Why Educational Standards Are Not Truly Objective

Matthew Metzgar*
University of North Carolina at Charlotte

Abstract
Educational standards have become a popular choice for setting clear educational targets for students. The language of standards is that they are "objective" as opposed to typical tests which may suffer from bias. This article seeks to further analyze the claims that standards are objective and fair to all. The author focuses on six issues which illustrate the problematic nature of educational standards. Examples from the Common Core standards are chosen to show the range of problems associated with standards-based systems. Given these arguments, it is questionable as to whether educational standards represent a better alternative to norm-referenced tests.

***Keywords*:** Educational measurement, educational standards

*** Matthew Metzgar** is a Clinical Assistant Professor of Economics at the University of North Carolina at Charlotte. He teaches undergraduate economic classes in the Belk College of Business. Previously, Professor Metzgar taught undergraduate and graduate classes in Ohio and New York. He also spent several years working in the private sector.

Correspondence: mmetzgar@uncc.edu

Why Educational Standards Are Not Truly Objective

The college instructor blames the high school teacher, the high school teacher complains of the grade teacher, each grade teacher above first grade finds fault with the poor work of the teacher in the grade below, and the first grade teacher in turn is chagrined at the shortcomings of the home training. Must this go on indefinitely? Whose opinion shall prevail? Is it not possible to get away from personal opinion to an agreed-upon consensus of opinion? May we not replace the constantly conflicting subjective standards with definitely defined objective standards?

—Wilson & Hoke, 1921

Educational standards are often seen as a way to induce higher student performance (Post et al, 2008). Standards can provide a clear target that may increase student motivation and outcomes. Many K-12 schools across the country are now actively raising standards as a way to improve performance. Some of this recent activity is linked to the Common Core standards which have been adopted by forty-five states.

The current discussion of standards-based education often uses the language that standards are "objective". This is in comparison to a norm-referenced test that typically ranks students in a relative manner. As norm-referenced tests often produce a distribution of outcomes ranging from high to low achievement, standards seem to offer an alternative where all students have an opportunity to meet a defined standard.

Yet are these educational standards truly "objective"? Do they set appropriate levels of student achievement? And how are these standards being assessed in practice? All these related questions are of great importance if standards-based education is to be equitable and objective.

This article seeks to place the notion of objective standards under greater scrutiny. On the surface, if a set objective standard is the educational goal, then all students will have to opportunity to reach this standard. However, the use of standards for assessment is, in practice, problematic for reasons I will discuss here.

The Oxford dictionary defines objective as "*Not dependent on the mind for existence; actual:*" (Oxford, 2014). As such, objective is in contrast to subjective, where personal opinions are employed. This paper uses these definitions for the forthcoming analysis.

The following six principles illustrate the problems with standards claiming to be objective way of assessing student performance.

1) The selection of a standard is not objective.

The goal of selecting a standard is to produce a clear educational target for students. Whereas a given exam on a subject can be made more or less difficult, once a standard is set the target for assessment should be clear. However, the selection of this standard must involve human judgment. Here is a Common Core standard for kindergarten math, CCSS.Math.Content.K.CC.B.5:

> Count to answer "how many?" questions about as many as 20 things arranged in a line, a rectangular array, or a circle, or as many as 10 things in a scattered configuration; given a number from 1–20, count out that many objects. (http://www.corestandards.org/Math/Content/K/CC/B/5/)

Focusing on the last part of the standard, "given a number from 1–20, count out that many objects", this seems very clear and transparent. However, how was the range 1-20 determined as being appropriate for a kindergarten level? Why were other ranges, such as 1-5, 1-10, 1-21, or 1-30 not chosen? All of these alternatives would be equally as clear and transparent for students. These alternate standards have different levels of difficulty, but based only on the criteria of being objective, all these standards would suffice.

