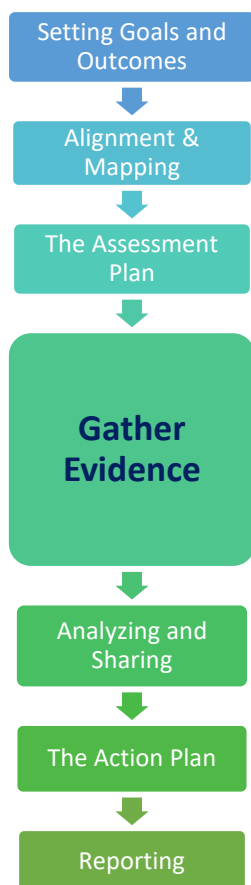
*Double and Blind Scoring:* While it is not always possible, the validity (and authority) of rubric scores is greatly improved by using multiple raters for each artifact being assessed (Suskie, 2018, p. 155). Each example of student work being assessed is reviewed and scored independently by two raters. Where their scores are closely normed (identical or differing only by one degree/level), the mean score of both ratings is taken. For each outcome where those two raters disagree by more than one degree or level, a third rater also assesses the artifact and offers their own score. The resulting scores are then averaged to produce a mean rubric score for each outcome. The average of all three raters will be the best *representation of the group consensus* on that student work, and generally produces the most reliable scores (Johnson, Penny, & Gordon, 2000). If a large number of student artifacts require third reads, it can also indicate that the group of raters is not well-normed and needs to be re-calibrated.

### Intra-rater Reliability:

Intra-rater reliability is the ability of raters to give consistent scores *over time*. There can be shifts in how raters are using a rubric from year to year, even if they are well-normed with each other in each year. If a program intends to compare their data with past data, this can be an important factor to consider. One strategy used by the Composition program at UWM was to include “controls”—examples of student work that were assessed in a previous year and masked to appear part of the current set. Current and past scores on the same examples were then compared to determine if there was a change in how raters were using the rubric.

Where time and resources permit, one way to set up a double scoring system is to create a separate assessment event in which a random sample of student work (preferably a portfolio, or a common final project or essay) from the academic year is assessed with a common rubric by a group of expert raters (faculty, teaching academic staff, trained TAs). The student artifacts will often be anonymized, removing identifying information for both the students and their instructors (called *blind scoring*), to reduce any possible bias in the raters. The event begins with a norming or calibration exercise to ensure raters are applying the rubric consistently. While ideal as a research practice, double and blind scoring do have some drawbacks. They literally double the assessment work being done, and it can be difficult to get enough raters to donate their time. Funding to pay lecturers or TAs to participate may be needed.

*Using Expert Raters:* Program assessment rubrics should ideally be used by expert raters familiar with disciplinary norms in the program. Faculty and teaching academic staff themselves are the best choice, followed by trained TAs. Generally, programs should not use undergraduate readers or outside raters unfamiliar with the program. However, when using national or standardized rubrics, expert outside raters may be an option.



## Planning for Sustainable Assessment

Each program should have a general assessment plan that allows for the *ongoing* collection and analysis of evidence related to the program outcomes. The ongoing plan might involve gathering data each term from a capstone course, or some other fairly low-maintenance (and thus sustainable) process that allows a program to actively monitor how things are going. Programs may also have periodic, special assessment projects designed around *specific* questions, or might set up a system where they do an intensive special assessment project every 3 to 5 years in order to gather more detailed information about their program.

Ongoing assessment plans need to be sustainable, which means they also need to be manageable for the faculty and staff involved. To develop a sustainable plan, programs might ask:

- What is your program already doing that can be used in the assessment plan?
- What is manageable, given time, faculty, and resources available?



- What needs to be assessed annually, and what could be assessed on a rotation?
- Are there outcomes that can be assessed at the same time (e.g., using a single exam or rubric)?
- Where are opportunities to use **embedded assessment**? (Required elements of a course that can provide evidence for program outcomes)

The curriculum map is often a good starting point for determining workable points of assessment. For summative assessment, courses in which students are expected to demonstrate advanced competencies or content mastery (such as capstone courses) work well. Programs and courses already produce evidence of learning in the form of assignments, exams, essays, presentations, and other forms of student work.

### Embedded Assessment

Embedded assessment is when assignments or exams at the course level, done as a graded requirement of the course, provide the necessary evidence of learning for program assessment. Designing embedded or built-in assessment saves the most time and labor in the long run, but initially may require some extra work to ensure that all faculty and staff are involved and have made the necessary adjustments to allow their in-course assignments to work as evidence in the program assessment process. Student assessments, such as exams, might need minor adjustments to also provide program evidence tied to *particular* learning outcomes. Once set up, a system of embedded assessment is almost self-sustaining.

Embedded assessments should make explicit how exams or assignments relate to program outcomes and should ensure that data for each included outcome is being collected separately (e.g., an analytic rubric that provides evaluation of individual outcomes is preferable to a holistic one that lumps outcomes together). The grading process, if it is sufficiently broken down so that *individual* learning outcomes are graded separately, can provide usable program assessment data. However, because the goals of grading and of assessment differ, overlapping these processes can be problematic. Grading processes may be too narrowly tailored to the course or assignment, account for factors beyond proficiency with particular outcomes, combine outcomes, or weight some outcomes as more important than others. Most embedded assessment strategies work to **keep student grading processes and assessment data collection separate**. Grades for an assignment can work as assessment data if that grade *only* reflects proficiency in one specific learning outcome.

A student artifact, while graded for the course, might be assessed using the program rubric well after the course has ended (particularly if the program is using a double scoring system), or might be scored simultaneously using separate grading and assessment rubrics. Many GER

courses at UWM use an assessment method like this, where faculty are asked to evaluate a final paper or project using a common GER rubric and report the data to their department.

### Signature Assignments

One embedded assessment strategy is to develop a [\*signature assignment\*](#), in which a common assignment is given to students in several different courses that gives them an opportunity to demonstrate proficiency in one or more program learning outcomes. Faculty might adapt the signature assignment for their courses but maintain core elements of the assignment that allow them to be assessed using a common rubric. This strategy can be used for final projects, papers, or even specific exam questions that are shared across different courses. The shared element is evaluated using a common rubric or standard, and the resulting data is aggregated at the program level to measure student learning of the associated learning outcome.

### Student Evaluations of Instruction (SEIs)

Student evaluations, or end-of-term evals, offer an opportunity to gather qualitative or indirect evidence for program assessment. Some departments divide their SEIs into sections, with one asking students to provide traditional feedback on instruction while another section gathers feedback on the *course*. They can include self-evaluation questions about skill improvement or knowledge acquired (although survey fatigue can be an issue if the evaluation form gets too long or has too many open-ended questions). Particularly for programs where many faculty, TAs, or staff teach the same course, this strategy can provide useful information to the program when data from the course and self-evaluation questions is aggregated.

### Add-on assessments

Add-on assessments are assessments outside of normal graded course requirements. For example, students may be asked to sit for a special exam, take a Qualtrics survey, or participate in a focus group. Add-on assessments can generate powerful and useful information, but they can also require careful planning. Students may not be willing to give them much time or energy if they perceive them to be unimportant or “extra,” and they may require an incentive to participate. The students who participate may also not be truly representative of the student population.

Suggestions for improving participation in add-on assessments:

- Convince participants of the importance of the assessment activity
- Appeal to student’s self-interest
- Be sensitive to survey fatigue
- Minimize the inconvenience of the assessment

- Keep the assessment short
- Keep the assessment clear
- Make the assessment engaging and fun
- Provide a material incentive to encourage students to participate
- Give recognition to top scorers or the first students to return a survey
- Make participation in the assessment a requirement of the program or course (when appropriate)

*Adapted from Suskie, L. (2018). Assessing student learning: A common sense guide, 3rd edition (p. 280-282). John Wiley & Sons.*

### Sampling

It's not always possible to gather assessment data from every student in a program. Some general education courses have thousands of students each year and may be more practical to assess using sampling methods. A simple random sample is one in which each student in a program has the same chance of being selected. There are a lot of software or online random number generating tools that can help with this. For example, if all of the students in a program are listed and numbered on a spreadsheet, a random number generator can be used to select students for the sample.

