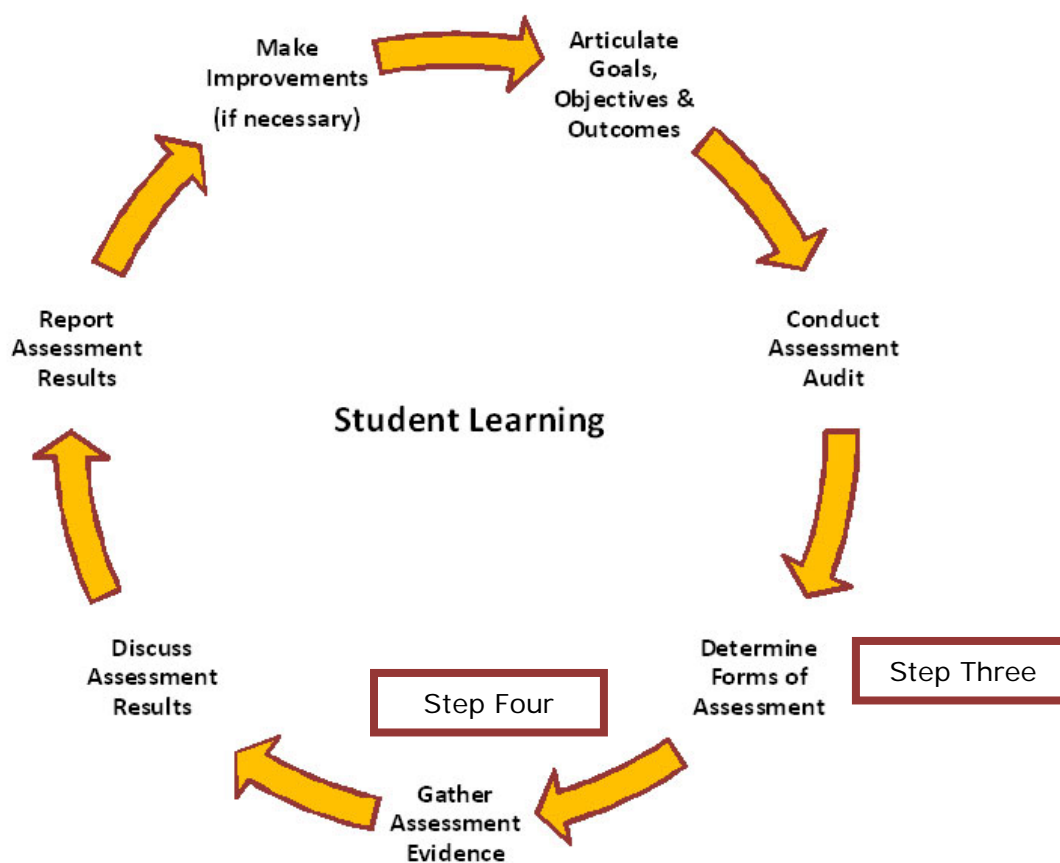# Outcomes Assessment Essentials: Test Blueprinting, A Course-Embedded Tool

No. 4
©January 2012

Course-embedded tools provide direct measures of student learning; the variety of tools is discussed in more detail in *Outcomes Assessment Essentials No. 2: Types of Outcomes Assessment Measures*. One course-embedded tool is test blueprinting. It enables faculty to map results of test items back to student learning objectives (SLO) identified for program-level or general education outcomes assessment. Deciding upon and using test blueprinting as one of your program's assessment tools are steps three and four of the assessment process.

Test Blueprint Defined

1) Linda Suskie defines a test blueprint as "an outline of the test that lists the learning goals that students are to demonstrate." The results reveal if you are testing what you claim is "most important."[1] **2)** Barbara Walvoord and Virginia Anderson define test blueprinting as "the process of linking tests to learning goals."[2]

---

[1] Linda Suskie, *Assessing Student Learning: A Common Sense Guide*, 2nd ed (San Francisco: Jossey-Bass, 2009), 167.
[2] Barbara E. Walvoord and Virginia Johnson Anderson, *Effective Grading: A Tool for Learning and Assessment in College*, 2nd ed (San Francisco: Jossey-Bass, 2010), 13.

ALE _____ http://www.bloomu.edu/tale

Look closely at the above definitions: the authors focus attention on student learning, not outcomes assessment *per se.* In short, test blueprinting represents best practice when you are writing objective exams. Indeed, you may already track or map exam items to ensure they will effectively and comprehensively test course content. What is more, you may also track the cognitive challenges for individual test items (e.g. demonstration of knowledge compared to the ability to analyze). So if you are already test blueprinting, making this data available for outcomes assessment becomes so much easier.

