

Do Graduation Tests Measure Up?

A Closer Look at State High School Exit Exams

 The American
Diploma Project
An initiative of


Achieve, Inc.

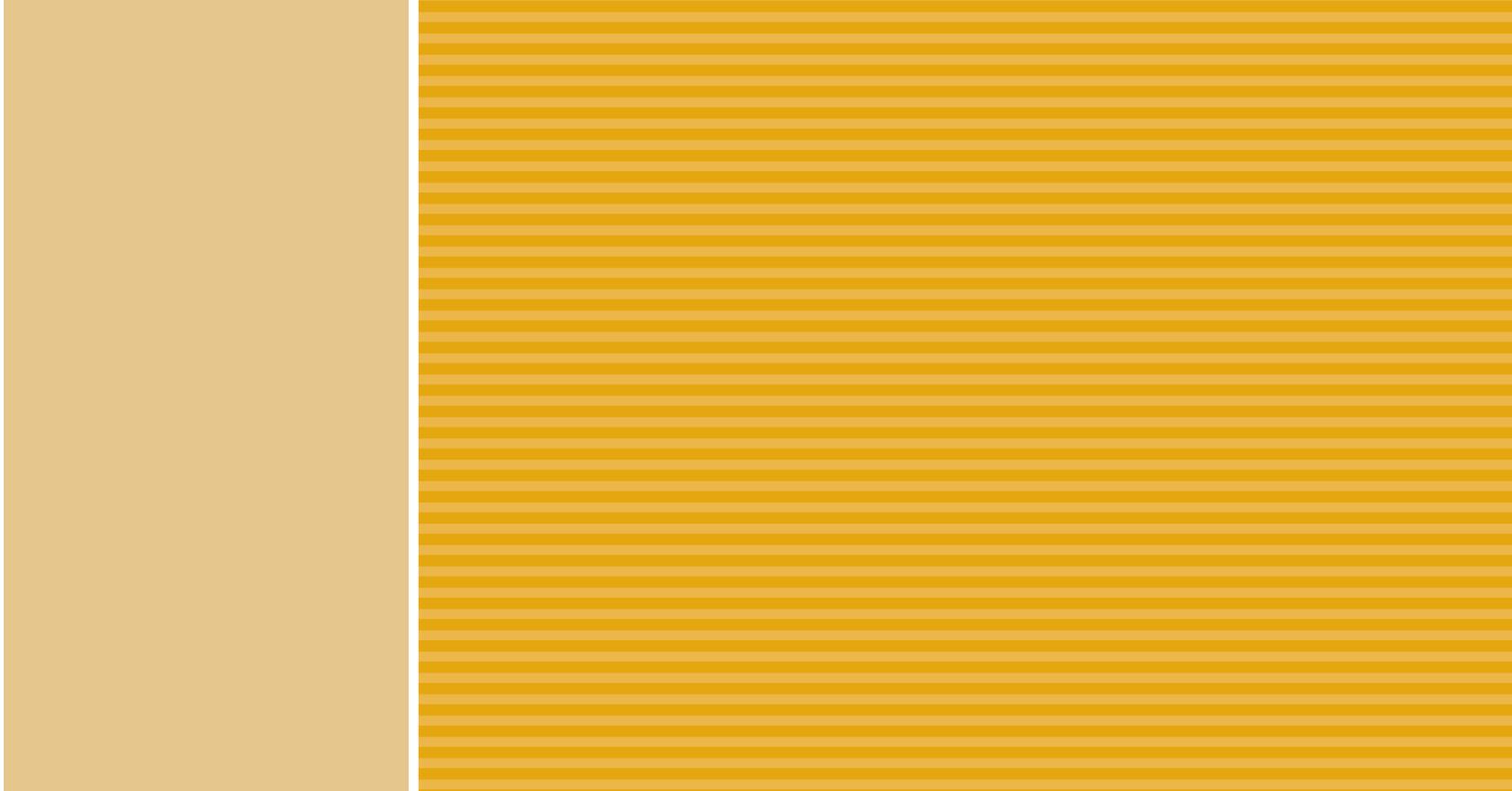
About Achieve, Inc.

Created by the nation's governors and business leaders, Achieve, Inc., is a bipartisan, non-profit organization that helps states raise academic standards, improve assessments and strengthen accountability to prepare all young people for postsecondary education, work and citizenship. Achieve has helped nearly half the states benchmark their standards and tests against the best examples in this country and abroad and work in partnership to improve teaching and learning. Achieve serves as a significant national voice for quality in standards-based reform and regularly convenes governors, CEOs and other influential leaders at National Education Summits and other gatherings to sustain support for higher standards and achievement for all of America's schoolchildren.

Copyright © 2004 Achieve, Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information or storage retrieval system, without permission from Achieve, Inc.

Editorial assistance and design: KSA-Plus Communications, Inc.



Do Graduation Tests Measure Up?

