What Is Assessment, and Who Is It For?

W. Blaine Dowler

May 28, 2011

1 Introduction to Assessment

Assessment is one of the most important aspects of an education system. However, many of the world's education systems do not have the funding or equipment to perform assessments the way they should be done. Even when new and preferable systems are developed, they are not always implemented due to the time and financial costs of educating parents and students about the new methodologies. Many of these methods will be detailed in this nine lesson course, so that parents, students, and educators reading it will better understand the options and the need for changes to be made worldwide.

1.1 What is Assessment?

Most commonly, assessment is the measuring and reporting of student achievement or ability. Assessment techniques can be used for many other purposes, such as employee performance evaluations and more, but the most common uses are in the field of education, and that is the series of applications this course will focus on.

1.2 Why Do Assessments?

Assessments are done for a few broad purposes. The main purposes are as follows:

- To give a general picture of how successful a student was in a particular course.

- To inform students of their strengths and weaknesses.

- To screen candidates for a particular placement at work or school.

- To evaluate the effectiveness of instruction.

There is no single assessment tool or reporting technique that does all these jobs well. The current public school systems in North America tend to focus on the general picture (grades) and the screening (university and job placement) while, unfortunately, leaving out the tools detailing a particular student's strengths and weaknesses. This highly important aspect of assessment is the one students need to improve their grades and skills in the future, and it is the goal least effectively served by traditional grade reporting methods in North American public schools. Little Johnny may come home with a 68% on his report card, but that report card probably doesn't say which skills are among the 32% of the course he can improve on.

1.3   Who Are Assessments For?

Assessments can be done for a number of individuals. The most common individuals are students, parents, teachers and school administrators, and prospective employers or educational institutions. Although there is certainly overlap in the goals of these groups, there are goals that are distinct to each of them.

- **Parents and Students:** Parents and students are the groups most likely to have overlapping goals. (In most cases, though certainly not all, differences arise when the parental goals are loftier than the student goals.) The primary goals of these groups are to make sure the student is equipped with all the skills required to succeed in the future. Thus, though the prevailing opinion among parents and students is that overall grades are the most important reports, research and logic dictate that a skill-based reporting scheme is, in fact, the most effective reporting methodology.

- **Teachers and School Administrators:** Many teachers look at grades the same way parents and students do, with the intent to help student performance improve. However, there is an additional purpose for assessment, and that is the evaluation of instructional effectiveness. If a class usually averages 65% on unit exams in a particular course, and then averages 50% on a quiz based on a particular lesson, this is a strong indicator that the lesson delivered was not very effective, and needs to be revised and/or retaught. School administrators can also use overall class averages on assessments to determine the effectiveness of teachers as a whole, and determine which teachers are best suited to the course assignments. This can be difficult in situations in which individual teachers produce their

own assessment tools. If each teacher creates his or her own unit exam, then the marks may not be directly comparable, as the exams may not place the same emphasis on the same skills. Standardized assessments, be they standardized throughout the school, district, state or province, country or world tend to be the preferred assessments for this purpose.[1]

- **Prospective Employers and Educational Institutions:** This group has their own goal in mind: filtering out the "best of the best." This is a completely different type of evaluation. Rather than looking at individual skill performance, they are more concerned with how prospective employees or students performed relative to the class average.

Each of these goals is best served by a different reporting scheme.

2 Types of Assessment Items

An assessment item is an individual task or question used to evaluate student performance on a skill or group of skills. By "item," we typically refer to a single question rather than a complete test. Assessment items can be grouped into two broad categories of objective and subjective items.

2.1 Objective Assessment Items

Objective items are the questions which have a definitive right or wrong answer, which means these are the questions which are eligible for machine-scored assessments. The most common questions of this type are now listed, in order of difficulty levels for assessment item creation.

