

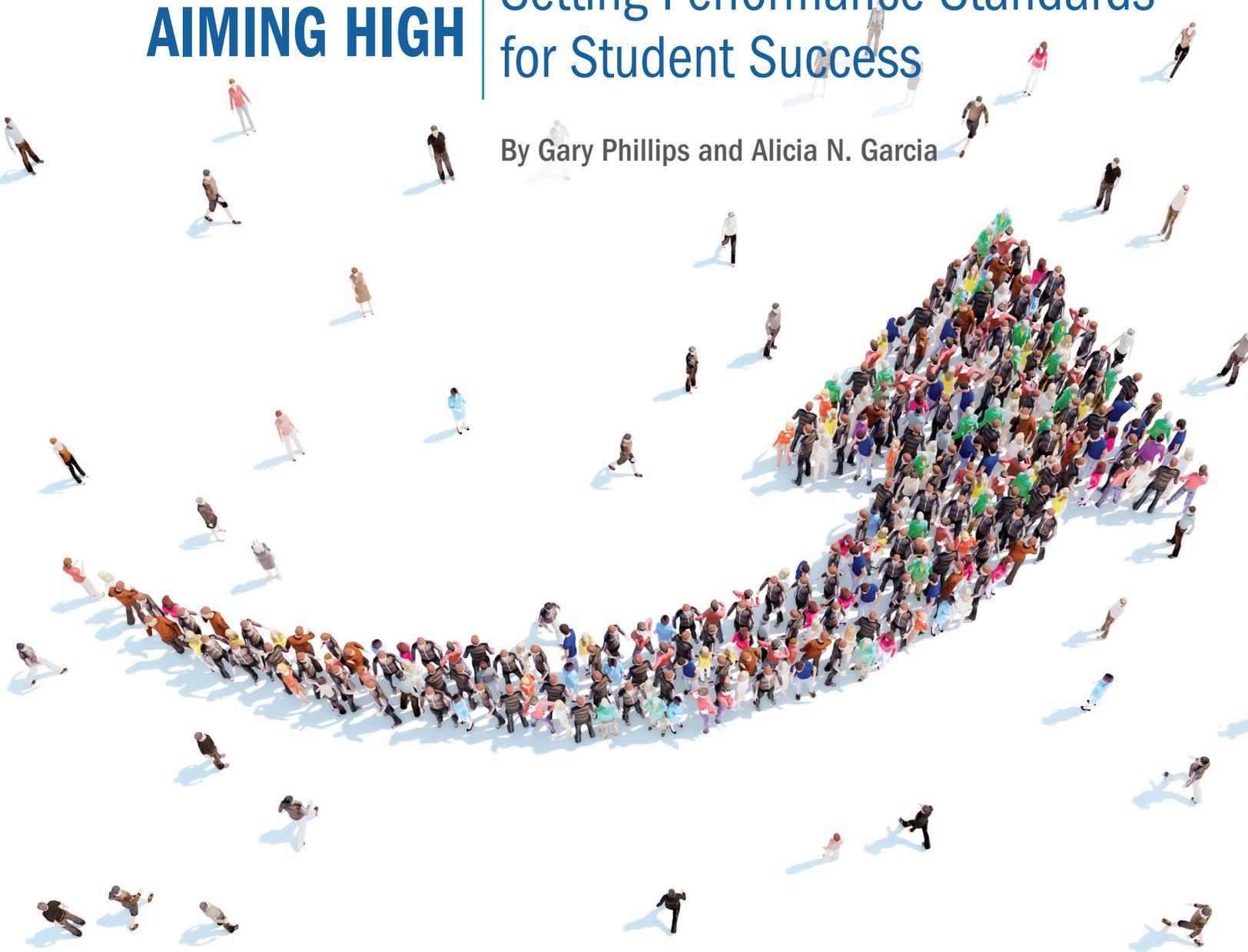
EDUCATION POLICY Center

at American Institutes for Research ■

AIMING HIGH

Setting Performance Standards for Student Success

By Gary Phillips and Alicia N. Garcia



AMERICAN INSTITUTES FOR RESEARCH®

FEBRUARY 2015

The **Education Policy Center at American Institutes for Research (AIR)**

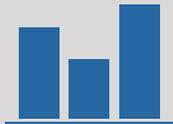
provides rigorous research- and evidence-based perspectives on education issues spanning prekindergarten to careers, including reports, briefs, legislative guides, and our InformED blog—all written by AIR experts. Visit our site at www.edupolicycenter.org regularly for current information on how research and practice can provide much-needed evidence to inform your policy decisions.

Setting Performance Standards for Student Success



THE ISSUE

Content standards, not performance standards, have been almost the sole focus of state policies and recent conversations about academic standards. Although the movement to adopt rigorous education *content* standards is evidence that states are motivated to raise academic expectations, current *performance* standards do not give accurate measures of student achievement. Without rigorous content and performance standards, we cannot adequately prepare students for the global marketplace.



THE RESEARCH

A recent American Institutes for Research (AIR) study shows that state performance standards are consistently low and extremely variable across states. Low state standards allow many students to be described as “proficient,” when they are not adequately prepared for success beyond high school. In addition, the variability across states results in students in states with the lowest standards performing three to four grades levels below those in states with higher standards.



THE RECOMMENDATION

States should use evidence-based methods of standard setting, such as the *benchmark method*, to create and adopt rigorous performance standards that prepare students to compete in the global marketplace.

Since the release of the Common Core State Standards four years ago, policy debates have raged across the nation about the adoption of state education content standards. Although the conversation surrounding the Common Core State Standards has been contentious, policymakers and stakeholders consistently agree that each state should have rigorous content standards that