

# Thoughts on Developing College and Career Ready Achievement Standards that are Internationally Benchmarked

---

*Marianne Perie, Center for Assessment. April 15, 2010*

## Purposes and Uses

Achievement standards are important mechanisms for communicating policymakers' goals. In the context of Common State Assessments, achievement standards exemplify in concrete form the goal of having all students graduate from high school prepared for college and career success. They can also be used to benchmark the performance of U.S. students taking these assessments to performance of students in other countries.

Achievement standards can be used as accountability measures, reporting tools, course-readiness indicators, and signals to educators about the level of each student's performance. They involve both written descriptors and cut scores and have more impact when released with items that depict performance expected in each level.

## Link to NIA

Section A.3.b.(ii) in the evaluation criteria states: *How the assessment system will produce the required student performance data (i.e., student achievement data and student growth data (both as defined in this notice) that can be used to determine whether individual students are college- and career-ready (as defined in this notice) or on track to being college- and career-ready (as defined in this notice)).* This paper is intended to address that issue of producing the performance data.

## Challenges:

1. NIA says states have to adopt achievement standards by the 2014-2015 school year.

NIA also says achievement standards must be data driven.

NIA also says the first operational year of the assessment should be no later than 2014-2015.

Therefore, standards may have to be set before the first operational year using data that are not yet available.

2. For the next generation assessments, policymakers intend to move away from traditional, judgment-based methods of setting cut scores towards more holistic data-driven methods incorporating information on student outcomes and convergent data from other sources.

We know how to do traditional standard settings using a judgmental approach such as Bookmark, Modified Angoff, etc. We are still learning how best to set standards using data driven methods, such as the Briefing Book method used by Achieve on the ADP Algebra II test.

Therefore, while we can sketch out a design now, some of the specifics will need to be determined by the level of data received and how it can best be organized and presented.

3. Currently, the average scores on international assessments (PIRLS, TIMSS, and PISA) of U.S. students compared to those in other participating countries are above average for elementary reading, average for high school reading, above average for elementary math, average for middle school math, and below average for high school math (Provasnik, Gonzales, & Miller, 2009).

Currently, approximately two-thirds of high school graduates go on to college and approximately 40% of those need remedial courses. So, approximately 40% of high school graduates enroll in college and need no remedial instruction.

Therefore, the internationally benchmarked standard that would lead us back to being a top performing nation would most likely be a much higher standard than the one needed to indicate college readiness at the current time. USED has talked about wanting standards that are both indicative of college and career readiness and internationally benchmarked, but those two goals are not necessarily compatible.

## Preferred Scenario

Each consortium would have at least three years of data on which to base their achievement standards.

Data might include:

- ACT/SAT scores
- IB/AP enrollment and completion data
- Graduation and degree statistics
- Percentage of students applying, being accepted, and enrolling in IHE
- Percentage of students enrolling in IHE who needed remediation
- Scores from certification tests and/or career placement tests
- Percentage of students applying, being offered, and taking a job out of high school
- ASVAB scores
- Percentage of students enlisting in the military after high school

Sufficient states in each consortium would participate in international assessments to receive state-level benchmarking scores.

- Next TIMSS administration: 2011
- Next PIRLS administration: 2011
- Next PISA administration: 2012

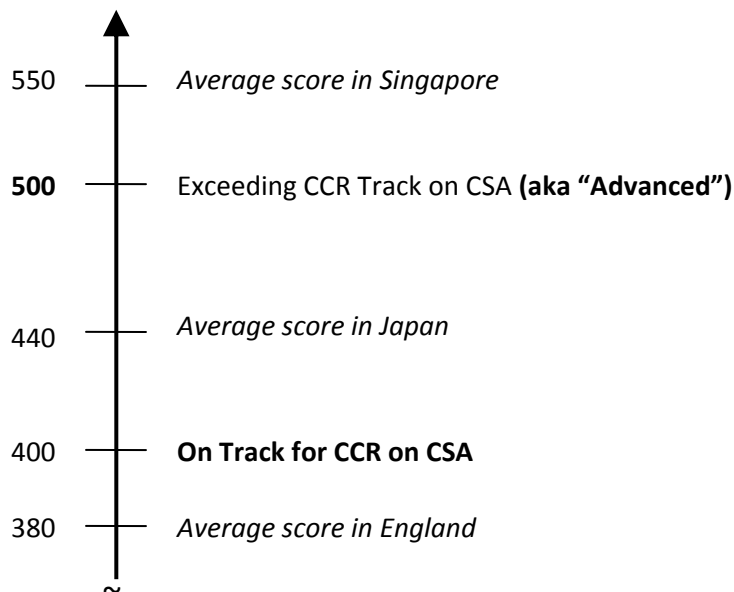
Ideally, we want to determine a data-based cut score that indicates readiness for college success, defined as preparedness for credit-bearing courses at a 2-year or 4-year institution without remediation. We also want a cut score that represents career readiness, defined as prepared to pass an entrance exam into a technical program or credentialing course or the readiness to enter a job training program in either the military or civilian sectors (including apprenticeship programs). Policymakers may decide to merge those cut scores into one college-and-career-ready cut score for all students, but for purposes of design, it is helpful to think of them separately.

While the data provide much information, judgments will need to be used to determine where to best put that cut score. For example, the data may tell us that 60% of all students receiving a score of 450 or above are college ready (as defined in the previous paragraph), and that 85% of all students receiving a score of 500 or higher are college ready, but to determine whether 450 or 500 (or somewhere in between) is the best cut score still involves human judgment. Moreover, when other indicators are added, such as enrollment in AP courses, correlation of scores with college entrance exam scores, etc., the judgment becomes more complex.

