

Original Paper

Measuring What You Test Testing What You Intend to Measure

Wade Smith, PhD¹

¹ Southeastern Louisiana University, Hammond, LA, USA

Received: February 5, 2022 Accepted: February 28, 2022 Online Published: March 19, 2022

doi:10.22158/fet.v5n1p76

URL: <http://dx.doi.org/10.22158/fet.v5n1p76>

Abstract

Public school teachers across America are supposed to differentiate instruction while simultaneously adhering to a highly scripted curriculum. Teachers routinely comment that they know they are teaching materials that are not retained, but they cannot adjust instruction due to the preset pacing of the curriculum. Adding to this problem is the expectation that teachers will stay on pace to prepare students for end-of-year exams. Understandably, these expectations are a significant contributor to stress incurred by teachers. This paper provides an analysis of these contradictions as well as suggesting a strategy for teachers to use that can provide for differentiated instruction while following highly scripted curriculum.

Keywords

active engagement, assessment, curriculum, instructional alignment

1. Introduction

Jeopardy is a long-running television game show. The premise of the game is that people can select dollar values in a subject category. When a value is chosen from a category an answer is revealed, and it is up to the player to provide the question that applies to the answer. Correct questions are awarded their assigned value and incorrect answers are penalized by the same amount. Whoever has the highest score wins at the end of the day.

The notion that more difficult challenges should be rewarded with higher gain is intuitive. Return on investment, where greater performance is met with higher compensation is a standard business practice in companies across America. That said, the risks associated with merit pay have been well known for some time. Lawler (1973), for example, points out that merit pay can result in portions of job descriptions being ignored or neglected if they are not in the performance plan. Additionally, merit pay

can result in employees artificially inflating data used in their performance evaluations. Finally, there is the possibility that group work norms may suffer when one person's work is singled out as exemplary.

Recognizing these risks is important to framing an optimal work environment that rewards extraordinary work while protecting against unintended consequences. Schools, like businesses, strive to recognize and reward excellence, primarily through the rewarding of letter grades and most specifically, letter grades on exams. As we shall see, there are also risks associated with providing harder test items with higher weights.

2. The Problem

Identifying mastery on a test is surprisingly easy when compared to identifying a basic level of achievement. As long as the test is aligned with instruction, the entire test can be written at the mastery level and those who have actually mastered the information will be able to demonstrate that fact. But what about the student who may be performing at a basic level? If the test is heavily weighted towards mastery questions, then it is quite likely their grade will be very low and do a poor job of demonstrating what they actually learned. For the first student, the test provides multiple opportunities to demonstrate mastery. For the second student, the test affords no opportunities to demonstrate a basic understanding of the material on the exam.

Schools are now under intense pressure to demonstrate their effectiveness vis-à-vis student success on standardized and high-stakes tests. In order for schools to score at the highest levels of a state's performance rankings, they need as many students as possible to score as high as possible. As a result, standards-based curricula with higher expectations have enjoyed heightened interest. The No Child Left Behind act of 2001 (NCLB) was perhaps the culmination of what has become a prolonged effort producing many standards-based reforms in school accountability. Supporters of these efforts believe that heightened accountability would result in improved instruction, more rigorous curriculum, and ultimately increased student performance (Berger, 2000).

As NCLB began to take hold, states began an effort to create a common testing platform so that scores could be compared from state-to-state in a direct manner. Ultimately, efforts by the National Governors Association and Council of Chief State School Officers, created the Common Core standards in 2009 and 2010 respectively. A primary expectation for Common Core was, as noted, creation of a platform that would allow for state-to-state comparison of student progress. This expectation has been essentially scuttled due to most states creating their own accountability measures and setting their own benchmarks for success (Nelson, 2015).

