

# Measuring Student Learning

Lisa Harris & Marshall G. Jones

21

## Measuring Student Learning

Lisa Harris & Marshall G. Jones

Measuring student learning is critical in the teaching and learning processes and can serve many purposes. Instructors can use assessment results to plan future instruction, adapt current instruction, communicate levels of understanding to students, and examine the overall effectiveness of instruction and course design. The measurement of student learning can take place before, during, or after instruction. Before lessons are even developed, instructors need to know what students already know and can do related to the content. There is no point in wasting time teaching something students already know, or in starting at a level that is so advanced students don't have the prerequisite knowledge necessary to be successful. To that end, the learner analysis in instructional design could be considered a type of assessment. Giving a pre-assessment, also called diagnostic assessment, can provide instructors with this valuable information. Measuring student learning during instruction, a formative assessment, provides instructors with important information about how students are progressing towards the learning objectives while there is still time to adjust instruction. Instructor may ask questions such as:

- Are students getting it?
- Are they confused about something that needs to be retaught?
- Is it time to move on with new material?

Finally, measuring student learning at the end of instruction, a summative assessment, provides information about the degree to which students mastered the learning objectives.

This chapter outlines practical strategies instructional designers can use to develop high-quality assessments to measure student learning. Best practices are the same for constructing diagnostic, formative, and summative assessments. Links to additional tools and resources are also provided.

# Constructing High-Quality Assessments

High-quality assessments are those that lead to valid, reliable and fair assessment results. Validity refers to the trustworthiness of the assessment results. For instance, if a student gets 80% of test items correct, does that mean they understand 80% of the material taught? Does the assessment measure what it purports to measure, or is the final score polluted by other factors? For example, consider a test that assesses mathematical ability and is made up of word problems. When taken by an English language learner or by an emerging reader, does the test assess math, reading, or a combination of both? The reliability of an assessment refers to the consistency of the measure. Multiple-choice test items, when properly constructed, are highly reliable. There should be only one correct answer and it is easy to grade. Essay items or performance assessments, on the other hand, are more subjective to grade. Finally, the extent to which an assessment is fair is a characteristic of a high-quality assessment. Fairness is the degree to which an assessment provides all learners an equal opportunity to learn and demonstrate achievement. While some aspects of validity and reliability can be measured through statistical analysis, it is uncommon that such complex measurement procedures are used for typical classroom assessments. Attending to best practices in assessment alignment and test item and assessment construction helps instructional designers increase the validity, reliability, and fairness of assessment instruments.

## Assessment Alignment

One of the most important concepts in assessment is alignment. It is critical that assessments and assessment items are aligned with goals and objectives. It is impossible to determine the extent to which learners have met course or workshop goals and objectives if their knowledge and skills have not been assessed. Assessment alignment tables and test blueprints are two tools instructional designers can use to align assessments and assessment items with learning objectives.

## Learning Taxonomies and Learning Objectives

Learning taxonomies assists instructional designers in constructing both learning objectives and assessment items. Bloom's Revised Taxonomy and Webb's Depth of Knowledge (DOK) are two frameworks commonly used by educators to categorize the academic rigor of an assessment as a whole or individual assessment items. To increase the content validity of an assessment, the complexity of the individual test questions should align with the level of knowledge or skill specified in the learning goal. If a learning objective states that a student compares and contrasts information, it is not appropriate for test items to simply ask students to recall information. Likewise, if the learning goal states that students will be able to synthesize information, a paper-and-pencil test will likely not be a sufficient measure of that skill.

Bloom's Revised Taxonomy divides learning into three domains: cognitive, affective, and psychomotor (Anderson et al., 2001). This chapter focuses on the cognitive domain which consists of six levels that vary in complexity. The three lower levels (remembering, understanding, and applying) are referred to as lower order thinking skills also called LOTS. The top three (analyzing, evaluating, and creating) are referred to as higher order thinking skills, or HOTS. Lists of verbs associated with each of these levels are readily available on the web and are very instrumental in helping instructional designers write measurable learning objectives and test questions that go beyond recalling definitions. (For an example, see: <https://edtechbooks.org-EZbp>.)