A Closer Look at State High School Exit Exams

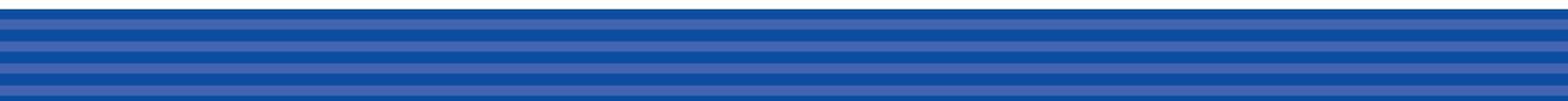




Table of Contents

Executive Summary	1
Introduction	4
Findings	10
Overview of Methodology <i>(10)</i>	
What Do the Mathematics Exams Measure? <i>(12)</i>	
What Do the English Language Arts Exams Measure? <i>(19)</i>	
How Difficult Is It To Pass the Tests? <i>(27)</i>	
Recommendations for States	30
Don't Lower the Standards or Delay Implementation <i>(30)</i>	
Continue Improving the Tests Over Time <i>(31)</i>	
Build More Comprehensive Assessment Systems <i>(33)</i>	
Appendix: Summary of Methodology	36
Acknowledgments	44

Executive Summary

High school graduation exams are in place in nearly half the states, and more than half the nation's high school students have to pass them to earn a diploma. There have been countless commentaries about the fairness of such tests, but very little research into what the tests actually measure and whether those expectations are reasonable. With the participation of six states, which together enroll nearly a quarter of the nation's high school students, Achieve set out to determine how high a bar high school graduation exams set for students.

After a detailed analysis of the mathematics and English language arts exams in Florida, Maryland, Massachusetts, New Jersey, Ohio and Texas, Achieve reached three conclusions: *First*, it is perfectly reasonable to expect high school graduates to pass these tests — they are not overly demanding. *Second*, these exams will need to be strengthened over time to better measure the knowledge and skills high school graduates need to succeed in the real world. *Third*, states should not rely exclusively on these tests to measure everything that matters in a young person's education. Over time, states will need to develop a more comprehensive set of measures beyond on-demand graduation tests.

Tests Are Not Overly Demanding

None of the tests Achieve reviewed in this study present unreasonable expectations for high school graduates. On the contrary, the tests cover material that most students study by early in their high school careers. Given where the bar is set, it is perfectly reasonable for states to require students to pass these exams to earn a high school diploma.

- **The questions on these exams reflect material that most students study by early in their high school careers.** In math, the tests place a heavier emphasis on prealgebra and basic geometry and measurement concepts than on concepts associated with Algebra I and high school geometry. In English language arts, the exams are a better measure of basic reading comprehension skills than the more advanced reading skills students will need to succeed in college and the new economy.
- **The “cut scores” required to pass the tests reflect modest expectations.** To pass the math tests, students in these states need to successfully answer questions that, on average, cover material students in most other countries study in 7th or 8th grade. To pass the English language arts tests, students need to successfully answer questions that ACT considers more appropriate for the test it gives to 8th and 9th graders than its college admissions test.

- **The tests measure only a fraction of the knowledge and skills that colleges and employers say are essential.** The exit exams in these six states measure some of the skills essential for college and workplace success, but a significant number of those skills go largely unmeasured. The skills that do get measured are staples; students cannot succeed without them. But the large gap between these tests and the real-world expectations of colleges and employers suggests that current exit exams are not strong measures of college and workplace readiness.

Tests Need To Be Strengthened Over Time

The exit exams reviewed in this study are considerably more challenging than the exams these states once used. In Ohio, for example, the new math test places greater emphasis on algebra than the previous proficiency test did. In Maryland, algebra and geometry were not even measured on the previous graduation test. These states — and others around the country — have raised the bar to better keep pace with the new realities that high school graduates face. The Achieve analysis reveals that the bar needs to be raised even higher over time. Achieve recommends that states:

- **Emphasize more challenging content.** In math, states need to increase the rigor of the algebra and geometry on the tests; in English language arts, states need to focus on more advanced reading skills and increase the sophistication of the reading passages on the tests. Reading tests also should give greater attention to informational topics, which currently receive less emphasis than literary topics. Writing tests should require students to write more and rely less heavily on multiple-choice items.
- **Ask more challenging questions.** In both reading and math, a majority of the test questions tap basic skills — routine procedures in math and inference and literal recall in English language arts. States should work over time to ensure that a larger percentage of assessment items measure higher-level skills, such as mathematical reasoning and critical reading.
- **Phase in higher cut scores.** One way states can phase in more challenging exams over time is by raising the score required for passing. Texas is using this approach with its new graduation exam. This strategy will work only if the exams have enough range in what they measure, so that a higher score actually reflects more advanced knowledge and skills. If a higher cut score simply means that students can answer more of the same kinds of questions correctly, rather than questions tapping advanced concepts and skills, it is not very meaningful to raise the cut score.

Graduation Exams Cannot Measure Everything That Matters

Basic fairness requires that students have multiple opportunities to take high school exit exams, so it is reasonable for states to begin to administer the tests in the 10th or 11th grades. Ultimately, however, it is important for 12th grade students to be able to do 12th grade work, not just pass a 10th or 11th grade test. Over time, states will need to develop a more comprehensive set of measures beyond on-demand graduation tests.

States could, for example, develop 12th grade assessments that are well aligned to college and workplace knowledge and skills. Another strategy could be developing end-of-course exams for subjects such as Algebra II or upper-level English that are beyond the range of the exit exams.

Rather than attaching “high stakes” to these tests, the scores might be factored into course grades or simply included on high school transcripts. This would provide valuable information that postsecondary institutions and employers could use in making admissions, placement or hiring decisions.