> ## Test blueprinting - It's not just for outcomes assessment…

Advantages to Test Blueprinting. If you do not systematically blueprint a test before administering, consider the advantages in doing so for you and your students. Linda Suskie highlights several reasons:[3]

- Affords an opportunity to reflect upon whether or not your test is measuring course learning goals.
- Ensures that you have written or selected test items that "give appropriate emphasis to thinking skills." As Linda Suskie notes, "Faculty writing test questions without a blueprint often find that questions asking for simple conceptual knowledge are easier to write than those asking students to interpret, infer, analyze, or think in other ways." Subsequently, you may have intended to test higher order thinking, but your questions over-emphasize the ability to recall correct facts or recognize correct answers.
- Facilitates reporting outcomes assessment results.

One additional advantage to test blue-printing that Middle States Commission on Higher Education points out: "After a student takes an examination, she can go back to the blueprint and reflect on her grade relative to the concepts she was expected to have mastered." In short, if you share the blueprint with students, test grades become more "meaningful."[4] Who knows, maybe students will be encouraged to focus as much attention on **learning** as earning grades.

A **test blueprint sample** follows these instructions and a blank template (word document) will be available on the TALE website if you find this visual a useful way to keep track of the steps.

Steps for Test Blueprinting

1. List the student learning objectives (SLOs) that the test will cover. For inspiration, consult your syllabus, lecture notes, readings, and student learning objectives listed in master syllabi, program goals, etc. For example, a course learning objective for Geomorphology (Geoscience 265) requires students "Interpret landforms to determine geomorphic processes acting on a landscape." In some cases, especially in the humanities and social sciences, learning objectives may be more broadly phrased. So when test blueprinting an individual exam, more specific concepts, themes, or topics should be listed with the SLO. For example, a course learning objective for Western Civilizations II (History 126) requires "Students will interpret historical evidence (from secondary and/or primary sources) regarding the regions and periods covered by this course." If the exam is focused on twentieth-century developments, you might list such topics as origins of the First World War, the

---

[3]Suskie, *Assessing Student Learning*, 167.

[4] Middle States Commission on Higher Education, *Student Learning Assessment: Options and Resources*, 2nd ed., (Philadelphia, 2007), 73. http://www.msche.org/publications/SLA_Book_0808080728085320.pdf

legacy of European imperialism, and the evolution of the European Union. (See **column 1** below.)

2. Decide what percentage of the entire exam will be dedicated to measuring each objective. The percentages assigned should reflect their importance to the learning objectives. (See **column 2** below.)

3. Determine what types of test items will most effectively measure the student learning objectives (multiple-choice, true-false, matching, fill-in-the-blank, ordering, Likert scale, etc). Be sure to consider whether the type of test item will create an appropriate cognitive challenge. For example, in most contexts, matching items will only assess student knowledge or ability to recall; it is unlikely to test their ability to apply. See *Outcomes Assessment Essentials No. 1* for more information about articulating learning goals and matching them to cognitive domains such as Bloom's taxonomy. (See **column 3** below.)

4. Determine how many test items you want to include in order to measure the learning objectives. Make certain that the number of items is proportional to the importance of the learning objectives. (See **column 4** below.)

5. Then, determine how each test item will be weighted. For example, you may use both multiple-choice and fill-in-the-blank to assess students' ability to interpret topographical maps, but one type may be more challenging than another. Will they be assigned the same value?

6. Write the test questions or select from a test bank and indicate the **cognitive level**. (See **column 5** below and the text box "Tips for Recognizing Cognitive Levels in Test Items.")

7. List the question numbers that will measure each SLO. Some test items may measure more than one SLO. (See **column 6** below.)

8. Before finalizing the test, Linda Suskie recommends that you review the exam with the following questions in mind:[5]

> • "Do the items follow the test blueprint?"
> • "Are the formats and content appropriate for the learning goals you

---

Tips for Recognizing Cognitive Levels in Test Items

If you are trying to determine the cognitive level of test items, consider the following two approaches.

First, compare the verbs you used to write the test items or test directions and locate them in the taxonomies for cognitive, affective, or psychomotor domains (e.g. Bloom's Taxonomy). Yet be careful not to reach hasty conclusions because judging the cognitive demands that you put on students is also influenced by your teaching and learning activities. For example, if you ask students to analyze a case study, in theory you will be asking them to engage in higher order thinking. However, if the case study on the exam is very similar to or identical to an example brought up during class instruction, students may only be engaging in recall.