Similar to Bloom, Webb divides levels of knowledge into increasingly complex categories. These include recall and reproduction, skills and concepts, strategic thinking, and extended thinking (Webb, 1999). Student tasks range from a student being able to recall facts to synthesizing information from a variety of sources. A description of tasks at each level can be found online at <https://edtechbooks.org/-bVW>. These descriptions can help instructional designers design assessment tasks that range in complexity.

## Assessment Alignment Tables

Regardless of the assessment method, instructional designers can ensure that learning goals, objectives, and assessments align by creating an alignment table. In the example below, course goals, student learning outcomes, and assessments are aligned in a table. This example is from a college level course on teaching with technology for pre-service teachers. This table indicates there is at least one learning objective aligned with each course goal and at least one assessment method aligned with each objective. If you find that a particular learning objective isn't being assessed, you can go back and develop an assessment to measure the learner's progress. A link to an Assessment Alignment Table Template is provided at the end of this chapter in the Additional Resources list.

**Table 1**

*Example Assessment Alignment Table*

Course Goal	Student Learning Objective (SLO)	Assessment(s)
Plan and implement meaningful learning opportunities that engage learners in the appropriate use of technology to meet learning outcomes.	SLO1. Develop a technology integrated activity plan that meets the needs of diverse learners (e.g. ELL, at-risk, gifted, learners with learning disabilities).	Technology Integration Portfolio
	SLO2. Explain how and why to use technology to meets the needs of diverse learners (e.g. ELL, at-risk, gifted, students with learning disabilities).	Technology Integration Portfolio Midterm
Use technology to implement Universal Design for Learning.	SLO3. Describe the elements of UDL included in the technology integrated activity.	Technology Integration Portfolio
Model and require safe, legal, ethical, and appropriate use of digital information and technology.	SLO4. Describe legal, ethical, cultural, and societal issues related to technology.	Midterm
		Final

## Table of Specifications

In addition to creating an alignment table for all assessments in the entire course, instructional designers can also create a table of specifications, or test blueprint, to align individual test items to course objectives. A table of specifications aligns the learning objective, all items on a single test, and the level of knowledge being assessed. This is evidence of content validity. This also helps the instructional designer see if the test includes items related to all the learning goals, and if the assessment items are written to elicit knowledge at the appropriate level of complexity. If you find that you have too many questions about one topic or not enough about another, or that you are only asking lower level questions when the learning objective is focused on higher order thinking skills, the test can be edited accordingly. The figure below shows a test blueprint for a 12-item test about

assessment. Each number represents the question number on the test. A link to a Table of Specifications Template is provided at the end of this chapter in the Additional Resources list.

**Table 2**

*Sample Test Blueprint for a 12 Item Test*

Learning objective	Level of Knowledge	
	Lower Order	Higher Order
Analyze learning objectives in terms of format, specificity, reasonableness, and alignment.	1, 2	8, 12
Explain the importance of alignment when designing lessons and assessments.	3, 5	10
Compare and contrast reliability and validity of classroom assessment	4, 6, 7	11, 9

## Assessment Formats

Common assessment formats include multiple-choice and essay questions, observation, oral-questioning, and performance-based assessments. This chapter focuses on paper-and-pencil tests and performance assessments. Best practices in constructing each are described below. These guidelines help increase the validity, reliability, and fairness of assessments.

### Multiple-Choice Best Practice Guidelines

Multiple-choice items are very easy to grade (assuming there is only one correct answer) but very difficult to write. Coming up with plausible distractors, or the incorrect responses, is the hardest part. If some answer choices aren't plausible (ones that are meant to be funny, for example), the probability that a student will be able to guess the correct answer increases. It is also difficult, but not impossible, to write multiple-choice questions that assess higher-order thinking skills. Tips for constructing multiple-choice test questions that assess HOTS are provided below.