States also will need to move beyond large-scale assessments because, as critical as they are, they cannot measure everything that matters in a young person’s education. The ability to make effective oral arguments and conduct significant research projects are considered essential skills by both employers and postsecondary educators, but these skills are very difficult to assess on a paper-and-pencil test. States should work with local districts to develop ways to incorporate research projects and oral examinations into instructional programs and to establish rigorous, systematic criteria for evaluating them across the state.



Achieve launched this study to help answer some basic questions about the expectations states are setting for their high school graduates through the use of exit exams. Do the tests reflect material that students should be familiar with by the time they complete high school? Is it reasonable to expect all students to pass these tests before they graduate? If they pass these tests, does it mean students are ready for their next steps in life?

The tests do indeed set a floor for students that states can responsibly defend as a graduation requirement. In states where the exit exams are being debated, Achieve strongly encourages policymakers not to lower the standards or delay implementation. If states stay the course with these exams and make the necessary investments to improve teaching and learning, they undoubtedly will find that their students will rise to the challenge. When sufficient numbers of students pass these tests, states should continue to raise the floor to reflect the demands students will face in postsecondary education and the world of work.

How High Is the Bar?

For the most part, the public debate over high school exit exams has occurred without vital information about how high a hurdle they actually set in front of high school students. There have been countless commentaries about the fairness of such tests, but very little research into what the tests actually measure and whether those expectations are reasonable. With the participation of six states, which together enroll nearly a quarter of the nation's high school students, Achieve set out to answer three fundamental questions about high school graduation exams:

- What do the tests actually measure?
- What does it take for students to pass the tests?
- How well do the tests measure what postsecondary educators and employers say first-year students and new employees must know and be able to do?

The states that participated in this study — Florida, Maryland, Massachusetts, New Jersey, Ohio and Texas — exhibited the courage and national leadership to open their exit exams to unprecedented independent scrutiny. From the outset, the states knew that Achieve intended to publish its findings widely to inform policymakers and others across the country in their discussions about high school exit expectations. In large measure, the states agreed to take part in the analysis to better understand how their expectations for high school graduates compare to one another and to gain valuable insights into how they can improve their exams — and their broader high school policies — to ensure that students graduate well prepared for the demands of postsecondary life.

The six states have made different policy choices around the timing of their exit exams. Florida, Massachusetts and Ohio each give their tests for the first time to 10th graders in the second half of the school year. New Jersey and Texas give their exit exams to 11th graders. Maryland has opted for a battery of end-of-course exams, with the English language arts exam given as early as the end of 9th grade. While these differences represent important context, in the end the tests — whenever they are given and whatever their format — represent each state's bottom-line expectations for what high school graduates must know and be able to do. Achieve chose to compare them based on this fundamental similarity.

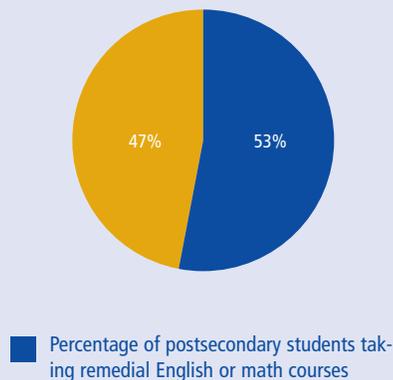
STATES PARTICIPATING IN GRADUATION EXAM STUDY

Florida	Maryland	Massachusetts	New Jersey	Ohio	Texas
TEST					
Florida Comprehensive Assessment Test	High School Assessments	Massachusetts Comprehensive Assessment System	High School Proficiency Assessment	Ohio Graduation Tests	Texas Assessment of Knowledge and Skills
GRADE FIRST GIVEN					
10th	End of Course	10th	11th	10th	11th
YEAR FIRST GIVEN					
1998	2001	1998	2002	2003	2003
REPLACED ANOTHER EXIT TEST					
Yes	Yes	No	Yes	Yes	Yes
SUBJECTS TESTED FOR GRADUATION REQUIREMENTS					
Reading, math	English I, Algebra/Data Analysis, Biology and Government	English language arts and math	Reading, writing and math	Reading, writing, math, science and social studies	English language arts, math, science and social studies
FIRST GRADUATING CLASS REQUIRED TO PASS					
2003	2009*	2003	2003	2007	2005
OPPORTUNITIES FOR STUDENTS WHO HAVE NOT PASSED TO RETAKE TESTS					
Yes	Yes	Yes	Yes	Yes	Yes
OTHER POLICIES RELATED TO STAKES					
Classes of 2003, 2004 permitted to substitute results on college admissions test.	Proposal is being considered to allow students to fail one of four tests and still graduate with cumulative score across tests. <i>* State Board of Education approval pending</i>	Appeals process uses statistical comparison of GPAs in subject area courses of passing and non-passing students.	State has alternative, performance-based assessment given and scored locally. Sixteen percent of class of 2003 statewide and up to 50 percent in some districts used this route to graduate.	State law allows students to fail one of five tests and still graduate if score is close to passing mark and GPA in subject is at least 2.5.	Passing score for first two graduating classes was lower than eventual passing mark.