Second, Barbara Walvoord and Virginia Anderson offer the following hints:
"**Higher critical thinking**. Questions would fall in the analysis, synthesis, or evaluation levels of Bloom's taxonomy. Course materials give needed background to answer questions. There is no directly visible connection between the course material and the test question."
"**Lower critical thinking**. Questions would fall in the application level of Bloom's taxonomy. Course materials give needed background to answer the questions. There is a directly visible connection between the course material and the test question."
"**Knowledge and comprehension levels of Bloom's taxonomy**. Material is directly from the course presentation, with some changes in wording and phraseology."
(Source: *Effective Grading*, 2nd ed., 55)

For more information about writing test question, consult TALE's Teaching Resources → Writing Test Questions

---

[5] Suskie, *Assessing Student Learning*, 179-180.

are assessing?"
- "Are the items at an appropriate reading level?"
- "Are the items of appropriate difficulty?"

Test Blueprint Sample
(this sample is available in template form on the TALE Website)

1	2	3	4	5	6	7
Student Learning Objectives (list SLOs tested on individual exam; if broadly phrased, include topics, themes, concepts)	Percentage of Test Dedicated to Measuring the SLO (column must total 100%)	Type(s) of Test Items to Measure the SLO (multiple-choice, true-false, ordering, fill-in-the-blank, short answer, Likert scale, etc.)	Number of Questions for Each Item Type & Item Weight (item weight here is expressed in point values)	Indicate Level of Cognition (e.g. knowledge, comprehension, application, analysis, synthesis, evaluation)	List Question Numbers for Each SLO (some test items may measure more than one SLO)	Percentage of students who answered correctly (use to reflect upon learning, teaching and outcomes assessment reporting)
SLO #1	10%	Multiple-Choice	10 (1 point each)	Knowledge	1-10	84%
SLO #2	10%	Matching	10 (1 point each)	Knowledge	11-20	86%
SLO #3	10%	Ordering	5 (2 points each)	Comprehension	21-25	82%
SLO #4	50%	Multiple-choice (interpreting graphs & maps)	5 (10 points each)	Analysis	26-30, 32, 34	80%
SLO #5	20%	Likert Scale: Strongly agree-Strongly disagree (in response to scenarios)	4 (5 points each)	Analysis	31-34, 27, 29	70%

❓ So how is test blueprinting a tool for Outcomes Assessment?

When the results of individual test items are linked to student learning objectives in outcomes assessment plans, then they are a course-embedded, direct measure of student learning. (See **column 7** above.) For example, if SLO #4 was either listed as a goal for general education credits or a program goal currently being assessed, then student scores on those individual test items provides evidence of learning. When reporting, Linda Suskie recommends that these items be aggregated "into an overall sub-score."[6] What is more, programs and departments do not report the results of all test items, only those which they have identified in their outcomes assessment plans or are listed as general education goals. For additional information on summarizing test results, consult Linda Suskie's *Assessing for Student Learning*, 2nd ed., 260-271.

Keep in mind: these results may not be the only department evidence to document student learning of SLO #4. It may also be measured through a capstone project, internship experience, or even with indirect measures. Subsequently, these results would provide corroborating evidence of student learning.

One last thought: since the goal of doing assessment is to improve learning, the aggregate results should be discussed in departments. If students scored poorly on a set of exam items, faculty should contemplate and explore the reasons and discuss solutions over which they have control. From the faculty perspective, using the results as an opportunity to reflect upon our teaching and student learning can provide sufficient intrinsic motivation even if we believe the results will disappear into a bureacratic black hole.

Below is the most direct and simple model of how course-embedded outcomes assessment begins with the individual faculty member. The two direct measures of student learning referenced in this model are test blueprinting and rubrics, which are the most commonly used course-embedded assessment tools.

The Most Direct and Simple Model of Course-Embedded Outcomes Assessment[7]

Instructor:
- By the beginning of each semester, the instructor identifies assignments, projects, tests, etc that fulfill course student learning objectives & will be used by him/her for department course-embedded assessment.
- The instructor provides this information to the department at the beginning of the semester and indicates if s/he is using rubrics and/or test blueprinting.

⬇

Instructor:
- The instructor develops the rubric criteria for projects, presentations, essays, etc and/or develops a test blueprint in the case of objective exams.
- If the course has multiple sections or is taught by multiple faculty, colleagues may want to work as a team to develop rubrics or blueprint tests. However, they do not have to use the same rubrics nor do they have to give the same tests. Each individual instructor identifies which test items and assignments will contribute to the department's assessment of student learning objectives (SLOs).
- If the course has received approval for general education points, the department, according to the General Education Guidelines, will have already developed a rubric to report results for outcomes assessment (i.e. the VALUE rubrics).