1. All answer choices should be similar in length and grammatically correct in relation to the item stem.
2. Avoid "all of the above", and "none of the above" answer choices.
3. Avoid confusing combinations of answer choices such as "A and B"; "B and C"; "A, B and C but not D".
4. Avoid negatively stated stems. If you must use them, bold the negative word to make it what you are asking clearer to the learner.
5. Avoid overlapping answer choices. (This most commonly occurs with number choices.)
6. The item stem should make sense on its own and not contain any extraneous information.
7. Don't include any clues in the item stem that would give the answer away.
8. Don't include too many answer choices. Typically, multiple choice questions contain four options.
9. Ensure the correct answer is the best answer.
10. Randomize the order of the correct answers.

**Table 3**

*Examples of Poor and Improved Items*

Poor Item	Improved Item	Explanation
<p>If a boy is swimming two miles an hour down a river that is polluted and contains no fish and the river is flowing at the rate of three miles per hour in the same direction as the boy is swimming, how far will the boy travel in two hours?</p> <p>a. four miles b. six miles c. ten miles d. twelve miles</p>	<p>A boy is swimming two miles per hour down a river relative to the water. The water is flowing at the rate of three miles per hour. How far will the boy travel in two hours?</p> <p>a. four miles b. six miles c. ten miles d. twelve miles</p>	<p>The poor item contains extraneous information and a confusing sentence structure. In the improved item, the extraneous information was removed. In addition, the prompt was broken up into several sentences and the actual question stands on its own.</p>
<p>Which one of the following is not a safe driving practice on icy roads?</p> <p>a. accelerating slowly b. jamming on the brakes c. holding the wheel firmly d. slowing down gradually</p>	<p>All of the following are safe driving practices on icy roads <u>EXCEPT</u></p> <p>a. accelerating slowly. b. jamming on the brakes. c. holding the wheel firmly. d. slowing down gradually.</p>	<p>When reading the poor item, a test taker may not recognize that they are being asked to pick a non-example of a safe driving practice. In the improved item, the word "except" is in all caps and underlined to call attention to what is being asked.</p>
<p>In most commercial publishing of a book, galley proofs are most often used _____ .</p> <p>1. page proofs precede galley proofs for minor editing. 2. to help isolate minor defects prior to printing of page proofs. 3. they can be useful for major editing or rewriting. 4. publishers decide whether book is worth publishing.</p>	<p>In publishing a book, galley proofs are most often used to</p> <p>1. aid in minor editing after page proofs. 2. isolate minor defects prior to page proofs. 3. assist in major editing or rewriting. 4. validate menus on large ships.</p>	<p>In the poor item, each answer choice is not grammatically correct in relation to the item stem. Often, a test taker can pick out the correct answer choice because it is the only one that is grammatically correct and not because they actually knew the answer. In the improved item, the item stem and answer choices have been edited so that they are all grammatically correct.</p>

## Tips for Writing Higher Order Thinking Multiple-Choice Questions

**Tip 1:** Use scenarios or provide examples that are new to learners. This allows you to ask learners to do more than simply recognize the correct answer. (Note that this can be problematic if you are assessing struggling readers or ESL learners. Know your audience!)

**Tip 2:** Develop multiple-choice questions around a stimulus you provide such as a map, graph, diagram, or reading passage. These are called interpretive exercises. Interpretive exercises include a set of data or information and a series of multiple-choice questions having answers that are dependent upon the information given.

## Best Practice Guidelines for Writing Essay Items

Essay questions are a good way to assess deep understanding and reasoning skills. Students can provide more in-depth answers in essay questions. Essay questions are also much easier to write than multiple-choice items. They are, however, harder to grade. Below are best practice guidelines for constructing and grading essay items and some real-world examples.

- Select the most important content in the workshop or unit to assess with essay times. Using essay items limits the amount of content you can cover on any one test because they take more time for a learner to answer. If one topic is less important than another, consider only asking multiple-choice questions about it.
- Write the prompt to focus learners on the key ideas they should address in their response. For example, tell learners how many reasons should they give, or how many examples should they provide. Stating directly what you want means that the learner doesn't have to try to interpret how much is enough.
- Break multi-faceted questions up into individual items. If the question is very long, make it more than one essay question on the test. This helps focus both the test taker and the grader.
- Include scoring criteria with the prompt and assign appropriate point values. If you want someone to provide three reasons why the Renaissance began in Italy, decide how many points each reason should count and make that clear to the learner. It is very difficult to objectively grade an essay question worth 10 or 20 points without first determining the grading criteria.
- Only include essay items that require higher-order thinking. Essay questions are too time consuming to grade. If it can be assessed with a multiple-choice question instead, don't waste valuable time reading essay answers.
- Avoid allowing learners to select which essay items they answer. This keeps learner scores comparable. If learners can choose which essay questions to answer, the test is not assessing the same thing for all students.