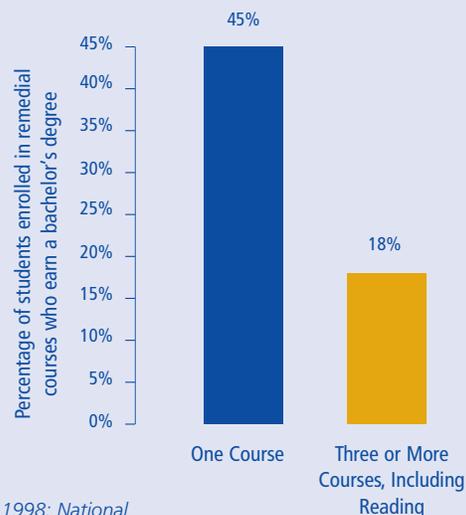
Why High School Exit Exams Are Important

While some people question the use of tests as part of states' graduation requirements, the *need* for higher high school exit expectations should be clear because the *cost* of poor high school preparation to students and states is so clear. While roughly three-quarters of high school graduates pursue postsecondary education within two years of earning a diploma, fewer than half ever earn a degree.¹ Twenty-eight percent of those entering two- and four-year colleges immediately need a remedial English language arts or mathematics course,² and half of all students take at least one remedial course during their college careers.³ For their part, more than 60 percent of employers rate high school graduates' grammar, spelling, writing and basic math skills as only "fair" or "poor."⁴ The cost for one state's employers to offer remedial English language arts and math was estimated at about \$40 million a year.⁵

More Than Half of All College Students Take Remedial Courses ...



... And Students Who Take Remedial Courses Are Less Likely To Finish College



Source: National Center for Public Policy and Higher Education, 1998; National Center for Education Statistics, 2001.

High school exit exams allow states to set a "floor" of academic achievement for all students to ensure that they meet standards before earning a diploma. Such exams provide evidence of achievement that is more comparable and aligned with standards than Carnegie units or course grades alone. Most important, like all high-quality, standards-based tests, exit exams can help schools focus teaching and learning on essential knowledge and skills while targeting extra attention and resources to those students who need them the most.

¹National Center for Education Statistics, *Access to Postsecondary Education for 1992 High School Graduates*, 1997.

²National Center for Education Statistics, *Remedial Education at Degree-Granting Postsecondary Institutions in Fall 2000*, 2003.

³National Center for Education Statistics, *Condition of Education 2001*, 2001.

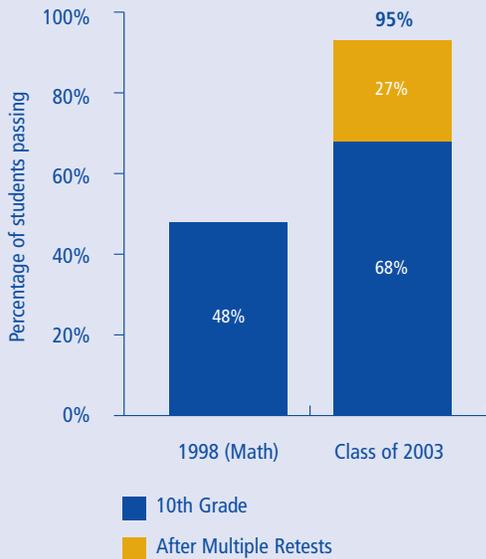
⁴Public Agenda, *Reality Check 2002*, 2002.

⁵Mackinac Center for Public Policy, *The Cost of Remedial Education*, 2000.

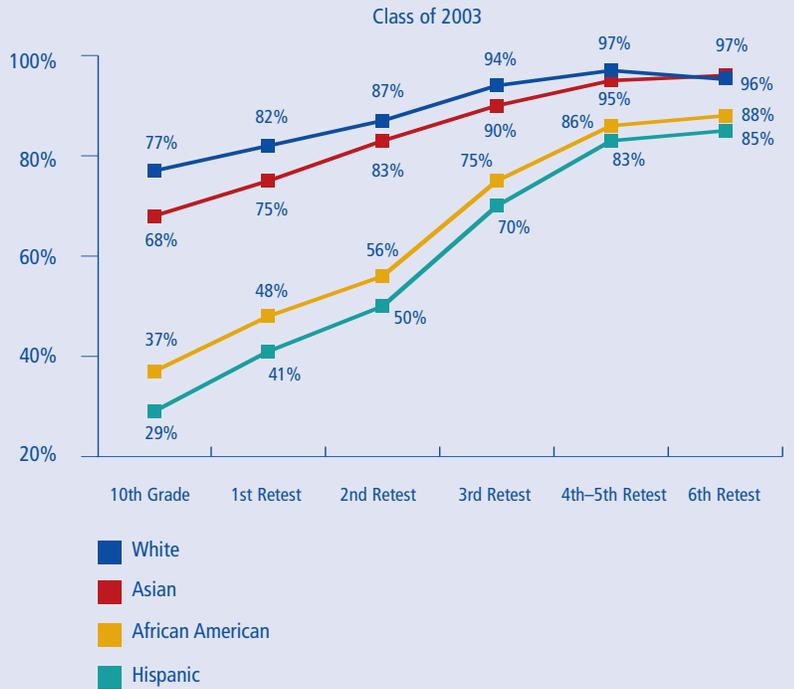
Rising to the Challenge

Although most states have been using standards-based tests as exit exams only for a short time, the emerging record is one of students rising to the challenge of higher expectations. The most promising story may be that of Massachusetts. Only 48 percent of students passed the math portion of the Massachusetts Comprehensive Assessment System (MCAS) in 1998. By providing extensive extra assistance and sparking substantial reforms in urban high schools, Massachusetts dramatically raised student performance over time. Five years later, 95 percent of the class of 2003 — the first students to face the MCAS graduation requirement — were able to pass. While the passing rates for black and Hispanic students remained lower than that of whites, the achievement gap closed dramatically — from 48 percentage points to 12 in the case of Hispanics. Those who initially failed the test and took advantage of the extra help available to them were significantly more

Student Performance on MCAS Improving Dramatically ...