*Representation:* When using a sample, keep in mind the *margin of error*. The error margin is about how representative of the whole group your sample really is. If 75% of the sample scored "satisfactory," with a margin of error of +/- 5%, this means that we can be 95% sure that the "actual" figure for the whole group is somewhere between 70-80%. To achieve a margin of error of 5% for a population of 1,000 students, you would need a sample size of 278 students. There are several margin-of-error and sample size calculators online that are free and easy to use. Another issue with representation is ensuring the sample demographically represents your student population. It is possible to construct the sample to be representative by trying to ensure that the right proportion of key demographics are present in the sample. If the sample isn't representative, it may be worth weighting the data to make it more representative.

## The Assessment Plan Essentials

The essential component of an assessment plan for accreditation purposes at UWM is one that is ongoing, systematic, and produces **direct** evidence of learning for each program learning outcome. To meet accreditation requirements, departments and programs are required to have an assessment plan that allows for **direct evidence of student learning to be collected on at least one program learning outcome each academic year**, on a rotating cycle, for both

undergraduate and graduate programs. Ideally, all of a program's learning outcomes should be assessed within a five-year period.

You will need to make decisions regarding:

- What type of direct evidence you want to collect
- What the benchmarks will be
- What the targets will be

#### Determine the type of evidence

What type of evidence should you collect? The evidence collected should be narrowly tailored to measure each program learning outcome separately (e.g., course grades are not direct evidence because they reflect multiple outcomes and behaviors together). The evidence should also enable the program to detect patterns, strengths, and weaknesses in relation to specific learning outcomes, and so should ideally provide a more detailed evaluation than pass/fail. Some examples of assessment focusing on the essentials is below:

#### **Example #1: Using a Rubric**

Final projects from a program's capstone course are evaluated using a developmental program rubric that lists 5 program learning outcomes. The program rubric is separate from the grading rubric the instructor uses. Each outcome is rated on a scale of 1-5.

*1= Not present*

*2= Beginning competency*

*3= Developing competency*

*4= Advanced competency*

*5= Expert competency*

Faculty assessing this course are well-normed, having defined each level carefully and achieved consistency in how they interpret and apply the rubric. They report to the program how many students scored at each level (and note any enrolled students that were not scored, e.g., did not complete the final project). The program has established a score of 4 as the **benchmark**, the minimally acceptable score for a graduating senior for each outcome. Their **target** is for 80% of all capstone students to score a 4 or better.

#### **Example #2: Using Exam Questions**

A science program has an advanced course, offered every semester, that all majors are required to take. Both the midterm exam and the final exam have *specific essay questions*, worth 50pts each, aimed at assessing two of the program's essential PLOs, student's mastery of disciplinary knowledge and ability to apply scientific method to answer real-world questions. Graders are provided with clear guidelines for grading, giving all essay

responses that meet the minimum benchmark a score of 30/50pts or higher, regardless of other factors. The instructors report the scores for these **two questions** to the department, but not the overall exam score (which also tests for other course learning outcomes). Because these outcomes are considered essential for all majors in the program, their target is for 90% of all students to score at least 30pts on these exam questions.

#### **Example #3: Using Committee Evaluation**

After each PhD student completes their dissertation proposal defense, the dissertation committee members meet to discuss their performance. After each defense, they systematically consider three learning outcomes they have identified for the proposal phase, rating candidates on *each* outcome as Outstanding (3), Satisfactory (2), Needs Revision (1), or Unacceptable (0). They also note on the form the specific strengths and weaknesses they noted. They turn in these forms to the graduate program coordinator, who keeps track of the scores and the frequency with which particular strengths or weaknesses are noted in the comments, allowing the program to note any patterns that emerge. The program requires 100% of Ph.D. candidates to earn at least a score of (2) on all outcomes to proceed with their dissertation, and so only rarely does the program fail to meet its target of 100% of students hitting this minimum benchmark. Instead, the program sets an exemplary target of 80% of all Ph.D. candidates earning a score of (3) on these learning outcomes (while 100% earn at least a 2).

In Example #1 above, a rubric is used in a capstone course to gather assessment data about five of the program's PLOs. In Example #2, a small number of standard exam questions are included in all versions of a course from semester to semester, even though other exam questions may change. In Example #3, faculty formalize and capture things they were already considering in evaluating a dissertation proposal, in a way that allows the program to track and detect patterns that could be addressed.

#### Setting Benchmarks and Targets

**Benchmarks** (sometimes called standards) establish what level of performance for a program outcome is acceptable for a program graduate. The **target** establishes what percent of students should be achieving that benchmark for the program to consider itself as succeeding in its core educational mission.

In short, both measure success: **benchmarks allow us to judge student success, while the target allows us to judge the program's success.**

Benchmarks should be set for each program learning outcome. Benchmarks are closely linked to the methods of assessment or measurement: if a PLO is measured with a rubric using a scale

of 1-5, is the minimal competency expected of a graduating senior going to be a 3 or a 4? If the PLO is measured using a national standard exam, what should be the minimum expected score of a program graduate?

To set useable benchmarks, Suskie (2018, p. 297) suggests that programs:

- Ask what would not embarrass you?
- Ask how will the assessment data be used (and by what audiences)?
- Ask what are the relative risks of setting the bar too high or too low?
- When in doubt, set the standard relatively high rather than relatively low
- If you can, use external sources to help set standards (disciplinary organizations, professional licensing requirements, etc.)
- Consider the assignment being assessed
- Consider a sample of student work and past experience

Benchmarks are necessary to make use of assessment data, although they are primarily aimed at the most basic assessment questions (What and to what extent did students learn?). In creating a program rubric, the benchmark will be whatever column on the rubric defines the *minimally acceptable* level of performance.

**Program targets** identify what percent of students should be achieving the minimum competency on each learning outcome for the program to be succeeding. To do so, Suskie suggests faculty distinguish between *essential* outcomes and *aspirational* outcomes. Essential outcomes are so fundamental that nearly 100% of students should achieve the minimum standard before graduating. Aspirational outcomes, however, are those where some students could still be successful even though they did not master the outcomes completely. Aspirational outcomes may therefore have a lower target percent, although Suskie argues that no target should ever be below 50%. Programs will also benefit from establishing a secondary, “exemplary” target that identifies what percent of students a program would like to see achieving the *highest* or most exemplary standard (Suskie, 2018, p. 300-302). Exemplary targets can help programs make more nuanced determinations about where they are succeeding and where they could do better. **Programs can set different and appropriate targets for each learning outcome.**

Example Program Targets		
	Minimal Target: (% achieving the minimum competency)	Exemplary Target: (% achieving exemplary competency)
Essential Outcomes	90%	50%
Aspirational Outcomes	70%	30%

### Essential Graduate Program Assessment

Graduate programs should report the outcomes of the thesis or dissertation defenses as part of their assessment plan. However, **this data needs a finer grain of detail** than pass/fail in order to be of use in evaluating and improving the program. To use this process effectively for program assessment requires capturing slightly more information—information that faculty already have in hand when they assess the *student*, but which may be lost if all that is recorded is whether the student passed. For each *specific learning outcome* the program has established, a simple scoring system such as Outstanding (3), Satisfactory (2), Needs Revision (1), or Unsatisfactory (0) can be used by the committee. The committee can agree on a single consensus score for each outcome, or each member of the committee can record their own score on a rubric. While this process may help focus discussion about the candidate, its purpose is to capture a more detailed picture *for the program*. Even the dissertation or thesis itself can be scored with a rubric (see Lovitts, 2006, p. 174-176 for an example of a dissertation rubric). While passing a thesis defense *is* direct evidence of a minimally acceptable level of competency in conducting original research, this data is only useful in the accountability/compliance mindset. By itself, this data provides graduate programs with almost no useable information that could help them improve the program. As a good first step, many programs provide a holistic, overall score similar to the above scale, that provides slightly more information. However, a more ideal assessment plan offers assessment data on *each* learning outcome individually so that the program can identify patterns of programmatic strengths and weaknesses.