Note: Essay items can also be assessed with rubrics. See Performance Assessments and Rubric Development for more information on how to construct a rubric.

## Essay Item Examples

Below are examples of high- and low-quality essay items. Note that the high quality examples include explicit instructions about what needs to be included in the answer. In addition, how the points will be allocated is clear. The low quality essay items are both very broad in scope. A test taker could easily answer the question without touching on any of the topics the instructor wanted them to

include in their answer. In addition, it isn't clear to the test taker or the instructor how the points are allocated. This can lead to inconsistencies in grading.

### High-Quality Examples

1. Proof 1: Given ABC is equilateral, and BD is the angle bisector of angle ABC. Prove that the measure of angle ADB and angle CDB is equal to 90 degrees. Provide the statement and reason for each step using the two-column proof format. (1/2 point for each correct statement and 1/2 point for each correct reason given. 8 total points.)



2. Compare and contrast large-scale assessment and classroom assessment on the dimensions of frequency and nature of feedback. (2 points frequency, 2 points feedback. 4 total points)

### Low-Quality Examples

1. Explain weather and climate. (20 points)
2. Describe the three principles of Universal Design for Learning. Do you believe they should be used to guide instruction? Why or why not? (10 points)

### Best Practice Guidelines in Developing Performance-Based Assessments

Performance-based assessment allows learners to apply knowledge and skills in authentic situations. Performance-based assessment results in the creation of a performance or a product. Performance examples include public speaking, inventing something to solve a problem, putting on a play, or playing in a basketball game. Public service announcements, digital videos, and infographics created by learners are examples of products. Consider the following guidelines when constructing performance assessments:

1. Design a task that applies to real-world situations. The more authentic a performance-based assessment can be the more meaningful it will be to the learner, although access to resources and time will certainly impose project limitations. For example, writing a paper on gardening, designing a garden, and creating a garden are all examples of performance tasks with varying degrees of authenticity.
2. Develop a task description that includes the following:
  - a. Purpose/learning objectives. Why are the learners completing this task? Write the learning objectives in learner friendly language.
  - b. Clear directions. Break down the task into its component parts. Don't assume learners know how to jump immediately into creating the final performance or product.
  - c. Perimeters and constraints. How much time do the learners have to complete the project? What resources are they allowed to use? Is it a group or individual project? Who are they allowed to ask for help?
  - d. Assessment criteria. How will the performance or product be graded? This is discussed in more detail below in the Rubrics section.

3. Develop any job aides learners will need in order to complete the task. Do you need to teach any additional skills such as how to locate articles in a database, how to measure volume, or how to use a particular piece of software?
4. If at all possible, provide learners with an example.

## Rubrics

As discussed earlier in the chapter, reliability is related to scoring consistency. One way to help ensure scoring consistency is to use rubrics for grading subjective assessment items, including essay questions and performance assessments. Rubrics focus the attention of a grader on what is most important about the assignment. Rubrics include topics or elements and descriptions of levels of performance. This provides a roadmap for how to assess an assignment that is more subjective than a multiple-choice question. Without a rubric, it is easy for a grader to grade for one thing for the first 10 papers and grade for something else the last 10 papers. This occurs when an instructor has a lot of papers to grade, grading takes place over several days, and if more than one instructor is grading the same assignment. Providing a rubric up front is also beneficial to the student. They communicate to the student from the beginning what is important, on what to focus, and where to spend time and energy.

There are three types of rubrics: holistic, analytic, and single-point. This section will focus on analytic rubrics, because they allow instructors to assess the component parts of the performance assessment individually and provide the clearest grading criteria. Several additional resources about the different types of rubrics are provided below.