... Achievement Gap Closing



Source: Massachusetts Department of Education, Progress Report on Students Attaining the Competency Determination Statewide and by District: Classes of 2003, 2004, 2005, 2004.

likely to pass the test the next time, underscoring the importance of student effort and responsibility. Perhaps even more promising is the trend in the two subsequent groups of prospective graduates, which shows a steady climb in the percentage of students passing the test on their first attempt.⁶

Making Continuous Improvement

The exit exams reviewed by Achieve remain works in progress, according to state leaders themselves. The exams are a step forward from minimum competency exams they replaced. In Maryland, for example, the new High School Assessments are replacing the Functional Literacy Tests, which students took as early as 7th grade and which many students had passed before they ever reached high school. The new Ohio Graduation Tests place more emphasis on algebra and geometry than the earlier 9th grade Ohio Proficiency Tests did. Ohio's previous tests also used only multiple-choice questions. The new tests require students to answer open-ended questions and show how they arrived at answers. The other states in the study also contend that their exams have become more ambitious over time.

Given the current levels of student performance, all states using exit exams must balance the need for increased rigor with the capacity of their education systems to raise performance statewide — particularly for the most disadvantaged students. These tests are intended to help promote sorely needed instructional improvements and to focus resources and attention where they are needed most. They are not intended to penalize students.

States that have exit exams should not back away from them. They do students no favors by awarding them a high school diploma if they are unprepared for the opportunities and challenges that await them. That, after all, is the ultimate goal of these exams and the other graduation requirements states have established. The states in this study and other states with exit exams are pushing their school systems to improve out of a sense of responsibility for young people. Achieve designed this study with a similar sense of responsibility in mind. Our goal is to help inform the national conversation about high school graduation standards, so that all youngsters who earn a high school diploma are well equipped for the future.

⁶Massachusetts Department of Education, *Progress Report on Students Attaining the Competency Determination Statewide and by District: Classes of 2003, 2004, 2005, 2004.*

Findings

Overview of Methodology

The six states in the study provided Achieve with the most recent version of the English language arts and mathematics assessments they use as exit tests, as well as data about the level of performance required to pass the exams.⁷

The foundation of Achieve's analysis was a thorough description of each item on the tests grounded in several dimensions of test quality.⁸ Two reviewers who are trained to use coding schemes for each dimension examined each test item and rated it. These reviewers worked independently and reconciled any differences in their judgments before final characterizations were assigned to each question. With these detailed analyses of each question, Achieve was able to aggregate the data to build an overall picture of each test, which allowed for cross-state comparisons.

One dimension examined was the content the tests measure. In math, for example, how much algebra appears on each test, and what kind of algebra is it? In addition, two independently devised benchmarks proved useful in estimating the grade level of content. In math, Achieve used an international scale created as part of the Third International Mathematics and Science Study (TIMSS). In English language arts, Achieve adapted a scale used by ACT, Inc., to describe questions on its college preparatory and admissions tests.

Another important dimension considered was the complexity or cognitive demand of each question and of the items as a group — what the questions ask students to do with their knowledge in English language arts and math. In reading, for example, students can be asked simply to recall information from a text — a relatively low-level skill — or they can be asked to perform a more complex task such as comparing imagery across different passages.

In addition, in evaluating the English language arts tests, the complexity of reading passages also figured into the analysis because the interaction of cognitive demand and the difficulty of a passage contributes to the rigor of reading tests. To address this dynamic, Achieve developed a Reading Rigor Index to rate items.

⁷The following tests were reviewed for this study:

*Florida: Grade 10 FCAT (Florida Comprehensive Assessment Test) for Mathematics and Reading, 2003

*Maryland: Maryland High School Assessment for Algebra/Data Analysis, Geometry, and English I, 2002

*Massachusetts: Grade 10 MCAS (Massachusetts Comprehensive Assessment System) Mathematics and English Language Arts Tests, 2002

*Ohio: Ohio Graduation Test for Mathematics, 2004; and Reading-English Language Arts, 2003

*New Jersey: Grade 11 HSPA (High School Proficiency Assessment) for Mathematics and Language Arts Literacy, 2003

*Texas: Grade 11 TAKS (Texas Assessment of Knowledge and Skills) for Mathematics and English Language Arts, 2003

In all states Achieve reviewed the most recent test for which the statistical data necessary for this study was available.

⁸To conduct the analyses described here, Achieve, with the assistance of a committee of psychometricians and mathematics and English language arts experts, created several new analytic tools, which permitted Achieve to describe and compare various aspects of these tests. See "Appendix" for a description of these scales and procedures.

This analysis also probed how well the exit tests measure what matters most to postsecondary education institutions and to employers in high-growth, high-performance industries. Through its work on the American Diploma Project (ADP) over the last two years, Achieve has defined concretely the English and math skills and knowledge high school graduates must have. This research, which involved more than 300 front-line managers and faculty members from two- and four-year postsecondary institutions, found a striking convergence in what it takes to succeed in credit-bearing college courses or in “good” jobs — those that can support a family well above poverty and offer benefits and career advancement. ADP’s benchmarks — which cover math content and skills roughly equivalent to Algebra I and II, Geometry, data analysis and statistics; English equivalent to four years of grade-level coursework; and analytical and reasoning skills typically associated with honors classes — were used to evaluate how well these tests measure college and workplace readiness.