will prepare students for postsecondary success. However, these debates have failed to address an important piece of the standards' conversation. To ensure that students are prepared for the future, states must also adopt improved performance standards. If states adopt rigorous *content* standards but retain low *performance* standards, the number of students identified as "proficient" will give a false picture of the nation's progress toward educational excellence in the global marketplace.

To ensure that students are prepared for the future, states must also adopt improved performance standards.

WHAT IS THE DIFFERENCE BETWEEN CONTENT STANDARDS AND PERFORMANCE STANDARDS?

CONTENT STANDARDS	PERFORMANCE STANDARDS
<p>Content standards outline the scope and sequence of the academic content a student is expected to learn at each grade level—WHAT WE EXPECT STUDENTS TO KNOW when they leave the classroom at the end of the school year.</p>	<p>Performance standards measure a student's progress toward learning that content—HOW MUCH of the academic content do we expect a student to know to be considered proficient. States use performance standards to monitor yearly progress and the success of each classroom, school, and district.</p>



THE ISSUE

States Focus on Content Standards

The movement of states to adopt new, rigorous content standards is part of an effort to raise academic expectations for American students. An examination of education content standards in the top-performing nations reveals that these standards share three common characteristics: focus, rigor, and coherence (National Governors Association [NGA], Council of Chief State School Officers [CCSSO], & Achieve, 2008).

- ▶ Focus in education standards requires covering fewer topics at each grade level but going into greater depth on each of those topics.
- ▶ Rigor requires introducing high-level topics at an early grade level (e.g., Algebra in the eighth grade).
- ▶ Coherence requires laying out an orderly progression of standards rather than repeating topics across grades (NGA, CCSSO, & Achieve, 2008).

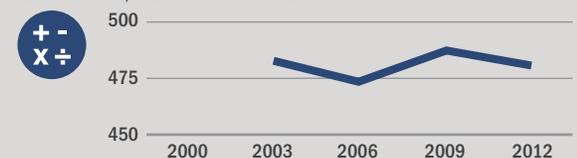
Prior to 2008, these characteristics were not often included in state education content standards throughout the United States. The addition of these characteristics is expected to improve U.S. student achievement to match that of the current top-performing countries.