An analytic rubric consists of criteria, levels of performance, and descriptors.

### Figure 1

*Example of an Analytic Rubric*



Rubric				Levels	Descriptors
Criteria	Exceeds Expectations	Meets Expectations	Below Expectations		
Accuracy of information	Product includes accurate references to the UDL Principles and Guidelines. 3 points	Product includes accurate references to the UDL Principles or Guidelines only. 2.5-2 points	Product includes inaccurate references to the UDL Principles and/ or Guidelines. < 2 points		
Depth of Knowledge	Product shows use of <b>complex thinking</b> about what Universal Design for Learning is and why it is important for teaching and learning. Explanation/ justification is clear and includes multiple and varied facets of UDL. 3 points	Product shows <b>application</b> of UDL concepts and why it is important for teaching and learning. Explanation/ justification may lack clarity. Justification includes multiple and varied facets of UDL. 2.5- 2 points	Product shows a <b>basic knowledge</b> of UDL. Or Information about UDL is inaccurate. Or Justification does not include multiple or varied facets of UDL. < 2 points		
UDL Examples	Product includes concrete examples of UDL Principles and Guidelines. 3 points	Product includes concrete examples of UDL Principles only. 2.5- 2 points	Product does not include concrete examples of UDL Principles or Guidelines. Or Examples may be general. < 2 points		

## Best Practice Guidelines for Creating Rubrics

1. Determine the criteria. Criteria can be written as a learning objective or category. Criteria should be measurable, important to the performance task, and taught. For example, creativity is often assessed in performance-based assessments. If creativity was not explicitly taught, it shouldn't be measured.
2. Determine the weight of each criteria. Will they all be worth the same amount of points or will some count for more than others?
3. Determine the number of performance levels. How many levels of the rating scale will be delineated on the rubric? Will they be numbers such as 4, 3, 2, 1 or descriptive such as developing, meets expectations, and exceeds expectations. Typically, analytic rubrics contain three to five performance levels.
4. Write descriptors for each of the performance levels. This is the hardest part! Descriptors should address the quality of the product. It is okay to count project elements for some of your criteria (i.e. number of references, number of graphs), but not for all of them. See examples of quality and numerical descriptors below.

## Numerical Descriptors vs Quality Descriptors Example

**Table 4**

*Numerical Descriptors in an Annotated Bibliography Rubric*

	4	3	2	1
--	---	---	---	---

<b>Quality / Reliability of Sources</b>	All sources cited are reliable and trustworthy. 5 points	At least 80% of sources cited are reliable and trustworthy. 4-3 points	At least 50% of sources are reliable and trustworthy. 2 points	Less than 50% of sources cited are reliable and trustworthy. 0-1 point
---	---	---	---	---

**Table 5**

*Quality Descriptors in a Technology Lesson Plan Rubric*

	<b>Exceeds Expectations (A)</b>	<b>Meets Expectations (B to C)</b>	<b>Below Expectations (C-and below)</b>
<b>Teacher candidate develops a learner-centered, technology-integrated activity that promotes creativity, collaboration, or communication, and results in a learner-created product.</b>	Activity promotes significant learner engagement through creativity, collaboration, and communication. Actively includes opportunity for learner to create a product. 5 points	Activity promotes creatively, collaboration, or communication and focuses on learner engagement with technology. Actively includes opportunity for learner to create a product. 2-4 points	Activity focuses on teacher-use of technology but lacks opportunities for learner engagement and/or product creation 1 point

Note also that the rubric element directly above is written as a learning objective rather than simply a category.

## Conclusion

Aligning test items and performance assessments to learning objectives, using best practice guidelines to create assessments, and using rubrics to grade complex tasks, are strategies instructional designers can use to develop high-quality assessments. High-quality assessments provide instructors with accurate information regarding the extent to which learners met the learning objectives, a critical component of the teaching and learning process. Accurate assessment results help instructional designers plan future instruction, adapt current instruction, communicate levels of understanding to students, and examine the overall effectiveness of instruction and course design.