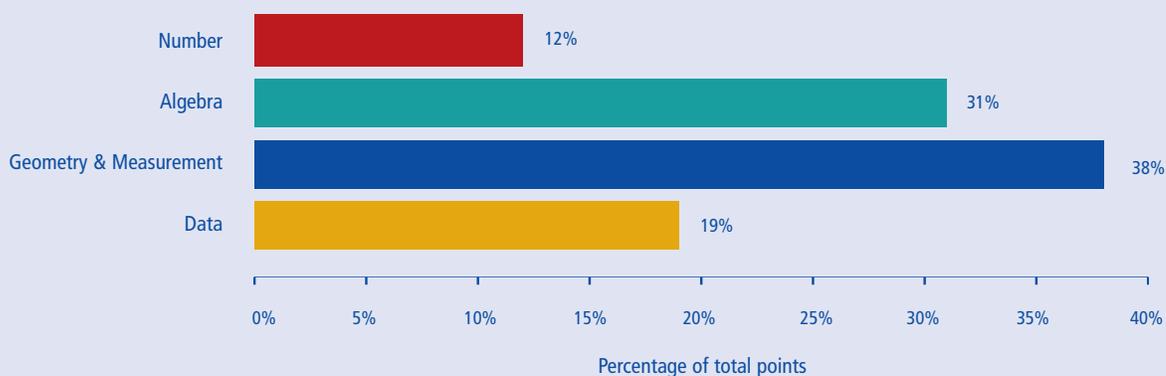
Finally, this analysis explored what it takes for students to pass each state test and how those expectations compare across states. Achieve and researchers from Michigan State University devised a statistical approach to allow cut scores from different states’ tests to be compared on the TIMSS and ACT scales. Using this approach, Achieve was able to compare across the states the difficulty of the questions that students scoring at the cut score answer correctly, which helps to paint an overall picture of how challenging the tests are to pass. For more information about the methodology used for this study, see “Appendix,” page 36.

What Do the Mathematics Exams Measure?

Tests Emphasize Algebra and Geometry, but at a Basic Level

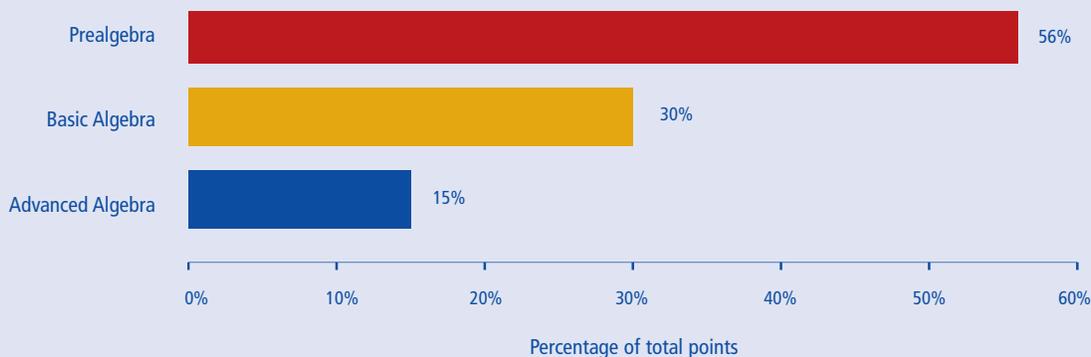
When Achieve divided the questions on the math tests into the discipline's four major domains — number, algebra, geometry/measurement and data analysis — the tests looked like one would expect at the high school level: The majority of the questions focused on algebra and geometry/measurement, followed by data analysis and number. That algebra and geometry receive top billing on these tests represents a welcome shift from the minimum competency tests states used in the 1980s, which focused heavily on arithmetic and computation. We know that students today need strong algebra, geometry and measurement skills to succeed when they graduate. These are considered prerequisites for success in college and in the new economy.

Content on Math Tests Across States



However, a closer look at the particular algebra and geometry/measurement topics being measured on these tests reveals that a majority of the points students can earn are associated with the least demanding topics. In algebra, for example, more than half (56 percent) of the points students can earn are based on facility with prealgebra skills — skills students should have mastered before high school. These include such basic skills as working with integers, rational numbers, patterns, representation, substitution, basic manipulation and simplification. Less than one-third of the points are dedicated to concepts such as linear equations, basic relations and functions typically associated with basic algebra or Algebra I — a course commonly taken in the freshman year of high school or even earlier. An even smaller proportion of the algebra points (15 percent) reflect advanced algebra concepts typically encountered in Algebra II courses. Few of the test questions, for example, address non-linear equations and inequalities or require students to demonstrate their understanding of real or complex