## Assessment Beyond the Essentials

To produce meaningful results, or to seek answers to burning questions, most programs will want to go beyond the “essential” plan in some way. For some that may mean setting up a more nuanced and complete system of ongoing assessment, and for others it may mean sticking to the “essential” plan annually, while planning special assessment initiatives every couple of years to dig deeper. Even on the most basic assessment plan, there is no reason to limit assessment to one outcome per year if you can do more—most programs will find that it is

possible to assess several outcomes each year, and some may have opportunity to assess *all* of them annually.

As a general best practice, **learning outcomes should be assessed with *both* direct and indirect evidence**, so supplementing the assessment plan with a student survey, focus groups, or exit interviews can greatly enhance both a program's understanding of itself and ability to interpret the direct evidence. Indirect evidence, such as alumni surveys, can also help a program understand how students perceived and experienced the program itself, including what they believe they learned and how they see themselves using what they learned down the road. Sometimes, the results of such investigations are surprising.

While GPA, course grades, pass rates, and retention are vital statistics, by themselves they can still leave much of what's going on in a program hidden. Digging deeper into the heart of a program can require different kinds of evidence to be gathered, including disaggregated data or using assessment data with identifying student information intact. As long as such data is handled carefully, in compliance with FERPA regulations, programs can track student learning across several courses and potentially engage in a systematic examination of the progress of learning across their course of study. Such detailed information can be vital for informing efforts to restructure a course of study, adjust curriculum, change policy, or make other meaningful changes to a program.

#### **Assessing a New Course**

When UWM's First Year Composition program replaced its non-credit 095 basic writing course with a mainstreamed English 100 course, they tested the new curriculum and course structure by setting up a series of comparisons. English 100 was now equivalent to English 101, but carried with it an extra credit hour and additional support structures for its students. EN100 portfolios were scored using a double and blind scoring method, and those scores were compared to EN101 portfolios scored in the same way. Because students from the old 095, the new 100, and from 101 eventually had to take English 102, it was important to compare outcomes in 102 as well. Portfolios from 102 were scored, with representative samples from students that took 100 and that took 101. These results were then compared with previous years' scores of 095 students who took 102. This process provided convincing evidence that students in the new 100 course did as well achieving the minimum benchmark on key learning outcomes as those in 101, and did better in 102 than those who had taken the old 095. Overall, the new course was a success.

The only limitations to what can be assessed are the time, resources, and desire to do so. If your program has identified an assessment need, but lacks resources or support to carry out



the assessment research, make sure that both your dean and the assessment coordinator in Academic Affairs know. While it is not always possible to receive additional financial support, garnering more funding for assessment work depends on decision-makers knowing that the need exists.

#### Additional Graduate Program Assessment

For outcomes assessment and regional accreditation, graduate programs should have some form of the “essential” assessment plan in place, gathering **direct evidence** tied to each learning outcome. As a part of program evaluation, most programs will want to also track *other* forms of evidence as well.

#### *Publications and Presentations*

Many programs track peer reviewed publications, funded grants or fellowships that require peer review, or the results of certification or licensure examinations. Publishing a peer-reviewed article demonstrates proficiency with some combination of writing, research, and other scholarly skills; however, by itself, publication statistics don’t allow faculty to evaluate specific strengths and weaknesses in a systematic way, and for this reason should be paired with other direct measures that allow for specific outcomes to be assessed.

Most programs also track the number of students presenting at conferences or at department events, and have established outcomes that require participation in these activities. These are *indirect* measures of learning since presenting at a conference by itself does not necessarily demonstrate proficiency in professional communication skills or mastery of disciplinary discourse. It is appropriate to include presentation and publication evidence as part of a program assessment plan, **but it should also be paired with other direct forms of evidence such as faculty evaluations of student work or presentations within the program** (e.g., the proposal, dissertation or thesis, course term papers, or the oral defense), which allow for a more systematic examination of program strengths and weaknesses.

#### *Program Review Evidence: Other Evidence for Graduate Programs*

For program review, and as part of a continuous process of improvement, most graduate programs will collect indirect and supporting evidence as well. The program may have goals aimed at job placement, career satisfaction, obtaining internships or fellowships, or goals for other kinds of student success and accomplishment. Relevant data for these kinds of goals can be collected by programs using things like exit interviews or exit surveys, or even alumni follow-up interviews or surveys for students that have been out of the program for a number of years. Programs should strongly consider tracking career outcomes and job satisfaction for their graduates, particularly in fields where graduates have a strong chance of going onto public

sector or non-tenure track jobs (Aanerud, Homer, Nerad, & Cerny, 2006). Student satisfaction with the program is also useful to track, particularly when broken down in to data points such as satisfaction with career advising, help with publishing, learning how to write grant proposals, or other salient factors. Additionally, surveys, focus groups, or self-evaluations can also be used to gather evidence of student learning. However, because these kinds of evidence are indirect, they don't stand on their own as evidence of student learning in a graduate program. They are, however, powerful secondary evidence that can supplement direct evidence and provide faculty with useful information for improvement (Funk & Klomparens, 2006, p. 155).

Kelly Funk and Karen Klomparens, in "Using the Assessment Process to Improve Doctoral Programs" (2006) suggest that graduate programs should also collect evidence from coursework throughout a graduate student's time in the program in order to track their "level of achievement along the continuum of their studies, not solely at the end" (p. 153). Peggy Maki (2009) takes this further, arguing that graduate program assessment should be anchored in "intellectual curiosity" about how graduate students develop and learn *over the course* of a program. Maki asks, "how do students become acculturated to the ways of thinking, knowing, and problem solving that your advanced degree values?" She also asks programs to consider what evidence along the way shows this process of acculturation, and what pedagogies best foster the development of expertise in the field. These questions seek to use the assessment process as part of a scholarship of teaching and learning, aimed at improving program efficacy, and suggest that graduate programs can gather and use a range of evidence from throughout graduate students' entire course of study in order to better understand (and improve) the nature and trajectory of learning and development in their program.

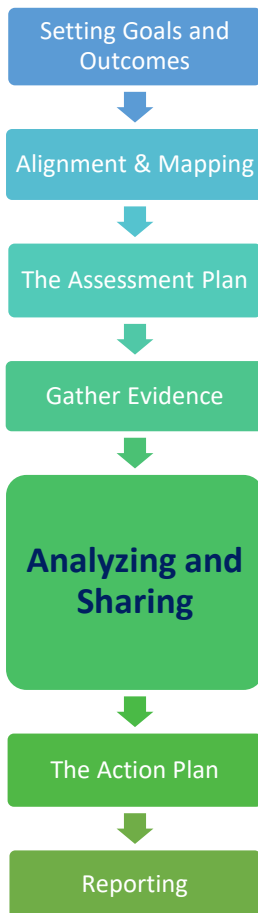
## Finalizing the Assessment Plan

Once decisions have been made about what data to collect and how to do it, the assessment plan should be **formalized in a document that outlines:**

- Program Goals and Learning Outcomes
- What, where, when, and how evidence of learning will be collected
  - Which student artifacts/performances will be assessed?
  - How will they be assessed? (rubric, exam score)
  - When will these outcomes be assessed? Annually? On a rotation?
  - In addition to the ongoing assessment, will there be any special assessment projects that gather different or more detailed data than the ongoing process does?
- Benchmarks and targets (e.g., 90% of students will score a 3 or higher) for each outcome
  - Benchmarks and targets for essential outcomes

- Benchmarks and targets for aspirational outcomes
- Minimum vs. Exemplary targets
- Department or program processes for gathering data, analyzing it, sharing it, and using it to improve learning
  - Who is responsible for gathering data at the point of assessment? (individual faculty?)
  - Who is responsible for aggregating and analyzing the data, and writing up or presenting the findings?
  - What are the processes for sharing findings and making a formal action plan?
  - What is the process for evaluating and revising the assessment plan itself?