Although Common Core has evolved into a much broader and diffuse set of expectations that when originally envisioned, many of the initial expectations still impact the classroom. Teachers, for example, are commonly given scripted curriculums in an effort to comply with state and federal mandates (Griffith, 2008; Milosovic, 2007). Common complaints from teachers are that these materials are not flexible enough to meet the needs of all their students and that their required level of instruction is not sensitive to the various ability levels in their classroom (Hudson, 2017).

What teachers are describing is a tension between what is taught and what is tested. Not only what is taught, but at what level of difficulty. The issue of how to reward successful completion of challenging work is straight at hand.

3. A Possible Resolution

The trend towards scripted curriculum is established, as is the trend towards increased instructional expectations. One can debate the merits or shortcomings of this reality, but the time spent towards that end will not affect what is currently the order of the day. What is needed is a means to work within the system so that students can remain engaged in challenging but doable activities while teachers can simultaneously document that they are supplying the rigor embedded in their curriculum.

The solution impacts both teaching and assessment. One of the ironies of scripted curriculum is the ongoing expectation that instruction be individualized to meet the needs of all students. The challenge created by expecting flexibility from a highly structured delivery model approaches the level of an oxymoron. On the one hand, the curriculum is highly scripted and adherence to the curriculum an expectation. On the other hand, the teacher is supposed to meet the needs of the student(s) not responding to the highly scripted curriculum.

This is where we need to revisit and challenge the notion that higher levels of achievement should be rewarded with more points. The concept may work well in the workplace, but it has little application to a classroom. To receive high compensation in the workplace for a difficult task you first must demonstrate that you have achieved what would be the normal expectations. Rewarding excellence means that you have also recognized basic expectations were achieved.

Assessments, on the other hand, can and often do try to document excellence without allowing for a demonstration of basic expectations. As mentioned at the beginning of this paper, the most egregious example of that type of test would be one where all items were written at the mastery level. That test assuredly documents mastery, but it has no capacity to demonstrate basic competencies. In this case, the assessment is measuring what a student does not know instead of what they do know.

If a test is generated with twenty questions, and every question is worth five points, then the implicit assumption is that every question holds equal value. When I first became a principal, I filled out my first self-evaluation. The criteria were straightforward: Do you dress properly? Do you speak properly? Are you punctual? Are you organized? And so forth. But then, there was one criterion that struck me as standing by itself: Do you exhibit sound decision-making strategies? I filled out my self-evaluation and then I noticed that out of the 25 odd questions, I was going to be deemed deficient if I received a negative ranking from my superior for more than two of the categories. According to the rules set forth in my evaluation, I could be unable to generate a proper response to a crisis as long as I was punctual, organized, well-spoken, and dressed to the nines.

This is the kind of stuff that happens when you fail to differentiate between what is mission critical and what might be important. Tests, like my first performance evaluation, are often guilty of this oversight. To remedy the situation, it is not enough to identify which questions are at the basic level and which questions demand more expertise. It is also crucial to weight the items appropriately. As noted, it is a common thought that harder questions should be worth more points. But, considering the example of a test that asks question solely at the mastery level, we can see where this thinking is misguided.

If a grading scale for an exam is 90, 80, 70 and so on, then we can see that the top 10% of the test is supposed to differentiate for the A. It makes sense then, that the problems or questions intended to demonstrate A level work should make up approximately 10% of the grade. If they make up more than 10% then the test is sliding towards the extreme example we have noted, and students are being overly penalized for what they do not know rather than being rewarded for what they do know.

There is, therefore, an inverse relationship between the level of difficulty and the points assigned to the item. Easy questions should be worth more and harder questions should be worth less. Of course, this thinking has to be balanced against the recognition that a test should discriminate between basic, above basic, and mastery levels of attainment. Since this is the case, it might be reasonable to have 75% of a test at the basic level (a middle C on a 10-point scale). About 15% of the test would be geared towards above basic material and about 10% would be identified as mastery level.