Once we can determine a college and career ready (CCR) standard for end-of-high-school in reading and math, we can regress those backwards to previous grades to set cut scores for “on-track” to college and career readiness. The growth model used in Colorado (Betebenner, 2009) provides an example of how those “on-track” standards can be set.

Superimposed on these CCR standards should be benchmarks comparing different levels of performance on the CSA to average performance in other countries – particularly high performing countries and countries with economies the size of the United States and that might be seen as economic and scientific competitors. For example, Figure 1 shows a possible chart of fourth-grade math performance.

**FIGURE 1: REPORTING SCALE FOR GRADE 4 MATHEMATICS ON COMMON SUMMATIVE ASSESSMENT**



Again, in an ideal scenario, this type of chart would be refined periodically as new data are gathered showing how students are progressing through learning math from third through eighth grade (which would increase the predictive function for “on track”) and with the release of new TIMSS data.

## Realistic Scenario

We will only have some of those data

Interim cut scores will need to be set at the end of the 2013-2014 school year to give each state time to adopt them. The performance level descriptors describing what is meant by college and career ready, on track, off track, exceeding track, etc. will need to be written earlier. Again, these descriptors will most likely be placeholders until we have more longitudinal data allowing us to write more complete descriptors.

States will need to collect data on student performance on the CSA and other indicators listed above over the next several years. That is, data systems will need to be created and maintained that track student performance on the CSA each grade, other assessments taken and the scores, credential(s) received at the end of high school, and information on application, acceptance, and enrollment in IHEs, career training programs, the military, or acceptance of a job.

A new achievement standard setting would need to be conducted within four years based on actual student data

Challenge: what if new standards are substantially different from interim standards? For instance, what it means to be CCR could change and information linking scores to success will become richer, but how will that be reflected in the full assessment and accountability system? Will accountability system support such a change or will some type of equipercentile smoothing need to be applied to maintain continuity of system? Tension between goals for students and needs for fairness toward those being held responsible, particularly if accountability system continues to be punitive. Some type of phase-in of high-level CCR achievement standards may be needed.

## Interim Approach

Look at the data we do or could have in addition to CSA results:

CCR:

- ACT/SAT scores
- IB/AP enrollment and completion data
- Scores from certification tests and/or career placement tests
- ASVAB scores

International benchmarking

- 2011 TIMSS scores available in 2012
- 2011 PIRLS scores available in 2012
- 2012 PISA scores available in 2013

## Other

- NAEP scores from 2011 available in 2012

We can certainly create tables linking average state scores on the CSA to their relative achievement on NAEP and conduct an “equating” such as that done by Braun & Qian (2007). This will provide information allowing us to set a cut score on the CSA comparable to the Proficient cut score on NAEP, if that is desirable.

In addition, if enough states participate in PIRLS, PISA, and TIMSS, allowing their students to be oversampled so that a state score can be generated, we can also benchmark scores on the CSA to scores on international assessments for the select grades in reading and math assessed by these measures. We should then be able to produce an NCES-type table that shows that a score of X on the grade 4 math CSA is roughly equivalent to the average score in Finland. This is a step state policymakers can start looking into now to determine what is required for full state participation in each of the international assessments.

We can use these types of data to inform the cut scores of grades 4, 8, and 10 (15-year olds in PISA). For end-of-high-school, we can also look at the range of scores of students who have enrolled in a passed at least one AP/IB course as an example of how to think about “advanced”. We can create concordance tables for scores on the CSA high school math and SAT math and ACT math and do the same for English/reading.

Overall, we can take the data from each of these three sources to help us set initial cut scores. We need at least two solid data points to create growth trajectories, so if we only have one year of data, we will need to set cut scores at certain years (e.g., 4, 8, and high school) and draw a line between them. That line can then be adjusted as we gather growth data over the years to show the “on-track” line indicating growth towards CCR.

Thus, performance level descriptors used for standard setting at this point should only reflect the extant data. For example, advanced performance might include reference to students taking and passing AP courses, but for now, the desire to have that level reflect students attending four-year colleges fully prepared for college credit will need to be de-emphasized until we have the data to support that claim. Cut scores can be set for grade levels in which we have external data to use in a standard-setting study, and the cut scores for the interim grades can be interpolated from those.

Caution will need to be taken in how these scores are publicized, described, and used these scores since we know they will be adjusted, and perhaps changed fairly substantially, as we collect more data on the relationship between CSA scores and college and career success.

## References

Betebenner, D. (2009). Norm- and criterion-referenced student growth. *Educational Assessment: Issues and Practices*, 48 (4),pp. 42-51.

Braun, H. I. and Qian, J. (2007). *Mapping state performance standards onto the NAEP scale*. Princeton, NJ: Educational Testing Service.

Provasnik, S., Gonzales, P., & Miller, D. (2009). *U.S. Performance across International Assessments of Student Achievement: Special Supplement to The Condition of Education 2009*. Washington, DC: U.S. Department of Education, National Center for Educational Statistics.

---

This paper is one of a series of brief documents produced by the Center for Assessment designed to assist state consortium leaders in thinking about and preparing their responses to the United States Department of Education Notice Inviting Applications for Comprehensive and High School Assessment System grants. Funding has been provided by the Bill & Melinda Gates Foundation. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Gates Foundation.