



## Accreditation Like a Boss

### Pro Tips for Connecting Accreditation and Assessment

---

As a higher education professional, you know that accreditation is a powerful tool for communicating your institution's value. Accrediting bodies provide external validation of the value and quality of your institution's programs. They develop educational standards and measure your programs against those standards in regular audits.

But how can you connect those standards, along with those of your program and course expectations to specific student learning outcomes you can measure? How can your measures prove that students are achieving these outcomes and goals? This article explores methods top assessment developers use to ensure their course assessments connect directly to both inside and outside measures of achievement.

### Setting the Stage

The first step is taking a hard look at the educational standards available. Those standards come from three key places:

- Accrediting body requirements
- Stated program expectations
- Individual course syllabi

The difficulty is that none of these documents are typically written with specific, *measurable* (i.e., testable) student learning outcomes in mind. They are often lofty—which is good; a person's reach should exceed their grasp, as Robert Browning says—and high-level because their goal is to identify the broader-reaching purposes of the department, program, and course. But in and of themselves, they are usually not specific enough to be measurable in an exam or assignment.

Because accrediting bodies require proof that your institution is achieving these goals, you need to connect assessment items directly to the goals. While you can retrofit existing assessments to these standards, you may want to consider a fresh approach.

That fresh approach involves stepping back and creating an assessment plan or blueprint for each key course involved in the accreditation. In an ideal world, you would do this at the program level. However, it can be difficult to get all the instructors for a program in the same room at the same time. We recommend starting at the course level, then rolling those course plans up to the program level to review for complete coverage and consistency.

### Identifying the Standards

The first step in creating a course assessment blueprint is to identify the standards and student learning outcomes you want to measure.

Let's look at a specific example from University of California Riverside, Bourns College of Engineering, *Engineering (ENG) 180W: Technical Communication for Engineers*. This upper-division, required course is a keystone in UCR's Accreditation Board for Engineering and Technology (ABET) accreditation.

ABET requires:

- An ability to function effectively on teams to accomplish a common goal
- An understanding of professional, ethical, legal, security and social issues and responsibilities
- An ability to communicate effectively with a range of audiences
- An ability to analyze the local and global impact of computing on individuals, organizations, and society
- Recognition of the need for and an ability to engage in continuing professional development

Bourns College requires in addition:

- A knowledge of contemporary issues

And ENG 180W further requires:

- An ability to participate in and contribute to discussions and meetings, both in leading and non-leading roles
- An ability to make cogent, well-organized verbal presentations with and without visual aids prepared via presentation software
- An ability to produce cogent, well-written documents (including email)
- An understanding of professional and ethical responsibility, particularly regarding well-designed human interfaces including documentation and other written communications
- An understanding of what is expected in the professional workplace, including the need for long-term professional development

That's a lot of expectations—and none of them are directly measurable as written. However, they all contain the seeds of assessment items that can be realized through assessments and assignments (which are a form of assessment, after all).

## Deconstructing the Standards

Next up is to create a plan that identifies specific foci for assessment, which will then drive item development.

Standard/ Objective/ Outcome Reference	Standard Text	Item Focus	No. of Items	Item Type
ABET Section IV	Students must demonstrate an ability to function effectively on teams to accomplish a common goal	Given a team scenario, address how you would respond in that scenario	2	ER
ABET Section V	Students must demonstrate an understanding of professional, ethical, legal, security and social issues and responsibilities	Answer a question about a specific ethical dilemma	1	SR
ABET Section VI	Students must demonstrate an ability to communicate effectively with a range of audiences	Given a specified audience, identify tactics for effective communication	3	MC/MR

This particular example shows one item focus per standard to provide a broad set of item focus examples. In a real plan, you would create several item foci for each standard.

Let's take a close look at one standard/item focus set. ABET Section VI requires that students must demonstrate an ability to communicate effectively with a range of audiences. That's too broad to measure in a single assessment item (although slightly easier to measure in an assignment or assignment suite). Instead, we want to identify a focus that can measure whether a student understands how to communicate effectively with *one* audience.

You can write an exam item much more easily for the standard as deconstructed above, because you're measuring only *one* thing (we'll talk more about item focus shortly). For example, you could create a multiple choice item as follows:

- When writing for a technology consumer (end user), which of the following should you AVOID?**
- A. Short sentences.
  - B. Illustrations.
  - ✓ C. Jargon.
  - D. Headings.