The assessment plan should be vetted and approved by the department. However, programs are also **encouraged to seek peer review and feedback of their assessment plan**, which can be facilitated by the assessment coordinator at [andrewsa@uwm.edu](mailto:andrewsa@uwm.edu).



## Part 4: Analyzing & Sharing Results

Once you have gathered evidence of learning, the next step is to make sense of it. While all program faculty will need to participate in interpreting data and deciding on an appropriate response, it's likely that someone will need to take point on making sense of the data, especially if statistical analysis is necessary. In some cases, that process will be fairly straightforward, and in others it may require some careful consideration.

### Analyzing Assessment Data

Data analysis can be exciting, especially if a well-designed assessment process has produced usable data that speaks to something faculty care about. Perhaps you finally understand why students seem to drop and retake Course D at an especially worrying rate, or you've uncovered evidence that students who take elective Course J tend to do better in the capstone course than those who don't.

Things to keep in mind when analyzing data:

- Present data in relation to goals and outcomes
- Select appropriate procedures for data analysis
- Use both quantitative and qualitative forms of analysis where possible
- Consider the original assessment questions your data was meant to illuminate
- Consider the needs of your audience(s) and stakeholders
- Consider possible recommendations arising out of your assessment data

Analysis might reveal:

- Strengths and weaknesses of your program
- Overall strengths and weaknesses of students in the program
- Whether and to what extent students are developing competency or mastery
- Areas of possible improvement, even where performance is acceptable
- Likely causes of issues with student performance
- New questions for future assessment projects to pursue

- Evidence about job placement, graduate school admittance, or student satisfaction with the program
- Shortcomings with the assessment plan itself that need to be addressed

Partially adapted from UMass Amherst's *Program-Based Review and Assessment* (2001)

Using the right tools can help make data analysis go a lot more smoothly. For the most basic assessment plans, Excel may offer enough functionality to tabulate data and calculate percentages. For more complex assessment projects, tools such as NVivo and SPSS can save time and work. UWM has a variety of software options that can simplify data analysis, detailed below.

### UWM Tools for Analyzing Data:

- ATLAS.ti (coding audio and video)
- Excel (basic statistics)
- JMP (statistics and graphics)
- Minitab (statistics)
- NVivo (coding survey, focus group, or interview responses)
- Qualtrics (built-in functions for basic analysis of survey data)
- SAS (statistics)
- SPSS (statistics)
- STATA (statistics)

For more information about these software options, see

<https://uwm.edu/software/softwarelist/>

### Analyzing Qualitative Data

Qualitative data analysis is often a recursive process, but not one that needs to be overly challenging or complex in program assessment. It often starts with going over your data (survey responses, focus group comments) to look for patterns, themes, ideas that get repeated. The analyst can then categorize or code the responses and count the number of times something is mentioned or the number of respondents who mentioned it.

### Using Qualitative Data:

A program administers a survey to students taking a required introductory course. Students are asked to rate how useful they perceive the class to be. While most find the course useful or very useful, about 25% give a neutral or negative response. Analysis of their *comments* reveals a pattern of complaints that the course covers material they learned in high school. The placement process may not be working well for all students, or the course may need to include more challenging options for some students.

**Simple or Basic Analysis:** Taking notes at a focus group with a colleague, and then going over and writing up the major themes in your notes into a report can be an easy and quick way to turn focus group data into something useable. Open-ended survey questions can be similarly read over by a group of faculty who then discuss (and write up) what they see as the major themes in the responses.

**More Advanced Analysis:** Larger data sets tend to demand more systematic or careful analysis. Survey responses or focus group transcripts can be coded using software like NVivo, and then used to better understand patterns or to explain other data.

### Analyzing Quantitative Data

Most forms of direct evidence will require some form of quantitative analysis in order to be useful. Numbers only make sense when compared to other numbers. This section does not present a complete or expert view on statistics, but only summarizes the most commonly used procedures for analyzing assessment data.

The most basic level of analysis is description. Basic descriptive statistics can include measures of central tendency (mean, median, or mode), standard deviation, or percentages. To make sense of this data, however, we need something to compare it to. **Benchmarks and targets** provide a point of comparison that allows programs to make use of their data without using other statistical methods. They allow for conclusions to be drawn about how well the program is doing, although by themselves, they may not be able to illuminate more complex questions about causes or differences within the program. For programs using a “basic” assessment plan, this may be the only analysis needed. However, particularly with larger data sets or when a program wishes to understand more complex issues, other kinds of statistics may be useful.

If a program notes a difference in scores among student groups (or between students using a new curriculum and an old one), **it’s important to find out if that difference is meaningful**. For interval-level data, such as exam scores, significance can be established using a t-test to compare two groups, or a one-way Anova to compare the means of three or more groups. For ordinal or ranked data, the Mann-Whitney (to compare ordinal data from two groups) and Kruskal-Wallis tests (to compare ordinal data from three or more groups) can be used. For comparing repeated ordinal measures (e.g., the same students scored at the start of the semester and again at the end of the semester) the Wilcoxon T test works for comparing two sets of scores. The Friedman test can be used for three or more repeated measures of the same sample group.

<b>Statistics:</b>	<b>Interval/Scale data</b>	<b>Ordinal/ranked data</b>	<b>Nominal/Categorical</b>
<b>Descriptives</b>	Mean, Standard Deviation	Median, mode, percent distribution	Frequencies, contingency tables
<b>Difference</b>	T-Test, ANOVA	Mann-Whitney, Kruskal-Wallis	Fisher's exact test, Chi-squared test
<b>Pre-Post, Repeated Measures</b>	Dependent sample T-Test, repeated measures ANOVA	Wilcoxon T test, Friedman test	McNemar test, Cochran's Q

### Common Issues in Analyzing Data

As exciting as data analysis can be, it can also be frustrating or misleading without some careful thinking, and sometimes outside perspectives. Assessment professionals are often optimistic about the immediate value and usefulness of assessment data. On the whole that optimism is borne out in experience: faculty observe data, draw conclusions, decide how to respond to that data, and then in repeating the process often discover if they were wrong about the implications of their data the first time. If they were right, they've made an improvement; and if they were wrong, they now know to search for other explanations and adjust accordingly. The pitfalls of that process can be more easily avoided with an awareness of where it can go wrong.

### *Misusing Data*

The first consideration is how your use of the data aligns with the original intention. UMass Amherst's program assessment handbook warns, "Data are misleading, and even threatening, when they are used for purposes other than originally intended and agreed upon" (Stassen, Doherty, & Poe, 2001, p. 50). In particular, the use of assessment data to evaluate individual faculty is problematic for a number of reasons. Faculty, TAs, and staff will not want to participate in assessment exercises if they feel threatened by how that data will be used. Or they may feel pressured to produce assessment data that inflates the view of what students are learning under their care. Using assessment data in this way may also fundamentally misrepresent the actual causes of the data. For example, data gathered from a capstone course will reflect strengths and gaps in student learning from across the *whole degree program*; it does *not* directly reflect on the quality of the capstone teacher's instruction. Even in introductory courses, student performance might be low because students are entering the course with inadequate preparation.

Assessment data should be structured and gathered in a way that makes clear that what is being assessed is student learning in the program as a *whole*, with the goal of improving teaching and learning. The integrity of the process is better served when data is used to guide improvement and not for evaluating individual instructors.