## References

- Anderson, L.W. (Ed.), Krathwohl, D.R. (Ed.), Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., & Wittrock, M.C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives (Complete edition)*. New York: Longman.
- Webb, N. (1999). *Alignment of science and mathematics standards and assessments in four states (Research Monograph No. 18)*. Washington, DC: CCSSO.

## Additional Readings and Resources

- [Assessment Alignment Table Template](#) (also included at the end of this chapter)
- [Making Data Driven Decisions using Assessment Data](#)
- [Single-point, Analytic, and Holistic Rubrics](#)
- [Rubric Wordsmith](#)
- [Table of Specifications Template](#) (also included at the end of this chapter)

## Templates

### Assessment Alignment Tables

Use this template to align your course goals, student learning objectives and assessments. This table helps instructional designers ensure that they have assessed all course goals and objectives. Delete the sample goals, objectives, and assessments in the blue font below and insert your own.

**Table 6**

*Example Alignment Table*

<b>Course Goal</b>	<b>Student Learning Objective (SLO)</b>	<b>Assessment(s)</b>
Plan and implement meaningful learning opportunities that engage learners in the appropriate use of technology to meet learning outcomes.	SLO1. Develop a technology integrated activity plan that meets the needs of diverse learners (e.g. ELL, at-risk, gifted, learners with learning disabilities).	Technology Integration Portfolio
	SLO2. Explain how and why to use technology to meets the needs of diverse learners (e.g. ELL, at-risk, gifted, students with learning disabilities).	Technology Integration Portfolio Midterm
Use technology to implement Universal Design for Learning.	SLO3. Describe the elements of UDL included in the technology integrated activity.	Technology Integration Portfolio
Model and require safe, legal, ethical, and appropriate use of digital information and technology.	SLO4. Describe legal, ethical, cultural, and societal issues related to technology.	Midterm Final

### Table of Specifications Template

Use this template to create a table of specifications for a test. This table helps you align the learning objective, all items on a single test, and the level of knowledge being assessed. Using this table helps you ensure that each item on the test is related to the learning goals and that you are asking higher order questions about the topics. Add rows to this table as necessary.

**Table 7**

*Test Blueprint Template*

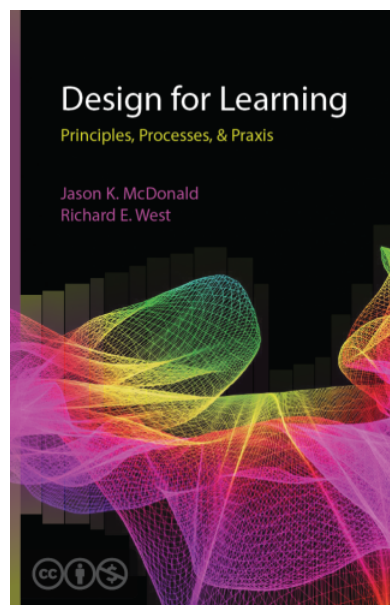
<b>Learning objective</b>	<b>Level of Knowledge</b>
---------------------------	---------------------------

**Lower Order Test Question #s****Higher Order Test Question #s**

- Learning Objective #1
  - Learning Objective #2
  - Learning Objective #3
  - Learning Objective #4
  - Learning Objective #5
- 

**Suggested Citation**

Harris, L. & Jones, M. G. (2020). *Measuring Student Learning*. In J. K. McDonald & R. E. West (Eds.), *Design for Learning: Principles, Processes, and Praxis*. EdTech Books. Retrieved from [https://edtechbooks.org/id/measuring\\_student\\_learning](https://edtechbooks.org/id/measuring_student_learning)



Harris, L. & Jones, M. G. (2020). *Measuring Student Learning*. In J. K. McDonald & R. E. West (Eds.), *Design for Learning: Principles, Processes, and Praxis*. EdTech Books. Retrieved from [https://edtechbooks.org/id/measuring\\_student\\_learning](https://edtechbooks.org/id/measuring_student_learning)



**CC BY-NC:** This work is released under a CC BY-NC license, which means that you are free to do with it as you please as long as you (1) properly attribute it and (2) do not use it for commercial gain.