Since 2008, all 50 states have adopted new content standards for English language arts and mathematics. As part of this movement, 40 states have adopted the new Common Core State Standards, which were

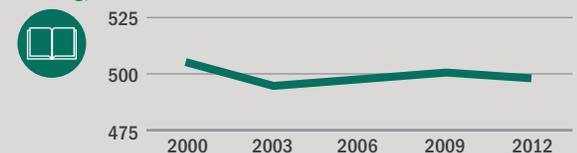
WHY ARE RIGOROUS PERFORMANCE STANDARDS SO IMPORTANT?

Performance scores on the Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS) show that American students lag behind students from other countries in the international marketplace. According to analysis of 2012 PISA results, U.S. students performed below average in mathematics and about average in reading and science (Organisation for Economic Co-operation and Development, 2013) compared to students in 33 other countries. More disconcerting, however, is that American students have not shown significant changes in mathematics, reading, or science since 2007 (Provasnik et al., 2012). These trends suggest that **American students are not acquiring the foundational skills needed to compete in the global marketplace.**

Mathematics, United States Mean Score



Reading, United States Mean Score

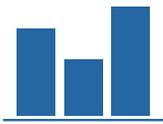


Science, United States Mean Score



released in June 2010 by the National Governors Association Center for Best Practices and the Council of Chief State School Officers. The standards include the three characteristics of the standards of top-performing nations and were designed to prepare students to graduate from high school ready to succeed in college courses and workforce training.

The movement to new standards has not been without controversy, but policymakers and stakeholders consistently agree that states need rigorous college- and career-ready standards. The increased expectations of these rigorous content standards are expected to improve student achievement. However, a recent [AIR study](#) found that most states' current *performance* standards are so low that they convey a false sense of student proficiency. Without an accurate measure of student knowledge, the adoption of rigorous content standards will not be enough to affect student achievement in a meaningful way because many students will be described as "proficient" when they are not adequately prepared for postsecondary success.