### *Choosing Appropriate Statistics*

Assessment statistics for some kinds of data are fairly straightforward. Using the mean from exam data as a way to understand overall student performance is likely an uncontroversial choice. But what if the data was produced by a rubric that rates student performance on a scale of 0-3 (0=Unsatisfactory, 1=Developing, 2=Competence, 3=Mastery)? Or a survey with Likert scale data?

Most statistics experts argue that it is inappropriate to take the mean of ordinal data. They suggest using the mode or median as a measure of central tendency instead and using non-parametric tests for this kind of data. For most, this is an open and shut case because of their discipline, view of measurement theory, or intended audience(s). Others have argued that this is too restrictive (e.g., Velleman & Wilkinson, 1993; Gaito, 1980; Lord, 1953), and in educational research, psychology, and other disciplines ordinal data (e.g., Likert scale data) is sometimes subjected to parametric procedures such as t-tests or ANOVA when certain conditions are met (Norman, 2010; de Winter & Dodou, 2010; Sullivan & Artino Jr., 2013). It is up to individual analysts or programs to decide how they want to work with the data they have. If the analyst understands this debate, they can make their own informed choices about how to deal with ordinal data from rubrics or surveys; if not, calculating the *percent distribution* is often enough for basic analysis.

Some tips:

- If in doubt about your ordinal data, go with percentages.
- If you calculate an average, be careful about what it means. The median may be a better indicator of central tendency, and in any case the percent distribution often tells us more (i.e., the % rated 1, 2, 3, 4, or 5). If the average is 3.0 from a scale of 1-5, does that suggest that there were a lot of 3s, or does it mean that there were a lot of 0s and 5s? Middling performance is very different than polarized data, which may mean the course is working well for some students and not at all for others.
- Avoid drawing interval level conclusions from ordinal data. Even if the difference between two groups is statistically significant, it doesn't make sense to say Group A is "2.40 more satisfied" with the course than Group B. But we can claim that Group A is more satisfied than Group B, and if we examine the numbers and percent distribution, we find that Group A students were 2x more likely to rate themselves as "Satisfied" or "Very Satisfied" than Group B students.
- The goal is *useful* data that can help a program make informed decisions--not unassailable data that will convince skeptics.



### *Common Sense Interpretations*

There is also some danger in “common sense” interpretations of assessment data. David Eubanks, in “A Guide for the Perplexed” (2017), offers the example of a foreign language program that notices that 16% of students are failing to show language proficiency. The root problem seems to be a lack of grammar and vocabulary knowledge, so the program decides to double-down on its efforts to reinforce grammar and vocabulary, assuming that the issue is a lack of sufficient instruction. This choice, however, as Eubanks points out, robs the majority of students of instruction time devoted to something more useful for them.

Eubanks reminds us that students come into courses with varying student traits and experiences, which may also account for this problem. A regression analysis examining high school GPA and how long students waited to take the courses shows that these scores are more likely the result of students who *waited* two or three years to enroll in language courses, during which time they forgot much of what they had learned in high school language courses. The better solution, Eubanks suggests, is a change to advising that ensures students enroll in language courses in their first year.

Eubanks’ example is about making a faulty causal inference--the obvious assumption, that a lack of instruction was to blame, didn’t account for other variables that could be causing students to fail proficiency.

### *Insufficient Data*

A second issue that Eubanks points to is issues with data. Small sample sizes and insufficient context can lead to problematic data.

- Small samples can fail to be representative or show meaningful patterns (where faculty may read meaning into data that isn’t actually significant), or they can fail to detect meaningful differences even when they *are* present. For small sample sizes, data can be collected over a span of years. But even if that isn’t an ideal solution, another approach is to supplement data collection with indirect and qualitative evidence that allows faculty to effectively evaluate their programs even with small numbers of students.
- Too often, programs may collect decontextualized data, where information about student preparation, identities, or other salient factors prevents them from making meaningful sense of the information they’ve collected. The more contextual data programs can collect in relation to assessment data, the more powerfully they can interpret the significance of their data. **Disaggregating data** by relevant groupings is generally helpful for making good sense of data, but it is also an important part of supporting **equity** by identifying and closing achievement gaps.

## Peer Feedback

These issues are not insurmountable, even for faculty who lack the necessary expertise. The best solution to all of these potential issues is to **seek out peer review of your assessment plan, findings, and interpretations**. Many colleges and universities have assessment committees whose sole function is to advise and review program assessment plans, data, and results, which can help programs to avoid these pitfalls. While UWM does not have a formal assessment committee, UWM does have many faculty who *are* experts in research design, statistics, or assessment, who can guide programs in the right direction and offer guidance on how to make sense of assessment data. **To get connected to faculty and staff that can help you make an assessment plan, or analyze and interpret assessment data, contact the assessment coordinator in Academic Affairs at [andrewsa@uwm.edu](mailto:andrewsa@uwm.edu).**

## Sharing Data

Assessment data needs to be shared with several different audiences. **The most important audience is your own program or department.** Faculty need to be able to understand the data, consider it, and decide on an appropriate response. They won't be able to do that with raw data, so it's important that someone takes the lead on presenting assessment data in a way that helps faculty make sense of it and make decisions. Departments should schedule a formal meeting where assessment results will be discussed.

Not all audiences will need a formal report. Assessment data can be presented informally, or using PowerPoint, at faculty or staff meetings. Visualizations, such as bar graphs or charts, help audiences more quickly understand and process assessment data. Results included in a program review may need to be more formally presented in a report.

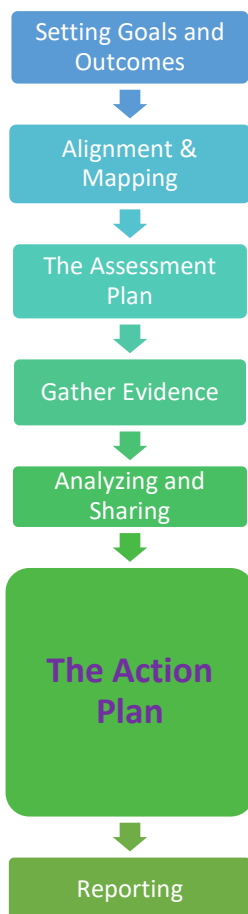
**Accreditation:** Departments and programs will share their assessment data for accreditation. They will report on goals, outcomes, measures and data, and any action plans resulting from their assessment. It's important to note that neither Academic Affairs nor the Higher Learning Commission is judging these materials in terms of a program's successes or failures: *what matters is evidence that departments have goals and outcomes, collect data about student learning, and use that data appropriately to improve student learning*. Department chairs or program coordinators can submit supporting documents, including rubrics, survey instruments, reports, meeting minutes (where results were discussed), or any other relevant documentation.

**Students and other Stakeholders:** The National Institute for Learning Outcomes Assessment's *Transparency Framework* encourages departments, programs, and institutions to make their learning outcomes, assessment processes, findings, and how they use evidence of student learning publicly available on their website. It encourages such information to be presented in

ways that are adapted to the intended audience(s), clearly worded, receptive to feedback, and adequately contextualized and explained to a lay audience. Such transparency can help make the case to students, donors, or other stakeholders about a program's commitment to student learning, to excellence, and about its successes. For more information, see <http://learningoutcomeassessment.org/TransparencyFramework.htm>.

**Academic Audiences Outside of UWM:** If you plan to publicly present (at a conference) or publish assessment data as part of a study, you should have your study reviewed by UWM's IRB before assessment data is collected. <https://uwm.edu/irb/>

## Part V: Using Results with an Action Plan



The most important and satisfying part of the program assessment process comes at the end: using assessment data to improve student learning. For the UWM accreditation process, improvement in student learning stems from the development of an “action plan,” a formalized statement of how faculty in the program are making use of their program assessment data to make changes, and how they plan to follow up to ensure that those changes are successful in improving student learning. That response can be to improve what isn’t working, and to preserve the things that are working as well as or better than expected.