Here is another Common Core standard for kindergarten math, CCSS.Math.Content.K.CC.A.3:

> Write numbers from 0 to 20. Represent a number of objects with a written numeral 0-20 (with 0 representing a count of no objects).
> (http://www.corestandards.org/Math/Content/K/CC/A/3/)

Again, the target is clear – write the numbers from 0 to 20. Yet why was 0 – 20 chosen? Why not other ranges? This objective standard of writing 0 – 20 is just one of many potential ranges that could be chosen for a kindergarten-level standard. This range was subjectively chosen using human judgment.

The use of language is paramount in describing a standard. From one perspective, it is an "objective standard" while from another perspective it is "subjectively chosen". Hence, the standard is not truly objective in the literal sense of the word, i.e., that it is free from any potential bias. These mathematical standards were not directly based on facets of reality – they were subjectively chosen using human judgment.

2) *The selection of a standard depends on the ability of the students trying to meet the standard.*

As discussed, the Common Core Standard CCSS.Math.Content.K.CC.B.5 revolves around kindergarteners learning to count objects. At some point, human judgment was used to determine that counting from 0 – 20, but not other ranges, was the appropriate standard. However, this judgment must, at some level, be based on the potential ability of kindergarteners to reach this standard. For example, it could be put forth as a standard that kindergarteners should be able to perform calculus. Such a standard would be viewed as unrealistic because kindergarteners do not have the mental ability to perform calculus. For another example, if a fitness standard was set that kindergarteners should be able to run a 4-minute mile, this would also be viewed as absurd because kindergarteners do not have the physical ability to meet the standard. Hence, a standard is indirectly chosen based on the ability of the subjects to meet the standard. If the students do not have the ability, the standard is not viable or realistic.

In many cases, the judgment as to whether a standard is appropriate or not is the result of experience. If educators witness generations of kindergarteners where most students are able to

count up to 20 objects, then such a standard may seem reasonable. The standard is implicitly based on what students can do given their abilities at that point in time and reasonable effort.

The role of expectations should also be discussed here. Some argue that standards should be set as to stretch what students can do, and not just rely on their current level of performance. Research has shown that expectations can affect student performance (Muller, 1997). However, a standard must be within reach of the student to be effective. If a standard is beyond the ability of a student, it may in fact be demotivating to the student (Harlen & Crick, 2003).

3) Standards ignore the fact that academic ability varies across students.

Assume a standard is chosen that is realistic given the student body. If all students possess equal academic ability, and if all students then put in equal effort or time, then such a standard could indeed be fair. However, academic ability is not equal distributed among individuals.

A number of studies have showed that academic ability in various subjects tends to have a normal statistical distribution. For example, reading ability has a normal statistical distribution (Shaywitz et al, 1992). Mathematical ability has a normal statistical distribution (Docherty et al, 2010). General cognitive ability also has a normal statistical distribution (Plomin, 1999).

This variation in student ability directly affects educational performance. A recent high school study showed that 52% of the variation in English scores and 58% of the variation in Mathematics scores was due to heritability (Shakeshaft et al, 2013). Differences in academic ability are also predictive, meaning initial tests can directly predict later levels of academic achievement in subjects (Furnham, 2009).

In short, if it is known that student ability varies and explains a large portion of educational outcomes, why is one standard the appropriate measure for student outcomes? A mid-range standard might be easy for high ability students and impossible for low ability students, and thus it would only effectively serve student in the middle ranges. The only appropriate single standard given a normal distribution of abilities might be what the lowest-ability student could achieve with reasonable effort. Of course, such a standard could be passed by virtually all students with little effort and would most likely be perceived as one with low rigor.

4) Standards that reference "grade-level" materials are indirectly based on the abilities of students and/or rely on human judgment.

A number of the Common Core State Standards reference "grade-level" reading materials. For example, here is the Common Core standard CCSS.ELA-Literacy.RF.4.4a:

Read grade-level text with purpose and understanding.
(http://www.corestandards.org/ELA-Literacy/RF/4/4/a/)

As to what defines grade-level, the user is directed to Appendix A. In the appendix, it discusses how the approach to defining "grade-level" relies upon both qualitative, quantitative, and "reader and task" components (http://www.corestandards.org/assets/Appendix_A.pdf , Page 4) . "Reader and task" refers primarily to the student's motivation and interest in the text.