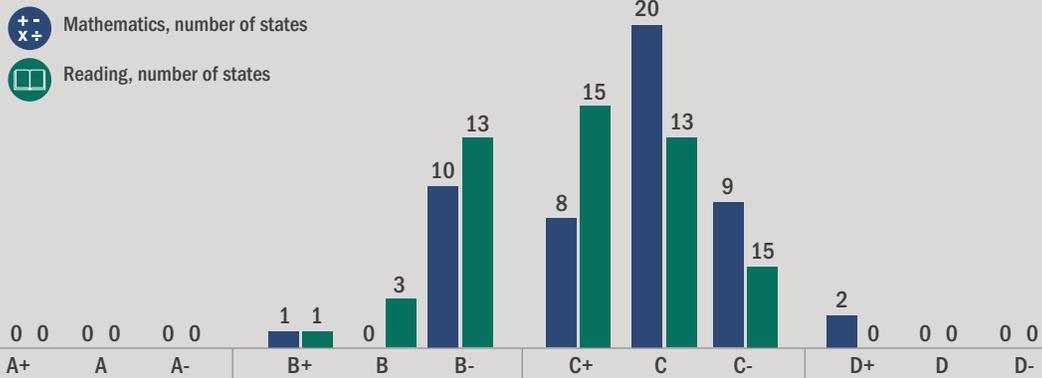
THE RESEARCH

Current State Performance Standards Are Low and Variable

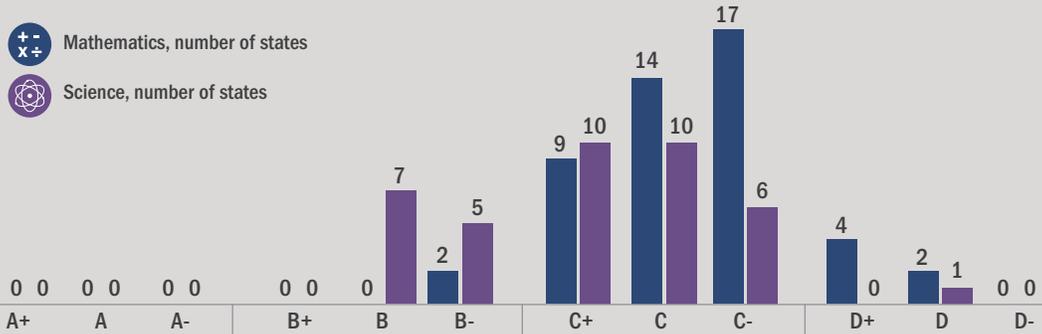
In his study, [International Benchmarking: State and National Education Performance Standards](#), AIR's Gary Phillips converted performance standards for 50 states and the District of Columbia to the metrics of Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Literacy Study (PIRLS). He assigned a letter grade to each state's performance standard. Phillips found that many states are using low standards to define proficiency (e.g., seven states require only the equivalent of a D or D+ for students to be considered proficient in eighth grade mathematics). Lower performance standards allow states to report higher student proficiency rates, but those proficiency rates overestimate real student achievement.

Phillips found that the difference in the rigor of performance standards between the state with the highest standards and the state with the lowest standards was three to four grade levels; the proficiency of an eighth grader in a low-performance standard state was equal to that of a fourth grader in a high-performance standard state. In statistical terms, this

International Benchmark Grade for Mathematics and Reading, Grade 4



International Benchmark Grade for Mathematics and Science, Grade 8



variation is approximately two standard deviations—more than twice the size of the national black-white achievement gap. “These wide variations and low standards bespeak a lack of credibility and lack of transparency in state and federal education reporting, confuse policymakers, and mislead the public in some states into believing that their students are proficient when they are not” (Phillips, 2014, p. 16). Decreasing the variation of performance standards among states is an important step in accurately assessing American student achievement.



THE RECOMMENDATION

States Should Use Evidence-Based Approaches to Standard Setting, Such As the Benchmark Method, to Create Performance Standards

Current approaches to developing performance standards “set standards based primarily (and in some cases exclusively) on test content” (Phillips, 2014, p. 16). This approach is often characterized by a three-step process. Each state does the following:

1. Develops content standards
2. Develops performance-level descriptors (i.e., statements about how much of the content standards students should learn)
3. Establishes performance standards (i.e., cut-scores on the test scale) that represent degrees of proficiency (Phillips, 2010)

Usually, these performance standards are believed to be rigorous and are recommended by a broad range of education stakeholders, but this approach does not include a process to ensure that the standards provide an accurate measure of how students compare to those in other states or other countries (Phillips, 2014).

To increase comparability, Phillips recommends that states use evidence-based methods of standard setting, such as the *benchmark method*, to set performance standards.

WHAT IS THE ROLE OF ASSESSMENTS IN PERFORMANCE STANDARDS?

Assessments used to measure student proficiency are a key component in setting high performance standards. Each state defines proficiency by identifying a particular “cut score” a student must achieve on the assessment used by the state (e.g., a student must respond correctly to 70 percent of the test questions to be considered proficient). If states use different assessments, it is difficult to determine whether the same cut score indicates the same level of performance for students across states. The use of a common assessment, such as those under development by the two Common Core assessment consortia (Partnership of Assessment of Readiness for College and Career and Smarter Balanced Assessment Consortium) can allow more accurate comparison of student proficiency across states.

Here is how states can use this five-step process:

1. Develop a pool of test items that align with state content standards.
2. Assemble those items into an ordered-item test scale¹ that covers the content standards.
3. Link the test scale to national and/or international test scales.
4. Use these statistical links to develop descriptions of what students are expected to know and master on the state test for each performance level—from “not proficient” to “highly proficient.” These descriptors should be established by content specialists and must be *comparable in rigor* to the performance standard for the national or international test.
5. Arrange for a panel of education stakeholders to review the performance descriptors in conjunction with the state’s content standards to make recommendations on where to set the cut-scores for the test (Phillips, 2010).