The “action plan” is a *programmatic* response to a pattern, and not about responding to *individual* students who may have failed to achieve minimum benchmarks. While programs should have processes in place to help such students, the “action plan” is about how the program will respond to patterns revealed in the assessment data. Action plans are then included in each program's assessment archive as part of the assessment data and the accreditation process.

As programs implement their action plans, this effort becomes evidence that programs are *using* assessment data to inform decision-making. While accreditors are not judging programs by their successes or failures, the way that programs conduct assessment and *use* the results to making ongoing improvements is of particular interest in the accreditation process. In addition to including action plans as a part of assessment data, programs can also include other kinds of evidence, such as meeting minutes where assessment data was discussed, flyers or other documentation for resulting professional development work, or other evidence of the ways that a program has responded to their assessment findings. This enables the faculty to map their efforts and focus attention to the results, while also providing the necessary accreditation evidence.

### Action Plans: Making Changes to the Program

Once faculty have analyzed and interpreted their findings, they must decide what warrants a response, and how to respond. Even when assessment data suggests that things are working as they should, it might also show that they could still be working better than they are. Assessment allows programs to discover if change is needed, and it also provides a way to measure if change has been successful or produced the results that faculty intended.

Possible ways to use assessment data:

- To inform curricular changes
- To inform pedagogical changes or professional development initiatives
- To inform structural or policy changes for the program
- To improve advising
- To identify and address unmet student needs
- To improve or adjust the assessment plan itself
- To create a special assessment project to investigate a question raised by the data

Below are some examples of programs at UWM that have made successful programmatic changes as a result of assessment:

Using Assessment Data to Improve Programs:
<p><b>Art Education</b> The School of Education at UWM conducts exit surveys for students in all its programs. Among Art Education graduates, they noted a pattern of concern about being inadequately prepared to teach art to students with disabilities. The Art Education program responded by hiring a new faculty member with expertise in Exceptional Education, creating space in their courses for addressing the use of adaptive tools, and by creating a new course focused on using technology for instruction. Subsequent graduates reported feeling more prepared to work with students with disabilities.</p>
<p><b>Materials Engineering</b> In the Materials Engineering program PhD students must pass two days of qualifying exams. The first day is dedicated to demonstrating mastery of core concepts and questions, while the second day of the exam focuses on the student's area of specialization. Faculty compiled data on how well students did on each required topic for the first day of qualifying exams. They noted a pattern of low student performance on questions related to thermodynamics. In response, the program required all students to enroll in the thermodynamics course. Analysis of subsequent qualifying exams showed marked improvement in student performance.</p>
<p><b>Master of Nursing Program</b> The Master of Nursing (MN) program requires that 100% of students complete nursing care plans for two patients that demonstrate consistency with the American Nursing Association Standards. The program noted that out of 28 students, only half were able to submit a proficient care plan on their first attempt, while the other half required significant revisions and additional work. The teaching team addressed this by seeking a consultation with CETL, which resulted in both curricular and pedagogical changes. They added a care plan requirement to the next course in the program's sequence to reinforce the importance of the skill and ensure continued engagement with this skill. They also implemented a new technique called "clinical debriefing," where each week after clinical, students discussed a different American Nursing Association Standard of Care and how that standard applied to their patients of the week. This helped students to better "connect the dots" between the Standards and the students' care plans. The program also implemented a new protocol for guiding student discussion that promoted inclusion and equal participation for all students and trained their instructors to use the technique. Ultimately, care plan development is foundational for students to</p>

demonstrate use of critical thinking to care for patients with acute and chronic conditions across the lifespan.

#### First Year Composition

In the early 2010s most sections of English 101 were taught by new TAs in the English department. Each fall they were given a new assignment sequence and mentoring to help them teach the curriculum for the first time. The program collected assessment data annually by double and blind scoring portfolios on learning outcomes using a sampling method that allowed them to compare the results of students of first year TAs with those of more seasoned TAs and lecturers. This ongoing assessment process allowed the program to track when the TA training program was working well (when new TAs produced scores that were comparable with more experienced instructors, and when smaller standard deviations demonstrated the standardizing effect of the training program on the TAs and their students). It also allowed the program to track when performance on particular learning outcomes shifted as a result of curricular changes in the assignment sequence for the course. While the curriculum and structure of the program has changed over the years, assessment continues to function as part of an ongoing process of evaluation and improvement.

#### Action Plan: Revising the Assessment Plan

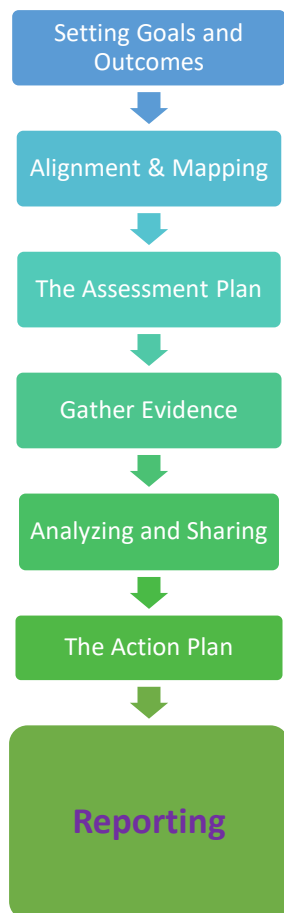
The purpose of assessment is to produce usable results that facilitate program improvement. When it is working, it is a powerful tool in faculty hands for evaluating, understanding, and improving the learning of their students and the effectiveness of their program. But when faculty argue that assessment isn't really helping them, it may stem from an assessment plan whose results actually *aren't* helping them. Rather than a failing, discovering this is one crucial use of the results of assessment: taking a hard, critical, look at your assessment plan to understand where it is serving its intended purpose, and where it may be failing.

It is possible, and desirable, to find from your assessment data that everything is working exactly as intended. It can be wonderful news to find that no changes are needed, and your well-designed curriculum is delivering exactly what it promises. However, it is also important to be wary of assessment plans that only ever deliver good news. If the data being gathered doesn't dig deep enough to reveal issues that may really be there, it can lull programs into a false sense of complacency or reinforce the notion that assessment work is a waste of time.

If faculty have designed a good assessment plan, addressed to the right questions and gathering evidence that allows patterns, strengths, and weaknesses to be discerned, then the results usually *will* demand a response at least some of the time. If your program's assessment work never reveals room for improvement, even over the course of several years, then it may be time to revisit the assessment plan itself to ask a different question, or to gather better evidence. In the long run, even a well-designed program and curriculum that has been running smoothly for years will eventually find that something is no longer working like it once did. With constantly changing student bodies, faculty lives, and shifting budgets, it's inevitable that at some point things in a program that once worked well may no longer be functioning as intended. If your assessment plan leaves these kinds of developments invisible, then the needed response to your assessment data is to **revise the assessment plan** itself.

### Results and the Purpose of Assessment

Faculty at universities across America have embraced assessment because they've seen first hand how it has helped them to improve their programs, their teaching, and outcomes for their students. It gives them a mechanism for monitoring, evaluating, and improving their programs, and for addressing questions about teaching and learning that matter to them. But assessment can only produce these kinds of benefits when faculty take control and make it work for them. Ultimately, the value of the assessment process is in producing results that faculty find valuable, and in using those findings to improve learning.



## Part VI: Reporting Assessment Work

UWM departments and programs are asked to submit documentation annually of their ongoing assessment work as part of the accreditation process. UWM uses a Qualtrics form for reporting and managing assessment data, stored in Sharepoint for each program. All faculty with responsibility for graduate and undergraduate program assessment, such as department chairs or program coordinators, should request access to their program's Sharepoint assessment archive from the assessment coordinator.