- **True or False Questions:** These questions are amongst the easiest to create. The possible responses are standardized for the entire section: it is either true or false. Students also rarely struggle with the instructions, and their responses are easy to mark. Unfortunately, because there are only two options, it is relatively easy for students to guess correctly. This is further compounded when the item is not well written, and specific determiners such as "never," "every," and "all" are used. When those

---

[1]No system is perfect, of course. A teacher who is aware of this evaluation scheme may be tempted to "teach to the test" in this situation if the test can be accessed in advance, thus preparing students for that assessment tool rather than for the general course outcomes. More details on this will follow in lesson three.

words are in place, the correct response is often "false," simply because there are exceptions to almost every rule.[2]

- **Fill in the Blank Questions:** These are only slightly more difficult to write than true or false questions. One simply thinks of a true statement, and them omits a word or term. The student must then complete the sentence. Marking is again efficient, and it is far less likely that the student answers the question correctly through a random and uninformed guess. However, as responses to these questions are typically hand written, they are difficult to machine score in a school not equipped with OCR technology, and need to be marked by humans. The human marking is very efficient, so it is typically not a problem when dealing with a single teacher's course load, but it gets dramatically more cumbersome with assessments at the school or district scale, such as year end finals.

- **Matching Questions:** These questions allow for a move diverse set of true/false question structure. The student is faced with a list of prompts as well as a list of possible responses, and needs to match one to the other. Again, marking is absolute and efficient, but there is a chance of guessing correctly. The odds of guessing correctly decrease as the number of options increases, and the odds increase when the matching is one-to-one, meaning every prompt matches a response (as they all should) and every response matches exactly one prompt. In those situations, a student who can correctly identify the responses to, say, nine out of ten prompts will likely match the remaining prompt to the remaining response, scoring 100% with an understanding of only 90%. This is further compounded when responses are not of the same type. For example, imagine a history assessment which attempts to evaluate a student's ability to match ten treaties to the years they were signed. If both the prompts and responses include years and treaties, then the odds of guessing correctly increase. In this situation, the best manner in which to measure this outcome with minimal odds of guessing in multiple choice would be to use the treaties themselves, sorted alphabetically,[3] as the prompts. The possible responses would then include all ten correct years, as well as a number off additional years from similar points in history.

- **Multiple Choice Questions:** These questions are the hardest objective test questions to create effectively, but they are amongst the easiest to mark, as most schools have the technology required to machine score the options. Again, the odds of guessing correctly are not zero, but four or five response options make this less likely. Also, if options "all of the

---

[2]The content of that sentence is the reason I included the word "almost" before the word "every" in that sentence.

[3]Human attempts to randomly order the treaties would result in subconscious associations, or in deliberate attempts to distance related treaties, so they wouldn't be random. An alphabetical sorting choice would randomize them effectively without providing any hints as to associations.

above" and "none of the above" appear on every single question in which they might apply, odds of guessing are reduced. (Most teacher created tests using these options tend to use them only when they are the correct responses.) Another tendency is to lack detail in the incorrect responses. As a student, if one is unsure about the correct response on a multiple choice assessment, selecting the longest answer tends to work out well. As a teacher, the greatest difficulty in creating multiple choice items is creating plausible distracters, meaning incorrect answers that are likely to appear correct to unsure students. (A well written distracter will tend to draw the attention of the weaker students, meaning the lowest performing students in a group have a strong tendency to choose that response instead of the correct one.)

Virtually all objective tests share a drawback: students are more often required to *recognize* responses than create them. These item types tend to be effective for many science and math assessments, but they are not very effective for language arts and social studies, where opinion based responses are more common, and must be generated by the students rather than the teachers.

2.2   Subjective Assessment Items

Subjective assessment items tend to be easier to create than objective assessment items, but can be far more difficult to mark. These include short answer and essay questions, in which the students must present ideas of their own. The prompts tend to be short, but effective scoring is far more difficult to achieve. For example, if the curricular outcome is "write a short story," and the teacher marking the results tends to prefer romance to science fiction, then two well made examples from the two different genres may get two different marks as a result of the teacher's subconscious bias rather than the product on the page.[4] However, these are the only assessment items which allow students to create and present their own ideas, so they are the only assessment items which can measure some curricular outcomes.