---

[6] Suskie, *Assessing Student Learning*, 260.
[7] Adapted from Walvoord, *Assessment: Clear and Simple*, 13-18.

Instructor:

- The instructor evaluates individual assignments during the semester and submits final grades to the registrar.
- The instructor reports assignment grades linked to criteria-based rubrics and test blueprints to the department for program-level and general education outcomes assessment.
- The instructor uses the opportunity to reflect upon what the grades reveal about student learning and teaching effectiveness, and when relevant revises assignments, methodology, sequence of material, pedagogy, etc.

Instructor:

- In the case of multiple section offerings, faculty may use their own rubrics to grade essays, projects, etc and have their own test items, but they will need to "translate" their results into the tools the department has chosen for reporting outcomes.

Example: Translating Course Results for your Department

In History 126 (Western Civilizations II), one of the course SLOs for which the department is seeking general education credits: "Students will identify and distinguish between important historical developments and movements, events, and actors relevant to the regions and periods covered by this course." This learning objective is linked to the VALUE rubric Element Existing Knowledge (Inquiry and Analysis). Some history faculty will decide to assess this SLO by administering objective exams, others will ask students to write essays.

How can individual faculty "translate" their results into the rubric which the department has created for reporting outcomes?

- If the instructor chooses to measure the SLO by objective exams, then s/he would test blueprint. The test blueprint, which the faculty member would submit to the department, links test items to student learning objectives. Then, s/he would report on the percentage of students who got those test items correct as aggregate sub-scores.
- If the SLO was measured through an essay exam, then the faculty member would provide the rubric that s/he used to evaluate student work and report the percentage of students who reached capstone (4), milestones (3 & 2), benchmark (1), or failed (0). These are the scales provided by the VALUE rubrics.

- **An extra level of complexity and time may be necessary.**
    - If your department is evaluating student learning through essays, presentations, portfolios, research papers, capstone projects, etc., they need to consider the following question: **Do all faculty in the department agree upon what A, B, C, D-level work looks like?**
    - To ensure "inter-rater reliability … [that is] the probability that several faculty raters will score a particular student's work in the same way," faculty may want to consider the following approach:
        1. Generate a set of anonymous samples in order to discuss grading standards. Be sure to remove student names and identity of course instructors.
        2. Faculty who participate in the sampling should be teaching the course(s), though not necessarily in that semester, and additional department faculty may want to participate.
        3. Have each faculty member rate the samples using a common rubric, then gather and discuss the results. Have you arrived at considerable agreement? If not, use this as an opportunity to discuss the differences. The goal is to get everyone "on standard," that is reach inter-rater reliability.
        4. Ensuring inter-rater reliability does add time to the process and some faculty may see this as an unnecessary burden but the results may stimulate valuable, intellectual discussion within your department about the standards of your profession. Still, course-embedded assessment can lose some of its appeal because scoring exceeds normal grading time.

Department:

- The department reports results annually through whatever means and format the university has developed (e.g. TracDat). For example, this particular SLO listed below is one of the History Department's general education goals which will be assessed. We would then report aggregate scores in the forms of percentages for each level of the scale.

Related Value Rubric Element(s)	SLOs for Hist. 126	Capstone 4	Milestone 3	Milestone 2	Benchmark 1	0
Existing Knowledge (Inquiry and Analysis)	Students will identify and distinguish between important historical developments and movements, events, and actors relevant to the regions and periods covered by this course.	Identifies key historical characteristics regarding relevant people, events, movements, or developments in a consistently accurate manner.	Identifies key historical characteristics regarding relevant people, events, movements, or developments in a predominantly accurate manner.	Identifies key historical characteristics regarding relevant people, events, movements, or developments in a manner that is more often accurate than not.	Identifies key historical characteristics regarding relevant people, events, movements, or developments with minimal accuracy	

- The department discusses what the results reveal about student learning, teaching effectiveness, and curriculum and contemplate making revisions, if deemed necessary or within its control, depending upon their assessment.

Useful Web Sources:

- See chapter 3, "Evaluating Student Learning," in Middle States Commission on Higher Education, *Student Learning Assessment: Options and Resources*, 2nd ed.
- Teaching and Learning Enhancement (TALE) Center; see especially Outcomes Assessment web pages and Teaching Resources → Writing Test Questions.

Written and sources gathered by L. M. Stallbaumer-Beishline, PhD
Associate Professor of History
TALE Director
Bloomsburg University of Pennsylvania

http://www.bloomu.edu/tale