

## “Good Assessment”

All educators, no matter the context or subject, have an interest in measuring their students’ learning. However, actually measuring learning is not a trivial task. This presentation will explore the importance of assessment, alternative modes of assessment, and practical techniques you can take to your jobs as educators.

### Why assess student learning?

- Is assessment absolutely necessary for students to learn? (Of course not)
- Will all students learn if they know they won’t be assessed?
- We should assess students to guarantee minimal learning (mastery of critical knowledge and skills)!
- We should assess students to provide feedback for their learning!
- Good assessment keeps everyone honest (i.e. keeps instructors “honest” in our teaching as much as it keeps students “honest” in their learning)!

### What is wrong with “teaching to the test”?

- Nothing, so long as the test is valid (i.e. measures the intended learning)!
- An analogy: API standards for petroleum products (fuel, lube oil, etc.)
  - Does the public availability of API specs and test procedures make the test invalid?
  - Is “refining to the test” a bad thing? (Only if the test is not valid!)

### What is a “valid” assessment?

→ **Open a discussion on what constitutes a valid test**

- A valid assessment accurately measures *practicable* learning.
- A valid assessment measures the degree to which learned knowledge has become a ready and comfortable tool for the student to use in real scenarios.
- A perfectly valid assessment may be openly published without fear of weakening its effectiveness (i.e. it cannot be “cheated”)!
- A perfectly valid assessment is not only appropriate to teach to; it is appropriate enough to serve as (part of) the teaching material!

### The challenge of creating a valid assessment

- You’ve probably heard the expression, “Be careful what you wish for, you just might get it!” Well, there’s an educational equivalent: “Be careful what you *test* for, students just might *learn* it!”
  - If you test for memorable phrases and words, that might be all your students learn.
  - If you test for the ability to solve specific types of problems, your students might not learn to solve any other kind.
- Actual test question in Instrumentation: “Identify the two types of valves.”
  - Actual answer to this question: “*Hand* and *Automatic*” (!!!)
  - This teacher obviously emphasized this one distinction during class sessions, and was assessing his students’ abilities to regurgitate that one distinction.
- Actual test question: “How much pressure will there be at the bottom of 1 cubic foot of water?”
  - Actual answer to this question: 0.434 PSI
  - While this answer is correct for a cube-shaped column of water 1 foot by 1 foot by 1 foot, it might not be if the 1 cubic foot of water took another shape (e.g. 6 square inch tube filled with water 288 inches high; here, 1 cubic foot of water would exert 10.4 PSI of pressure).
- Possibly apocryphal story of an Army effort to train a computer to detect camouflaged tanks in forested settings: all photos without tanks happened to be taken on sunny days, while all photos with camouflaged tanks happened to be taken on overcast days. The computer merely learned to distinguish sunny from overcast days!
- The “Black Belt” problem: can a person earn a black belt in a martial art without actually being able to defend him/herself in an actual hand-to-hand combat situation?

## The root of the problem: people “compartmentalize” what they learn

- We tend to connect new knowledge to specific contexts, and have difficulty abstracting that new knowledge to other areas of application. This helps us learn more quickly than if we had to mentally “connect” all new knowledge to all previous experience. However, this tendency inhibits mastery.
- Imagine, if every time you learned something new, you had to mentally compare the new knowledge with the sum total of all previous knowledge and experiences to see if it “fit” before you could retain it! So, as a rule, we don’t do that, and thus we are content to limit new knowledge to very specific examples and contexts.
- A good assessment forces people to cross these “compartmental” barriers, proving whether or not the learning has become practicable.
  - Example: an operator who knows how to manage a crisis in their own unit may not do as well with a crisis in another unit, even if the physics of the process are identical.
  - Example: a technician proficient in the maintenance of a particular device freezes when confronted with a similar device of different manufacture.
- Just as case examples help illustrate concepts during teaching, they also work well to bridge “compartments” of knowledge during an assessment. The key is to use enough varied case examples to avoid “overcontextualization” (where the student only applies knowledge to one context).

## Assessment criteria – a teacher’s perspective

- Validity
- Cost/effort/time to develop
- Cost/effort/time to administer (proctor and grade)

## Assessment modes

→ Rank each one according to the previous criteria

- Written
  - Essay
  - Short answer
  - Multiple choice and True/False (on paper)
  - Multiple choice and True/False (on computer)
- Oral review (interview style)
- Direct demonstration of knowledge/skill
- Randomization as a tool for enhancing validity of assessments
  - A bit of randomness ensures students must cross those compartmental barriers
  - Too much randomness creates inequity: some students assessed harder than others

## To grade or not to grade . . . different ways to mark assessments

- Proportional (purely a percentage score or letter grade)
- Mastery (purely pass/fail) – repeat attempts allowed
  - For *critical* knowledge and skills, re-demonstration at mastery level is required (e.g. First Aid / CPR every two years to maintain certification).
- Proportioned mastery – score based on number of repeat efforts
- Proportioned mastery – score based on time to complete
- Proportioned mastery – score based on multiple levels of complexity

## What to do with marks?

- Be careful with “minimum passing” scores on proportional assessments. It opens a door to non-mastery of critical domains.
  - In the academic world, grades are everything!
  - In the industrial training world . . . ?
- Poll attendees as to how (or even if) they use marks at their facilities
- Use proportional scores strictly for feedback during the training period?
  - Use proportional scores to set individual re-training intervals?

## The emotional side of assessment

- How many people like taking a math test? No one, right? Even people who like math generally don't like math tests. Why is this?
- Emotion is involved in every human endeavor, assessment included. Gasoline samples don't “feel bad” if they fail an API test, nor do they enter the test with anxiety. Human beings do, and this must be recognized.
  - Mastery assessments, with their re-take provision, generally help to overcome this anxiety while maintaining absolute standards of performance.
  - Recovering points on a proportional exercise (must explain incorrect reasoning – guesses don't count)

## Testing the test!

- Can an untrained person pass it? – try this!
  - If you are sure your new trainees are new to this, you may offer the assessment as a pre-test!
- Can a competent person pass it? – try this!
  - Take people with demonstrated (on-the-job) mastery and see how they score!
- Identify what would be required for a student to cheat. The fewer available modes of cheating, the better.

## Summary advice for good assessments

- Poll responses from the audience – what did they glean from this seminar?
- Always ask yourself the critical question, “How well will this assessment accurately measure students' practicable learning?” and its corollary, “How can the assessment be ‘cheated,’ whether by mistake or by intent?”
  - Make use of “case examples” in your assessments to force students to de-compartmentalize their knowledge
    - BP's mandatory safety test for contractors at Cherry Point: requiring test-takers to reference refinery unit maps to answer questions, rather than just ask abstract questions.
  - Look for ways to “randomize” the assessment
  - De-fuse the emotional barriers to good assessment
  - If possible, “test” your assessment by having experienced and inexperienced people alike take it!