Extending this thinking, an assessment might have multiple opportunities to document mastery of a skill. Even if there are five items that are linked to mastery level achievement, their collective score should be about 10% of the total test score, otherwise the assessment is beginning to measure what a student does not know rather than what they do. Importantly, the principle guiding the weighting of items can be applied to any grading scale. At its most extreme, a grading scale might be 98-100 for an A. If that is the case, the same five items would now be worth 2% of the total grading scale.

4. Linking Assessment to Instruction

Identifying difficulty levels for items and assigning proper weighting for these items is important, but insufficient. It is also important to align the instruction that takes place with the assessment. This is where the rubber meets the road for the typical teacher trying to balance the expected curriculum rigor with the various levels of skill and potential within their class.

Early on in Common Core's implementation, concerns were raised by leaders in the field. Initial results on Common Core assessments were generally dismal, with one principal remarking the difficulty for the exams were more akin to AP exams rather than a general classroom (Ravitch, 2013). Nearly a decade removed, the same concerns are still being expressed by classroom teachers and those concerns provided some impetus for the movement of states towards their own peculiar standards.

The standards may now vary state-by-state, but the curricula used by these states generally comes from a very small pool of suppliers. One of the selling points for those supplying curricula is their product is aligned to rigorous standards.

Aligning instruction to standards is all well and good. But one can reasonably question whether a standard may or may not have too high of a performance expectation. For example, a track coach might set a standard of 6'6" for the high jump, but that expectation might not be reasonable for an athlete just learning the fundamentals of the event. That analogy holds up well for the classroom. The high standard has been established, but it might not be reasonable for some students.

Fortunately, it is possible to maintain the standard as documentation of the highest level of achievement while also providing students with opportunities to demonstrate basic or above basic achievement. Just as the test must weight test items properly, it is also important to properly weight the time spent on the ascending levels of attainment.

So, for example, a class would spend a necessary amount of time making sure that students understand and can actively engage the basic expectations that will translate to being able to pass the forthcoming assessment with at least a C. How much time? That question has no set answer. There is a tension between pacing for basic and preparing for mastery, but that is nothing new, it has been this way since the first one-room schoolhouse.

Although it cannot be definitively stated how much time will be needed to address basic, it is quite possible that there will be an inverse relationship at work here as well, i.e., the basic expectations that will comprise the majority of the weighting for the test assessment would be expected to require less time to cover when compared to the higher-level expectations yet to come.

Explaining to students that the materials currently being covered will provide the majority of the points to be earned on the next exam would be a direct means of communicating to students that they must obtain this level of proficiency to pass the exam. Providing task specificity, such as suggested here, is known to contribute to heightened belief in one's ability to organize and execute successful courses of action, i.e., self-efficacy. That is no small achievement since self-efficacy has been demonstrated to influence how much effort is invested in a given endeavor, perseverance in the face of obstacles and failure, resilience to adversity, whether thoughts are self-aiding or hindering, levels of stress and depression while facing significant obstacles, which courses of action are pursued, and ultimately the courses of action chosen to pursue (Bandura, 1997).

If instruction is clearly identified as aligning with the basic, above basic, and mastery portions of the attendant assessment, teachers have also differentiated their instruction to meet the needs of the varying levels of ability and potential found in the class. As an aside, it is simply not possible to manage a class of 25 students, provide an effective lesson, answer the intercom, send the sick kid to the nurse, and find time to give significant individual differentiated instruction to the child who desperately needs it. It is possible, however, to provide that student and every other student with adequate opportunity to engage the tasks necessary to document basic proficiency.

4.1 Real World Application

I have collaborated with schools that have worked to properly weight their assessments while aligning their instruction to these assessments. These schools have ranged from college preparatory schools to schools serving high levels of poverty. In one particular case for the latter, a teacher was faced with the onerous task of teaching Algebra I to students that included 17-year-old students. She lamented their apathy and their resistance to engaging much of anything she was trying to present.

To address the problem, we introduced a criterion referenced learning guide (CRLG). This was a document given to the student that told them what basic looked like, the tasks associated with documenting basic, and when the assessment would take place. We also identified above basic and mastery, but the immediate task was to get students involved with the most basic level of skill documentation.