Assessment data for each academic year should be submitted annually, no later than Nov 1 of the following academic year. It will ask you to login with your ePanther ID and password to access the form. Data for each degree program should be submitted separately.

[Submit Program Assessment Data \(Qualtrics form\)](#)

For questions or assistance submitting data, please contact the Assessment Coordinator, Adam Andrews, at [andrewsa@uwm.edu](mailto:andrewsa@uwm.edu).



# Resources

If you have questions about program assessment at UWM, there are some individuals and offices on campus that can help:

Adam Andrews, *Assessment Coordinator*

Academic Affairs

Northwest Quadrant B, Room 2483

2025 E. Newport Ave.

Milwaukee, WI 53211

[andrewsa@uwm.edu](mailto:andrewsa@uwm.edu)

Devarajan Venugopalan, *Vice Provost*

Academic Affairs

Chapman Hall, Room 240

2310 E. Hartford Avenue

Milwaukee, WI 53211

(414) 323-9790

[dv@uwm.edu](mailto:dv@uwm.edu)

Connie Schroeder, *Senior Consultant for Instructional and Organizational Development*

Center for Excellence in Teaching and Learning

Engelmann Hall, Room B50B

2033 E. Hartford Avenue

Milwaukee, WI 53211

(414) 229-5764

[connies@uwm.edu](mailto:connies@uwm.edu)

Office of Assessment & Institutional Research

Chapman Hall, Room 236

2310 E. Hartford Avenue

Milwaukee, WI 53211

414-229-3194

<https://uwm.edu/institutional-research/>

Center for Excellence in Teaching and Learning

Engelmann Hall, Room B50

2033 E. Hartford Avenue

Milwaukee, WI 53211

(414) 229-6638 (phone)

(414) 229-6758 (fax)

[cetl@uwm.edu](mailto:cetl@uwm.edu)

### Other Online Assessment Resources

Assessment at Marquette

<https://www.marquette.edu/assessment/>

Assessment Commons: Internet Resources for Higher Education Outcomes Assessment

<http://assessmentcommons.org/>

Association for the Assessment of Learning in Higher Education (AALHE)

<https://www.aalhe.org/>

Association of American Colleges and Universities (AAC&U)

<https://www.aacu.org/>

Higher Learning Commission

<https://www.hlcommission.org/>

National Institute for Learning Outcomes Assessment (NILOA)

<http://learningoutcomeassessment.org/>

National Survey of Student Engagement (NSSE)

<http://nsse.indiana.edu/index.cfm>

UIPUI: Assessment Institute

<http://assessmentinstitute.iupui.edu/>

UW-Madison: Student Learning Assessment

<https://assessment.provost.wisc.edu/>

## Bibliography

- Aanerud, R., Homer, L., Nerad, M., & Cerny, J. (2006). Paths and Perceptions: Assessing Doctoral Education Using Career Path Analysis. *The assessment of doctoral education: Emerging criteria and new models for improving outcomes*. Stylus Publishing, Sterling, VA, 109-141.
- Adelman, C. (2015). To Imagine a Verb: The Language and Syntax of Learning Outcomes Statements. Occasional Paper# 24. National Institute for Learning Outcomes Assessment.
- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., ... & Wittrock, M. C. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives, abridged edition. White Plains, NY: Longman.
- Anson, C. M., Dannels, D. P., Flash, P., & Gaffney, A. L. H. (2012). Big rubrics and weird genres: The futility of using generic assessment tools across diverse instructional contexts. *Journal of Writing Assessment*, 5(1), 16.
- Banta, T. W. (Ed.). (2002). *Building a scholarship of assessment*. John Wiley & Sons.
- Banta, T. W., & Blaich, C. (2010). Closing the assessment loop. *Change: The Magazine of Higher Learning*, 43(1), 22-27.
- Banta, T. W., Jones, E. A., & Black, K. E. (2009). *Designing effective assessment: Principles and profiles of good practice*. John Wiley & Sons.
- Brennan, R. L. (2006). Educational Measurement. ACE/Praeger Series on Higher Education. Praeger. Available from: Greenwood Publishing Group, Inc. PO Box 5926, Portsmouth, NH 03802-6926.
- Brooks, R. L., & Heiland, D. (2007). Accountability, assessment and doctoral education: Recommendations for moving forward. *European Journal of Education*, 42(3), 351-362.
- Cizek, G. J. (Ed.). (2012). *Setting performance standards: Foundations, methods, and innovations*. Routledge.
- de Winter, J.C.F. and D. Dodou (2010), Five-Point Likert Items: t test versus Mann-Whitney-Wilcoxon, *Practical Assessment, Research and Evaluation*, 15(11).
- Donohue, W. J. (2017). Students Talking About Assessment: Insights on Program Learning Outcomes.
- Eubanks, D. (2017). A guide for the perplexed. Intersection of Assessment and Learning.

- Ewell, P. T. (2009). Assessment, accountability, and improvement. *NILOA Occasional Paper*, (1).
- Funk, K., & Klomparens, K. L. (2006). Using the assessment process to improve doctoral programs. *The assessment of doctoral education: Emerging criteria and new models for improving outcomes*, 145-162.
- Golde, C. M., Jones, L., Bueschel, A. C., & Walker, G. E. (2006). The challenges of doctoral program assessment. *The assessment of doctoral education: Emerging criteria and new models for improving outcomes*. Stylus Publishing, Sterling, VA, 53-82.
- Huot, B. (2003). *Rearticulating writing assessment for teaching and learning*. University Press of Colorado.
- Hutchings, P., Huber, M. T., & Ciccone, A. (2011). *The scholarship of teaching and learning reconsidered: Institutional integration and impact* (Vol. 21). John Wiley & Sons.
- Jankowski, N. A. (2017). Unpacking relationships: Instruction and student outcomes. *American Council on Education*.
- Jankowski, N. A., & Marshall, D. W. (2017). *Degrees that matter: Moving higher education to a learning systems paradigm*. Stylus Publishing, LLC.
- Jankowski, N. A., Timmer, J. D., Kinzie, J., & Kuh, G. D. (2018). Assessment that matters: Trending toward practices that document authentic student learning. *Urbana, IL: National Institute for Learning Assessment*.
- Jankowski, Natasha A. "EMERGING DIALOGUES IN ASSESSMENT." *Association for the Assessment of Learning in Higher Education*, 2018, [http://www.aalhe.org/page/ed\\_2018\\_principles](http://www.aalhe.org/page/ed_2018_principles).
- Johnson, R. L., Penny, J., & Gordon, B. (2000). The relation between score resolution methods and interrater reliability: An empirical study of an analytic scoring rubric. *Applied Measurement in Education*, 13(2), 121-138.
- Kemmis, S., and McTaggart, R. (2000). Participatory Action Research. In N. Denzin and Y. Lincoln (Eds.) *Handbook of Qualitative Research* (2nd Ed.) (pp. 567-605). Beverley Hills CA: Sage.
- Kuh, G. D., Ikenberry, S. O., Jankowski, N. A., Cain, T. R., Hutchings, P., & Kinzie, J. (2014). *Using evidence of student learning to improve higher education*. John Wiley & Sons.
- Litterst, J. K., & Tompkins, P. (2000). Assessment as a Scholarship of Teaching.
- Lovitts, B. E. (2006). Making the implicit explicit: Creating performance expectations for the dissertation. *The assessment of doctoral education*:

- Emerging criteria and new models for improving outcomes. Stylus Publishing, Sterling, VA, 163-187.*
- Maki, P. (2009). Anchoring Assessment in Intellectual Curiosity about Graduate, Professional, and Doctoral Students' Chronological Learning. Presentation at the University of Rhode Island.
- Maki, P. L. (2012). *Assessing for learning: Building a sustainable commitment across the institution*. Stylus Publishing, LLC.. (Kindle edition)
- Maki, P. L., & Borkowski, N. A. (2006). *The Assessment of Doctoral Education: Emerging Criteria and New Models for Improving Outcomes*. Stylus Publishing, LLC. PO Box 605, Herndon, VA 20172-0605.
- Montenegro, E., & Jankowski, N. A. (2017). Equity and Assessment: Moving towards Culturally Responsive Assessment. Occasional Paper# 29. *National Institute for Learning Outcomes Assessment*.
- Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in health sciences education*, 15(5), 625-632.
- O'Neill, P., Moore, C., & Huot, B. (2009). *Guide to college writing assessment*. University Press of Colorado.
- O'Neill, P., & Murphy, S. (2012). Post-Secondary Writing Assessment. In Secolsky, C., & Denison, D. B. (Eds.). *Handbook on measurement, assessment, and evaluation in higher education*. Routledge (p586-604).
- Oakleaf, M. (2009). Using rubrics to assess information literacy: An examination of methodology and interrater reliability. *Journal of the American Society for Information Science and Technology*, 60(5), 969-983.
- Palomba, C. A., & Banta, T. W. (1999). *Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education*. Higher and Adult Education Series. Jossey-Bass, Inc., Publishers, 350 Sansome Street, San Francisco, CA 94104.
- Secolsky, C., & Denison, D. B. (Eds.). (2012). *Handbook on measurement, assessment, and evaluation in higher education*. Routledge.
- Stassen, Martha L.A., Doherty, Katheryn, and Poe, Mya. (2001) *PROGRAM-Based Review and Assessment: Tools and Techniques for Program Improvement*. University of Massachusetts Amherst.
- Stevens, S.S. (1951). Mathematics, measurement, and psychophysics. In S.S. Stevens (Ed.), *Handbook of experimental psychology*. New York: John Wiley.
- Sullivan, G. & Artino Jr., A. R. (2013). Analyzing and Interpreting Data From Likert-Type Scales. *Journal of Graduate Medical Education*. 5(4), pp. 541-542.

- Suskie, L. (2012) "Rubric Development." In Secolsky, C., & Denison, D. B. (Eds.). *Handbook on measurement, assessment, and evaluation in higher education*. Routledge(p545-7).
- Suskie, L. (2018). *Assessing student learning: A common sense guide*. 3<sup>rd</sup> edition. John Wiley & Sons.
- Velleman, P. F., & Wilkinson, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. *The American Statistician*, 47(1), 65-72.
- Volkein, J. F. (1996). *Program evaluation and assessment: What's the question?* Albany: State University of New York.
- Walker, G. E., Golde, C. M., Jones, L., Bueschel, A. C., & Hutchings, P. (2009). *The formation of scholars: Rethinking doctoral education for the twenty-first century* (Vol. 11). John Wiley & Sons.
- Walvoord, B.E. (2009). The Useful, Sensible, No-Frills Departmental Assessment Plan. In *Essays on Teaching Excellence: Toward the Best in the Academy*. The Professional and Organizational Development Network in Higher Education.
- Walvoord, B. E. (2010). *Assessment clear and simple: A practical guide for institutions, departments, and general education*. John Wiley & Sons.
- Walvoord, B.E. (2012). The Role of Faculty Grading Processes in Program and General-Education Assessment. In Secolsky, C., & Denison, D. B. (Eds.). *Handbook on measurement, assessment, and evaluation in higher education*. Routledge (p573-585).

# A Glossary of Assessment Terms

<b>Alignment</b>	The process of ensuring that assignments correspond with course outcomes, course outcomes with program goals & outcomes, and that program goals & outcomes correspond with institutional goals.
<b>Artifact</b>	A product produced by a student that demonstrates their mastery, proficiency, or competence in relation to a learning outcome.
<b>Assessment</b>	A systematic process of research and evidence gathering aimed at understanding and improving student learning in relation to institutional, program, and course goals and outcomes.
<b>Calibration</b>	Also called “norming,” a process by which a group of raters using a rubric agree on how to use, interpret, and apply the rubric’s performance levels to produce consistent scores by looking at a series of examples and discussing how to evaluate them until consensus is reached.
<b>Curriculum Map</b>	A visual representation of the course of study that shows in which courses each learning outcome is introduced or reinforced, and in which courses students will demonstrate mastery.
<b>Direct Evidence</b>	Direct evidence is gathered by evaluating student work (artifacts or performances) in light of learning outcomes. Direct evidence is usually quantitative but can also be qualitative (e.g, collected faculty comments evaluating student artifacts in relation to an outcome).
<b>Formative Assessment</b>	Assessment that gathers evidence of student learning from at least one point before the end of the course or program, with the aim of understanding and improving the learning process for students.
<b>Indirect Evidence</b>	Evidence of learning gathered by evaluating student perceptions of their learning or experience. Indirect evidence can include interviews, surveys, focus groups, self-reports, or student reflections. Indirect evidence can be quantitative (e.g., likert scale questions from a survey) or qualitative (written responses to open-ended questions, interviews).
<b>Institutional Assessment</b>	Institutional Assessment is the systematic process of research and evidence gathering aimed at understanding student learning and experience in relation to institutional goals, initiatives, and outcomes.
<b>Inter-rater Reliability</b>	The degree to which a group of raters produce the same or similar scores when evaluating the same or similar student work or performances.
<b>Intra-rater Reliability</b>	The degree to which the same rater or raters, using a rubric, produce the same or similar scores over time (i.e., if rating the same or similar artifacts).
<b>Learning Outcomes Assessment</b>	Assessment focused on measuring student <i>learning</i> in relation to program goals and outcomes. Focuses on improving <i>student learning</i> .
<b>Mission Statement</b>	Explains why the program or institution exists, what its goals and values are, and articulates principles governing how those goals are achieved.
<b>Program Assessment</b>	The systematic process of research and evidence gathering aimed at understanding student learning and experience in relation to a program’s goals,

	initiatives, and outcomes. Includes both learning outcomes assessment and other program evaluation measures.
<b>Program Evaluation</b>	The systematic process of research and evidence gathering aimed at understanding, evaluating, and improving the functioning of the program in relation to its mission, goals, and outcomes. Focuses on <i>program improvement</i> .
<b>Program Goals</b>	Goals are general, broad, sometimes aspirational statements about the knowledge, skills, habits of mind, or values that graduating students from a program should possess. Goals should correspond to and help fulfill the program's stated mission.
<b>Reliability</b>	Describes how well an assessment method or tool provides consistent and accurate results.
<b>Rubric</b>	A form that guides faculty in scoring or evaluating student work or performances in relation to learning outcomes. Rubrics help make expectations clear to students, contribute to consistency and fairness in evaluation, and facilitate gathering program assessment data.
<b>Strategic Goals</b>	Goals articulated by a program or institution which are not about learning per se, but about maintaining or improving the program's ability to support student learning and fulfill its mission.
<b>Student Learning Outcomes (SLOs)</b>	SLOs are specific statements of what students will do to demonstrate their mastery of program goals. SLOs can be articulated at any level, as Institutional Learning Outcomes (ILOs), Program Learning Outcomes (PLOs), and as Course Learning Outcomes (CLOs). Goals may have multiple outcomes associated with them.
<b>Summative Assessment</b>	Assessment that gathers evidence from the end of a student's studies, to evaluate how well students graduating from a program are achieving its stated goals and outcomes.
<b>Supporting Evidence</b>	Evidence that provides context for other assessment data, or which speaks to strategic program goals. Supporting evidence can also help programs identify areas of concern. Examples include pass/fail rates, course grades, grade distributions, course syllabi, or other kinds of institutional data.
<b>Validity</b>	The degree to which an assessment method or tool actually measures the intended learning outcome.
<b>Vision Statement</b>	A vision statement explains how a program or institution plans to evolve, or how it is changing to better achieve its goals and fulfill its mission. Vision statements are future oriented.