These last three steps mark a significant change in the current approach used to set performance standards. Using this benchmark method rather than an internal content focus results in state performance standards that are “consistent and more on par with the high standards used by national and international assessments” (Phillips, 2014, p. 17).

SBAC ADOPTS EVIDENCE-BASED PERFORMANCE STANDARDS.

Smarter Balanced Assessment Consortium recently set its new performance standards using an evidence-based method. The process lasted a little more than three months and **involved multiple steps**. First, members of the public reviewed the test items online and rated their difficulty levels. Second, 500 reviewers (classroom teachers, principals, curriculum directors, and higher education faculty) met to examine the items, which were arranged in order of difficulty, to decide the points that distinguished four levels of achievement. Third, the reviewers’ results were aggregated into cut scores for each grade and subject and **compared to performance data from national tests, such as the National Assessment of Educational Progress and the ACT**. Finally, these scores were reviewed across grade levels for consistency before being presented to the states for adoption. The scores were approved by the member states in November 2014 (Gewertz, 2014).

¹ An ordered-item test scale orders the items based on difficulty level (easy to hard), which makes it easier for states to establish statistical linkages to other test scales.

Implementing the Recommendation: Considerations for Policymakers

Policymakers may face a number of challenges in adopting rigorous performance standards:

- ▶ **Political Discord.** States may still be struggling with the adoption of new content standards. The adoption of new performance standards may increase political discord among education stakeholders within the state.
- ▶ **Stakeholder Pushback.** Switching to an evidence-based method of standards setting, such as the benchmark method to determine performance standards, is a significant shift from the approach states are currently using. Policymakers may receive pushback from stakeholders.
- ▶ **Public Concern.** Higher performance standards are likely to result in a dramatic increase in the percentage of students not meeting proficiency.² Such a dramatic increase will likely result in public concern within the state.

As policymakers move to adopt more rigorous performance standards, they should focus on educating practitioners and the general public about the reasons for the change. Communication and dissemination of information about the need for more rigorous performance standards, and the benefits to American students, will help alleviate some of the expected concerns. Despite these challenges, it is important that policymakers consider adopting rigorous content *and performance* standards in order to prepare American students to compete in the global economy.



LESSONS LEARNED | Policymakers in states that have already adopted rigorous performance standards have prepared their constituents for the anticipated drop in test scores by communicating about the change months in advance. States looking for examples of communications to use should review the efforts by New York and Kentucky. Both states offer examples of incorporating statewide communication plans regarding expected results and the reasons for the changes both preceding and following the release of new test scores.

² As described in a recent *Education Week* article, under the new proficiency scores approved by the SBAC “41 percent of 11th graders will show proficiency in English/language arts, and 33 percent will do so in math” (Gewertz, 2014).

Want More Information About Performance Standards?

Read the full Phillips report, [International Benchmarking: State and National Education Performance Standards](#)

References

- Gewertz, C. (2014, November 17). Cutoff scores set for common-core tests. *Education Week*. Retrieved from <http://www.edweek.org/ew/articles/2014/11/17/13sbac.h34.html>
- National Governors Association, Council of Chief State School Officers, & Achieve, Inc. (2008). *Benchmarking for success: Ensuring U.S. students receive a world-class education*. Washington, DC: National Governors Association.
- Organisation for Economic Co-operation and Development. (2013). *Lessons from PISA 2012 for the United States: Strong performers and successful reformers in education*. Organisation for Economic Co-operation and Development Publishing.
- Phillips, G. (2010). *International benchmarking: State education performance standards*. Washington, DC: American Institutes for Research.
- Phillips, G. (2014). *International benchmarking: State and national education performance standards*. Washington, DC: American Institutes for Research.
- Provasnik, S., Kastberg, D., Ferraro, D., Lemanski, N., Roey, S., & Jenkins, F. (2012). *Highlights from TIMSS 2011: Mathematics and science achievement of U.S. fourth- and eighth-grade students in an international context* (NCES 2013-009 Revised). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.



1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
800.356.2735 | 202.403.5000

www.air.org
www.edupolicycenter.org