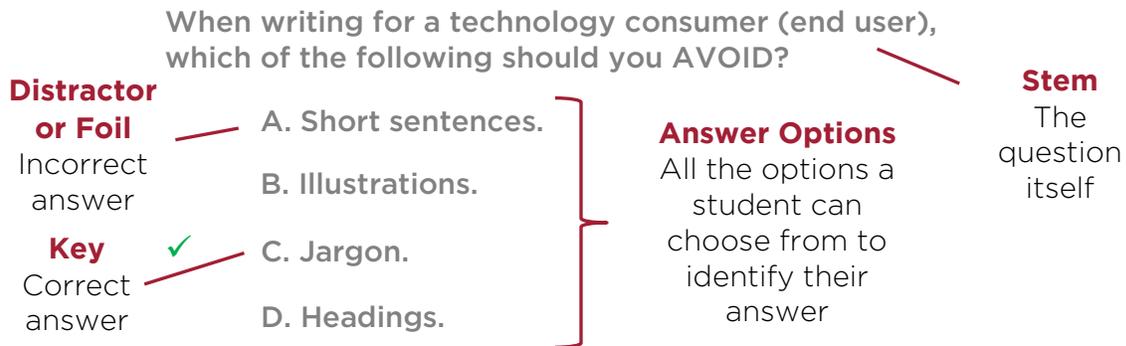
You could also create an extended response (essay) item as follows:

**Create an example of the introduction to a functional spec segment (intended to be used by other engineers on the development team) for a feature or function.**

We strongly recommend developing multiple items of multiple types for each item focus. This gives you flexibility to approach the concept from multiple depth of knowledge levels (more on that later). It also enables you to change up your exams term over term while maintaining equivalence of difficulty across exams.

## Developing Items

Once you have a complete plan covering your exam approach, it's time to start writing items. Before we provide tips, let's clarify some terms:



## Tips for Writing Items

- **Keep items independent**  
Items should be able to stand on their own rather than relying on previous items. This way, you can randomize questions and answer choices without fear of breaking any links. You also get cleaner statistics and a better idea of whether items are truly measuring what you intend them to measure. Also, be careful that no item contains an answer to another item on your exam.
- **Avoid trademarked or brand names**  
Companies can be justifiably touchy about preserving their trademarks and brands. Save yourself some research time and simply use generic terms instead.
- **Avoid living people**  
Nothing dates an item faster than using yesterday's pop culture phenom. Aside from avoiding potential privacy or bias issues, using fictional people makes your items more enduring.
- **Control cognitive demand**  
As you craft items, keep depth of knowledge or other cognitive complexity metrics in mind (see *Editing for Breadth and Suitability*, below). Further, avoid writing items with stems and answer options that are difficult to read and parse. The idea is to measure proficiency, not to trick students by including easy-to-miss clues like “now” vs “not” or including complex vocabulary and sentence structure (unless, of course, that's what's actually being tested). In addition, keep provided reading passages short and to the point. Do not include complex passages that add reading time without adding necessary information. Focus on making clear what's being tested.
- **Assess important content**  
Don't include extraneous questions that do not link back to your blueprint. We have limited space on assessments and we want to be sure that we only test on concepts that are directly related to our learning outcomes.

- Measure one concept at a time**

“And” is not your friend in item development. Any time you find yourself typing “and” in a question stem, ask yourself if you’ve fully deconstructed the standard being measured. Items that measure more than one concept at a time do not result in a true assessment of student proficiency because if students get it wrong, you don’t know which part the student didn’t understand. “And” in answer options may be acceptable, depending on what you’re trying to measure, although even then you may be better off with a multiple-response “choose all that apply” item type. Using the multiple response option helps you narrow in on exactly what element the student is not connecting to the concept.
- Phrase stems positively**

As noted earlier, words like “now” and “not” can be hard to distinguish in the time-constrained, stressful environment of a test. They have the effect of creating a “trick” question. While instructors have used trick questions as a pedagogical technique for decades (if not centuries), modern assessment development frowns on them, as they do not really contribute to distinguishing between students who know the material and students who do not. Looking back to our sample question, note that the stem not only uses “which should you avoid” instead of “which is not correct,” but it also emphasizes “AVOID” to ensure students see this important clue.
- Write clear answer options**

Make sure your answer options do not allow students to automatically eliminate distractors using grammatical clues. In the following simplified example, astute students can avoid two answer options immediately, reducing the effectiveness of the question to a 50-50 gamble:

**Jake brought a load of firewood in his truck.**

**In this sentence, the words “a load of firewood” can be described as a \_\_\_\_\_.**

- A. adverb phrase that modifies the subject**
- B. adjective phrase that modifies the subject**
- C. verb phrase that functions as a direct object**
- D. noun phrase that functions as a direct object**

## **Avoiding Bias**

We all have biases, whether conscious or unconscious. Since we can’t, as humans, avoid this, we *can* substitute careful thought and attention to control it. When you’re writing items, be sure to always ask yourself the following:

- Am I unconsciously applying gender stereotypes? For example, do I always refer to doctors as “he” and nurses as “she”?
- Do I fall back on using names that are familiar to me that betray an unconscious culture bias? For example, do I always use “Sue,” “Sam,” “Joe,” “Ann”? (And see how easy it is to have an unconscious bias? We just did is right here with an

assumption about what *your* culture is.) Try using common names from multiple cultures in addition, such as “Rashid,” “Humberto,” “Priya,” “Ming-Na,” etc. If you are not sure of common names from cultures other than your own, a quick internet search can turn up lots of options.