First, the qualitative part of defining a grade-level standard involves human judgment, and therefore may suffer from bias as discussed earlier in this paper. The qualitative component is explicitly defined as such relying on human judgment:

> Using qualitative measures of text complexity involves making an informed decision about the difficulty of a text in terms of one or more factors discernible to a human reader applying trained judgment to the task. In the Standards, qualitative measures, along with professional judgment in matching a text to reader and task, serve as a necessary complement and sometimes as a corrective to quantitative measures, which, as discussed below, cannot (at least at present) capture all of the elements that make a text easy or challenging to read and are not equally successful in rating the complexity of all categories of text. (http://www.corestandards.org/assets/Appendix_A.pdf , Page 5)

Several quantitative measures are discussed such as the Flesch-Kincaid Grade Level test and the Lexile framework. While these various formulas can calculate a score for a text based on objective factors (word count, etc.), assigning a grade level to these scores is based on the average performance of actual students in reading these texts.

From the Lexile website:

> Grade equivalents are scores based on the performance of students in the test's norming group. The grade equivalent represents the grade level and month of the typical (median) score for students. For example, a 5th-grade student who earns a 5.9 on a norm-referenced test has earned a score similar to the 50th percentile students in the test's norming group who were in their ninth month of fifth grade. Normative data are often collected at one point in the year from students in two or more grades. (https://www.lexile.com/about-lexile/grade-equivalent/)

It is clearly stated that Lexile grade levels are norm-referenced, not criterion-referenced. Therefore, the selection of grade-level materials by using this formula will be based on the average performance of students in a grade.

Grade-level measures, as defined by the Common Core, are then not truly objective in two ways. The qualitative measures involve human judgment and are subject to bias. The quantitative measures are based against average student performance, and therefore are norm-referenced.

5) The assessment of some standards is directly subjective.

The practical assessment of standards leads to other problematic issues regarding objectivity. Some of the earlier mentioned Common Core standards, such as counting 20 objects, should be relatively easy to score. However, many of the other standards implicitly rely on human judgment. Here is a Common Core Standard for English Language Arts, Grade 6, CCSS.ELA-Literacy.L.6.3:

> Use knowledge of language and its conventions when writing, speaking, reading, or listening. (http://www.corestandards.org/ELA-Literacy/L/6/3/)

Such a standard does not lend itself to the objective world of mathematics where answers can be definitively right or wrong. Grading under such a standard will be highly subjective, and it would be difficult if not impossible to make this grading consistent from classroom to classroom, much less from state to state. In practice, many teachers will be using their judgment alone for assessment creating even more potential for bias.

The sub-standard, CCSS.ELA-Literacy.L.6.3a, that follows with the main standard is also highly subjective:

> *Vary sentence patterns for meaning, reader/listener interest, and style.* (http://www.corestandards.org/ELA-Literacy/L/6/3/a/)

This sub-standard leads to many questions: how much variation? What audience? Whose definition of style? It is easy to see that these open-ended questions will lead to a wide variety of opinions, and hence a wide variability (Shavelson et al, 1993).

The validity of assessing non-absolute levels of student performance can be strengthened by using rubrics and other established procedures. However, the development of these rubrics and procedures involved human judgment and again do not represent something free of potential bias. Again, the use of language is paramount: an "objective" rubric has been subjectively developed using human expertise. As such, even though rubrics may reduce variability in assessment, they do not eliminate the problem of bias since they were created using human judgment.

6) *The "cut-off" score for meeting a standard across several questions is the result of a subjective process.*

With simple material, meeting a standard or not can be relatively clear. The previously discussed standard, CCSS.Math.Content.K.CC.A.3, essentially has students write the numbers from 0 to 20. This standard could be assessed by a single question or prompt. Yet as material gets more complex, it becomes more likely that several questions would be needed to assess a standard. For example, here is Common Core standard, CCSS.Math.Content.HSA.REI.B.3:

> Solve linear equations and inequalities in one variable, including equations with coefficients represented by letters.
> (http://www.corestandards.org/Math/Content/HSA/REI/B/3/)

This type of standard would most likely be assessed with a set of questions as opposed to a single question. So if multiple questions are used, what level of proficiency constitutes mastery of the standard? For example, if a student gets 7 out of 10 problems correct, has he or she achieved the standard? The answer to that question will involve human judgment.