The teacher handed out the CRLG and followed it diligently. When the assessment was administered, the students' performance was unchanged. She was disappointed, but we were not done. The next day, when she returned to class, she pulled out the CRLG and the assessment and showed the students exactly how the CRLG was preparing them to pass the exam. This demonstration led one of the 17-year-old students to ask: "You mean to tell me you are going to tell us what we need to know?"

The next time we met she was pleased to announce that engagement and interest had both risen significantly. Bandura would nod his head approvingly. The students, knowing what was expected of them and that these expectations were a direct path to success had acted rationally and engaged those expectations. To be clear, the likelihood of 17-year-old ninth graders finishing school is not reasonable. That said, the students who had been left behind for so long were able to see that there was a path to success that might include a GED.

Positive results such as these were not limited to at-risk populations. In one instance, a math teacher at a school where 98% attended college upon graduation employed the strategy in her class. The students quickly realized that they had a ready means of identifying how to optimize their test score. After the first exam, the teacher reported that she observed students answering the questions that secured the highest point values first. After they “banked” as many points as possible, they turned their attention to the harder questions where the higher levels of achievement would be documented.

Similar reports have come from elementary schools. One third-grade teacher implementing the CRLG spoke passionately about how students were often left frustrated by tests that stressed the highest level of achievement. When she introduced the CRLG and aligned her assessment’s weighting to reflect the inverse relationship between difficulty and point allocation suggested in this paper she saw immediate results. Students who had experienced test anxiety were now much more at ease. Interestingly, many students were now attempting the more difficult problems with a higher level of energy. Removing the stress from the attempt facilitated active engagement and encouraged the students to view difficult tasks as challenges to engage rather than landmines that would obliterate their test score.

5. Conclusion

An interesting trend is emerging in education. Teachers are leaving the profession in ever greater numbers and the pool of the next generation of teachers is shrinking (Walker, 2022). Asking teachers to follow a highly structured curriculum while simultaneously expecting them to differentiate instruction is undoubtedly contributing to the exodus. Meanwhile, the next generation of teachers is understandably reticent to engage the very tasks that are driving their predecessors from the profession.

Emphasis upon student excellence is admirable and necessary. However, emphasis upon excellence that is not properly sensitive to documentation of basic skill attainment is misguided and fraught with logical and practical unintended consequences. Undue stress worsens negative reactions to assigned tasks. That is true of students, teachers, administrators, engineers, doctors, plumbers, and lawyers. It is as close to a universal truth as you can find. One way to mitigate stress is to provide a clear means to success. For schools, that requires proper weighting of assessments and subsequent alignment of instruction to that test.

References

- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H. Freeman. New York.
- Griffith, R. (2008). The impact of a scripted reading program on teachers' professional spirits. *Teaching & Learning*, 22(3), 121-133.
- Hudson, C. (2017). *Stakeholders debate pros and cons of new curriculum*. Retrieved from <https://www.thewashingtondailynews.com/2017/11/10/stakeholders-debate-pros-and-cons-of-new-curriculum/>
- Lawler III, E. E. (1973). *Motivation in work organizations*. Brooks/Cole Publishing. Monterey, CA.
- Milosovic, S. (2007). Building a case against scripted reading programs. *Education Digest*, 73(1), 27-30.
- Nelson, L. (2015). *Everything you need to know about the common core*. Retrieved from <https://www.vox.com/2014/10/7/18088680/common-core>
- Ravitch, D. (2013). *Why I cannot support the common core standards*. Retrieved from <https://dianeravitch.net/2013/02/26/why-i-cannot-support-the-common-core-standards/>
- Walker, T. (2/01/2022). *Survey: Alarming numbers of teachers may soon leave the profession*. Retrieved from <https://www.nea.org/advocating-for-change/new-from-nea/survey-alarming-number-educators-may-soon-leave-profession>