- Do I tend to use unconscious cultural stereotypes in examples? For example, are my Latino names always showing up as gardeners or preparing tacos? That’s an extreme example, but you’ll also want to watch out for more subtle instances. We live in a multicultural society where people from all different backgrounds can and do participate in walks of life. Make sure your examples reflect that. While these types of examples may have their place if used consciously in classroom discussion, they have *no* place in assessments where you cannot provide the context of a broader discussion.

## Editing for Breadth and Suitability

As you’re drafting your suite of items, make sure you’re including at least one item for every identified standard/item focus in your blueprint. This ensures you have the breadth of coverage necessary to truly measure your student learning outcomes. If you only have one item per item focus, don’t worry! Item banks are living documents—you’ll be adding and refining items throughout the life of your course.

You also want to consider creating items that measure different levels of cognitive complexity. One example of a framework for cognitive complexity is Webb’s Depth of Knowledge<sup>1</sup> (DOK) scale. While developed for a K–12 environment, it can also be useful in a higher education settings as well. The levels represent a hierarchy of proficiency that you can reflect in your assessment items. Other examples of complexity frameworks include Bloom’s<sup>2</sup> or Marzano’s<sup>3</sup> taxonomy.

While it is beyond the scope of this article to explain the details of these methods of determining cognitive rigor, you can certainly use an internet or library search to begin or deepen your understanding of these systems (see the footnotes for some starter links). The key takeaway is to provide a consistent and balanced approach to cognitive rigor and complexity. For example, if your entire test delivers only items that fall into Webb’s DOK Level 1, you may not be measuring the full intent of the learning outcome.

However, don’t assume that levels with higher numbers are automatically better. The levels are intended to provide a scaffolding that enables you as the instructor to delve deeply into exactly where a student is struggling. If your test includes only items that fall into Webb’s DOK Level 3, you may not gain enough intelligence about exactly where your students are struggling, which will make it difficult to adjust instruction to correct that issue.

In addition, the balance may change depending on when in the term you give the test. It may be more appropriate to focus on Webb’s DOK Level 1 and 2 items early in the

---

<sup>1</sup> See <http://www.webbalign.org/> for more information about Webb’s Depth of Knowledge scale.

<sup>2</sup> See <http://www.bloomstaxonomy.org/> for more information on Bloom’s Taxonomy.

<sup>3</sup> See <http://www.marzanocenter.com/> for more information on Marzano’s Taxonomy.

term, when instructional adjustment is still possible, and tend toward a higher balance of Level 3 items on the final.

## Reviewing Items Before You Administer

Once you have your items drafted, it's a good idea for you and at least one other associate—if not your entire program team—to review all the items you have created. We've provided a checklist at the end of this document (if you're reading the PDF) or made the checklist available for download (if you're reading it online) to help. Refer to the checklist for further details.

## Analyzing Item Performance

One can (and many people have) written entire textbooks on examining item statistics to determine their quality. However, unless you are developing a high-stakes exam, such as an exam that determines certification or licensure, many item statistics deliver rapidly diminishing returns—that is, the work you put into collecting and understanding them may be overkill for your exam. In most cases, you only need two statistics to determine sufficient reliability and validity for interim and final exams in college courses:

- P-value, or distractor analysis
- Point-biserial, or stem/scoring analysis

### P-Value, or Distractor Analysis

The p-value, or item difficulty, is the percentage of the total group exposed to the item that got the item correct. This value is an indicator of whether an item is too easy or too hard. The optimal difficulty level for an item depends on the type of question and the number of distractors.

Optimal p-values are ideal but not always practical. For our purposes, items that have p-values less than 30% or more than 90% definitely need attention. These items should be replaced or revised as they are not reliably measuring the student learning outcome.

Another way you can use p-values is to check your distractors. Using the same basic range as identified above for items as a whole, you can further examine the p-values for each distractor and determine whether it is, indeed, distracting unprepared students.

Items with more balanced p-values for distractors as compared to the key are considered to be more reliable indicators of student proficiency.

## Point-Biserial, or Stem/Scoring Analysis

The point-biserial correlation coefficient measures the correlation between the correct answer on an item and the total test score of a student. The point-biserial identifies items that correctly discriminate between high and low groups, as defined by the test as a whole.