Algebra on Math Tests Across States

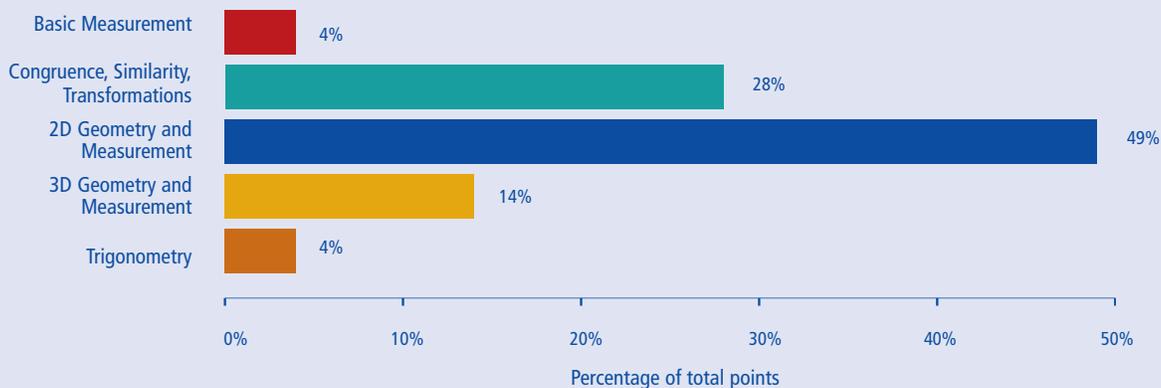


NOTE: Totals may not equal 100 percent due to rounding.

numbers, yet these are skills college-bound students will need to succeed in credit-bearing math courses.

The same pattern of emphasis holds true in geometry and measurement. Almost half the geometry/measurement points on these tests are associated with two-dimensional geometry and measurement, including understanding basic facts, shapes and properties. Only a small proportion of the points (14 percent) is attributed to three-dimensional geometry — concepts such as volume and surface area. In contrast, the National Assessment of Educational Progress (NAEP) includes two-dimensional geometry and measurement on its 8th grade assessment and expands to include formal three-dimensional geometry on its 12th grade assessment.

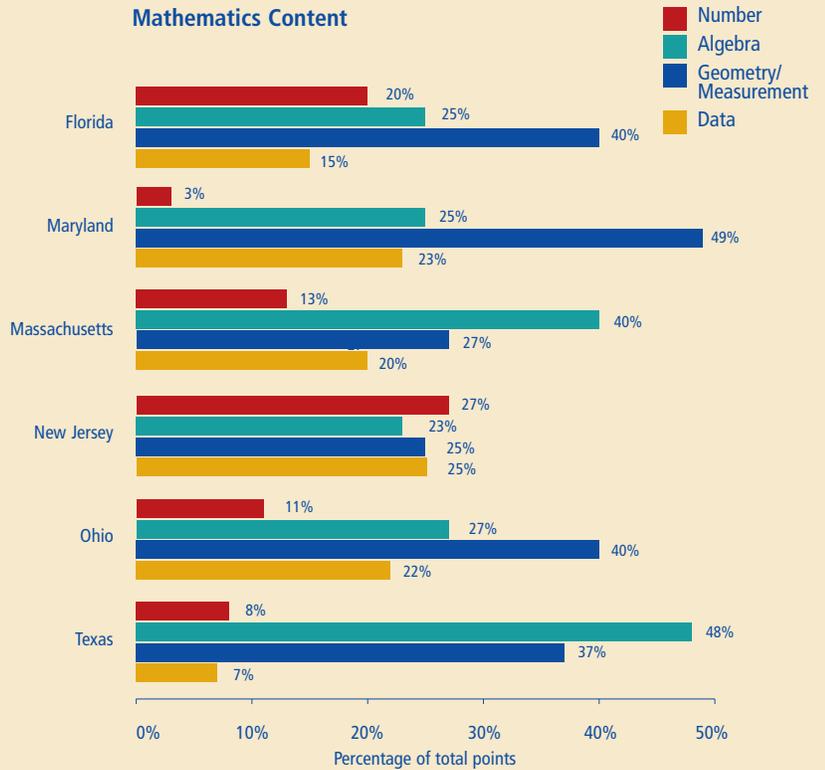
Geometry/Measurement on Math Tests Across States



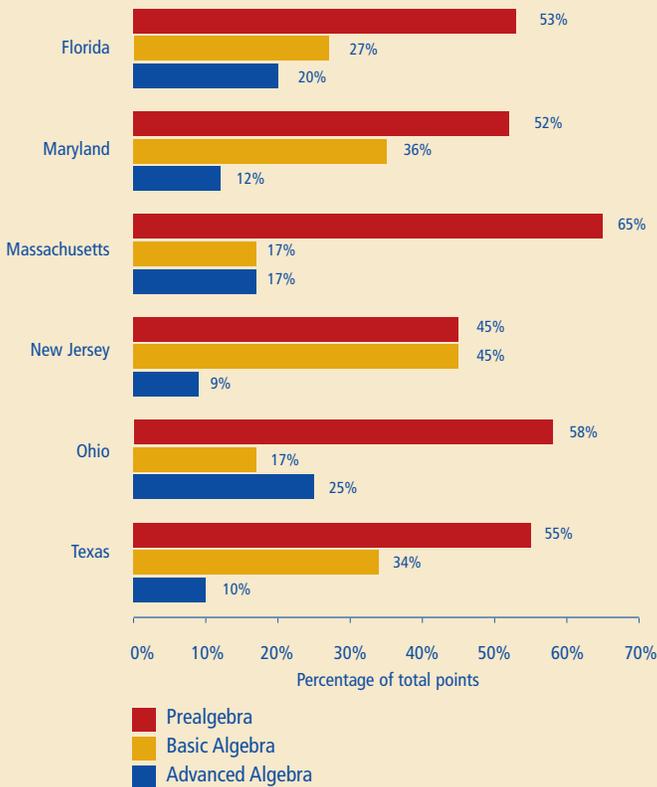
NOTE: Totals may not equal 100 percent due to rounding.