3   How Students Learn

Students learn in stages. Some outcomes and skills are easier to learn than others. These were detailed most effectively by Benjamin Bloom in what is called Bloom's Taxonomy. As students learn, they progress through six stages,

---

[4]This is why so many state, province or country wide standardized high stakes tests are marked by multiple individuals whose scores are averaged together.

assuming the skill is eventually mastered. If a student only progresses through three or four of those stages, then that student is more likely to regress in that skill area over time if that skill is not applied. Once the sixth level of the taxonomy is reached, the skill is truly mastered, and can be far more effectively retained over time despite disuse.

The six levels of Bloom's Taxonomy are as follows:

1. **Knowledge:** This is the lowest and easiest level to obtain, and the one which is the easiest to reach with objective assessment items. It only requires fact recognition and recall. For example, a student who does not understand that multiplication is the repeated act of addition may still memorize that $3 \times 4 = 12$, and answer such a question correctly with no true understanding of why 12 is the correct answer.

2. **Comprehension:** This is the next level, in which the student understands why the knowledge is true. This is the level often reached by students, particularly working at home with their parents. As the mass populace is not typically instructed in Bloom's Taxonomy, parents are often confused by the struggles students have when they reach this level but do not perform in school, or do not retain that information for tests. They begin to question student effort or physiology, as those are the causes for the struggles they are already familiar with, and may not identify the real cause of the problems: a lack of progression through the complete taxonomy. This is more pronounced when the parents make up questions to quiz students with while studying, as these questions are typically presented at this level.

3. **Application:** This is the next most effective level to reach. Students can then apply the knowledge to unfamiliar situations. For example, if a student has learned about multiplication as a mathematical abstract, and is then asked how many apples there are in the back of a truck with 30 crates of 144 apples each, then the student answers correctly by moving through the application level.

4. **Analysis:** This level is one in which the student examines the underlying processes of a skill and completely understands how they work. This is typically demonstrated by comparing multiple correct answers and choosing the best answer, or developing the best answer. This is the highest level of Bloom's Taxonomy which can be effectively evaluated with a machine scored or objective test, as all higher levels require responses to be created by the student rather than the test creator. Students typically complain about the "choose the best answer" questions because the students have not been taught about Bloom's Taxonomy, and their self study habits have only taken them to the application level.

5. **Synthesis:** This requires a level of convergence. Students need to take multiple skills and combine them to produce a new concept or approach not previously seen in class. Students tend to complain about these questions, too, with the argument "we were being tested on stuff the teacher never even taught us!" If the curricular outcome requires this stage of understanding, questions that challenging will appear.

6. **Evaluation:** This is the highest level of understanding. At this level, students are ready to evaluate their own ideas, and the ideas of others, to determine which are the most effective. If a student understands a skill at this level, the student will likely retain the skill for life.[5] As such, this is the level both students and teachers should strive for, particularly in the critical areas of reading and arithmetic.

4   Upcoming Lessons

Through the rest of this course, we will analyze the assessment tools needed for each of the primary purposes, with particular emphasis on how teachers and assessment administrators can ensure that the assessment items they use are appropriate to the goals at hand.

---

[5]Retention is absolutely critical for student success. If a skill is not retained, then it will not be present when a subsequent skill is taught, and that subsequent skill cannot be retained. For a rather obvious example, imagine trying to teach multiplication to a student who doesn't understand addition. If the foundation skill of addition is not present, the student cannot learn multiplication. This is the inherent cause for students to continue struggling after hiring tutors or getting extra help from the teacher outside of class: if a student is struggling in grade six, the problem is usually because that student has not retained a prerequisite skill from grade five, or four, or three, and so forth. When a typical tutor or teacher gives additional help on the grade level material, the prerequisite skills are not being rebuilt in the student's foundation, and the new skill will not be retained in the long term either. The student may be able to learn how to go through the motions and answer the homework, but that's operating at the first level of Bloom's Taxonomy only. The student will probably struggle with the same questions when the unit exam or year end final comes around.