Generally, the higher the point-biserial, the better the item discrimination, and thus, the better the item:

- If the point-biserial is less than the minimum medium value, then the discrimination level is **low** (i.e., the question does not discriminate well between those who have demonstrated proficiency and those who have not).
- If the point-biserial is greater than or equal to the minimum high value, then the discrimination level is **high** and the item thus more reliable and valid—that is, the item measures what we think it does.

Use the following general criteria to evaluate test items.

General Point Biserial Interpretation	
<b>.30 and above</b>	Very good item
<b>.20 to .29</b>	Reasonably good item
<b>.09 to .19</b>	Marginal item; needs improvement
<b>below .09</b>	Poor item; reject or improve

You can also use the answer option point-biserial to identify issues with either your distractors or your key. Your key should have the highest point-biserial. If your distractors have high point-biserials (even if they're not higher than the key), you may have an item where you either identified the wrong key (it happens) or your distractors are confusing or ambiguous. Unless what you're testing is a student's ability to cut through ambiguity, we strongly recommend reworking the distractors.

## Conclusion

Ultimately, your tests should demonstrate student proficiency against identified learning outcomes in a valid and reliable way. This evidence helps strengthen your case that your courses and programs are worthy of accreditation. It can even make the accreditation audit easier for all concerned. Increasingly, accrediting bodies and even students themselves are demanding evidence that your programs graduate students who are prepared to pursue a career or the next steps in their advanced education. Connecting assessment directly to accreditation standards makes providing that evidence much easier.

## How Can Scantron Help?

Scantron is a long-time leader in assessment solutions and services. We've served the education market from kindergarten to professional career development for more than 40 years. Our assessment solutions provide a broad range of support for higher education.

**Need a department- or program-level assessment management solution installed locally behind your institution's firewall?** ParSystem by Scantron helps you develop and deliver tests without testing your patience. Its interlinked modules provide tools to create collaboratively developed testing items and exams that ensure consistent assessment across a program or department regardless of instructor. All items can be linked to accreditation, program, or course student learning outcomes and standards. You can deliver exams online, on paper, or both—all results for a test appear in the same gradebook and can be analyzed as a unit. Robust reports ensure you can use your testing data, not just have it.

**Need a secure online assessment development and delivery environment for department or program exam collaboration?** Scantron provides that, too!

**Need help developing assessment and analyzing results?** Regardless of how you want to access your exams, Scantron's team of assessment development and psychometric experts stand ready to assist you with any or all of the assessment development process. We can create items and tests for you, provide professional development to further grow your current subject-matter experts, or analyze what you already have and suggest improvements.

**Need rapid scoring based on high-quality, guaranteed-scannable forms?** Scantron's test scoring machines fit the bill. Instructors can securely score class quizzes and exams, then port the results back to their desks for analysis in Remark Classroom Edition by Scantron. And we package our forms so you can provide them in your campus bookstore.

Whatever the assessment assistance you need, Scantron has the products, tools, services, and expertise to help you ensure that you have the right program for your students. Our desktop and web-based software, combined with our length and breadth of experience, help you get the most out of your assessment. Contact us today and see how Scantron can meet you where you are and help you get to where you want to be.

### Contact Us:

1.800.722.6876

[www.scantron.com](http://www.scantron.com)

Scantron Corporation, 1313 Lone Oak Road, Eagan, MN 55121



## CHECKLIST

# Item Review Checklist

- Do the items match the blueprint/assignment?
- Is the cognitive complexity level appropriate and accurate?
- Are the vocabulary and concepts appropriate to the course?
- Are any names used in examples appropriate for the group being assessed?
- Are the stems and/or answer choices too easy? Too difficult?
- Are all distractors plausible?
- Is the correct answer for each item keyed correctly?
- Are students able to answer the item based on the information in the item?  
Is anything missing?
- Are there any items that you cannot edit because of missing information or an inability to answer?
- Is the item content completely clear? If ambiguous, make recommendations.
- Are the graphics clear and required to answer the item?  
Identify items with extraneous or missing graphics.
- Are there enemy items?
- Are the items formatted correctly?
- Are the items consistent with one another in style and format?
- Are the options consistent/parallel in structure/format?
- Do the items contain accurate quotations?



**INFORM INSTRUCTION  
TO IMPACT STUDENT  
ACHIEVEMENT TODAY!**

For a free consultation to meet your academic goals, call **800.722.6876** or visit us at [www.scantron.com/higher-ed](http://www.scantron.com/higher-ed) to learn more.

### *About Us*

Scantron® provides technology to accelerate student growth. We offer software and services with a focus on assessment, course evaluation, and automated scoring, delivered with the quality and support you've come to expect from Scantron.