Many agencies set "cut" scores for determining the pass/fail level. Yet setting any "cut" score to determine mastery or proficiency will necessarily involve subjective judgment. The only objective cut score is 100%. All other agency-created cut scores, even if they are developed from an established procedure, must ultimately rely on human judgment at some level. The procedures for setting cut scores were developed using human judgment, and therefore are not free of potential bias.

Conclusion

The aim of this article was to investigate the claim that educational standards are objective, and that, as such, they represent a superior alternative to the typical norm-referenced tools of assessment. Under further scrutiny, this does not appear to be the case for the reasons discussed. Standards are subjectively chosen by individuals and groups, and the chosen standards are implicitly based on the ability of the student body. The use of singular standards ignores differences in academic abilities.

Standards that focus on grade-level materials are also implicitly based upon the ability of the student body. The assessment of some standards, such as in reading or writing, will automatically involve human judgment. Finally, the setting of "cut" scores to determine pass/fail status involves human judgment at some level.

Given the preceding discussion, it is questionable as to whether standards-based assessment offers a legitimate alternative to typical norm-referenced assessment. Norm-referenced assessment can show how students compare to one another, but these results may not be connected to any outside reference point. Standards-based education attempts to connect to absolute criterion, but as this article shows there are many issues in that regard. A sampling of Common Core standards shows them to be subjectively chosen, implicitly based on student norms, and subjectively assessed in some subjects. In conclusion, the available evidence suggests that the creation and assessment of standards is not an objective process, but one that relies heavily on human judgment and average student performance.

References

Docherty, S., Davis, O., Kovas, Y., Meaburn, E., Dale, P., Petrill, S., Schalkwyk, S. & Plomin, R. (2010). A genome-wide association study identifies multiple loci associated with mathematics ability and disability. *Genes, Brain and Behavior* 9(2): 234–247.

Furnham, A., Monsen, J., & Ahmetoglu, G. (2009). Typical intellectual engagement, Big Five personality traits, approaches to learning and cognitive ability predictors of academic performance. *British Journal of Educational Psychology* 79(4): 769-82.

Harlen, W. & Crick, R.D. (2003). Testing and Motivation for Learning. *Assessment in Education, 10*(2): 169-207.

Muller, C. (2008). The Minimum Competency Exam Requirement, Teachers' and Students' Expectations and Academic Performance. *Social Psychology of Education*, 2(2): 199-216

Oxford dictionaries online. (2014). http://www.oxforddictionaries.com/

Plomin, R. (1999). Genetics and general cognitive ability. *Nature*, 402(6761 Suppl): C25-9.

Post, T., Harwell, M., Davis, J., Maeda, Y., Cutler, A., Andersen, E., Kahn, J., & Norman, K. (2008). "Standards"-Based Mathematics Curricula and Middle-Grades Students' Performance on Standardized Achievement Tests. *Journal for Research in Mathematics Education*, 39(2): 184-212.

Shakeshaft, N., Trzaskowski, M., McMillan, A., Rimfeld,K., Krapohl, E.,. Haworth, C., Dale, P., & Plomin, R. (2013). Strong Genetic Influence on a UK Nationwide Test of Educational Achievement at the End of Compulsory Education at Age 16. *PLoS ONE*, 8 (12): e80341 DOI: 10.1371/journal.pone.0080341.

Shavelson, R., Baxter, G., &Gao, X. (1993). Sampling Variability of Performance Assessments. *Journal of Educational Measurement*, 30(3): 215-232.

Shaywitz, S.E., Escobar, M.D., Shaywitz, B.A., Fletcher, J.M.,& Makuch R. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *New England Journal of Medicine* 326(3): 145-50.

Wilson, G., & Hoke, K. (1921). *How to Measure.* New York: The Macmillan Company.