A state-by-state look at the content on the math tests shows that every state except New Jersey targets a majority of the points to algebra and geometry. Every state except for New Jersey emphasizes prealgebra over basic algebra (Algebra I), and few states emphasize advanced algebra (Algebra II) on their tests. (Of the six states in this study, Ohio and Florida give Algebra II the greatest weight.) In geometry and measurement, two-dimensional concepts are emphasized more than three-dimensional concepts such as surface area and volume.

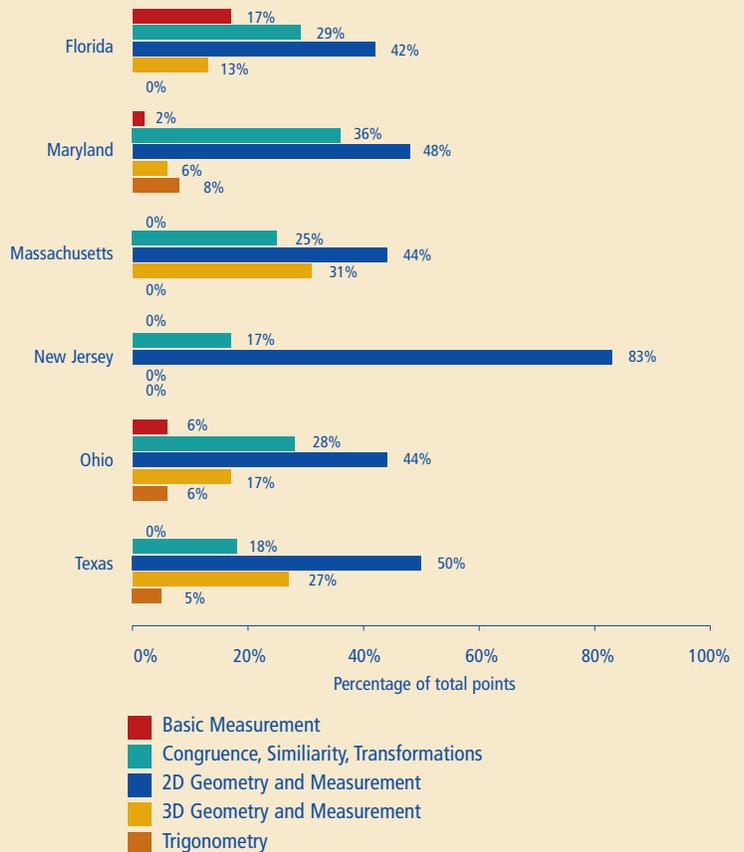
Mathematics Content



Algebra



Geometry/Measurement

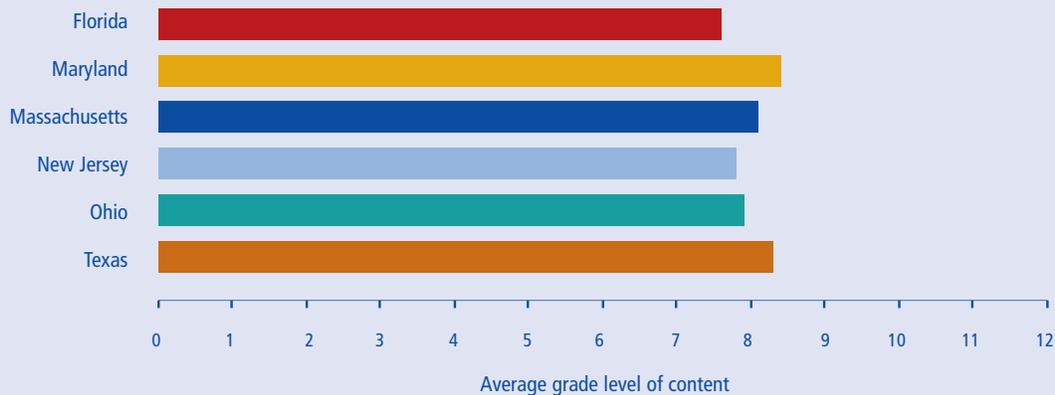


Tests Measure Math Concepts Students in Other Countries Study in 8th Grade

Because the performance of U.S. high school students in math lags behind that of students in other industrialized countries, it is valuable to compare what is expected of students on these tests with expectations in other countries. Achieve had the advantage in this study of looking at the math exams by means of the International Grade Placement (IGP) index developed by Michigan State University as part of its ongoing work on the Third International Mathematics and Science Study (TIMSS). The IGP index represents an “average” or composite among 41 nations of the world (both high-performing and low-performing countries) as to the grade level in which a mathematics topic typically appears in the curriculum. For example, decimals and fractions are typically focused on at the 4th grade level internationally. Therefore, this topic has an IGP rating of four. Right-triangle trigonometry, on the other hand, is most often taught in the 9th grade around the world, so it receives an IGP rating of nine.

When Achieve applied the IGP index to the state exit exams, it revealed that the content measured on the tests is taught, on average, at the 8th grade level internationally. In other words, the material on the exams states are using as a requirement for high school graduation is considered middle school content in most other countries. While there was some variation across the states, this finding held true for all of the math tests. No test had an average IGP rating higher than the 8th grade.⁹

Content on Each State’s Math Tests on TIMSS International Grade Scale



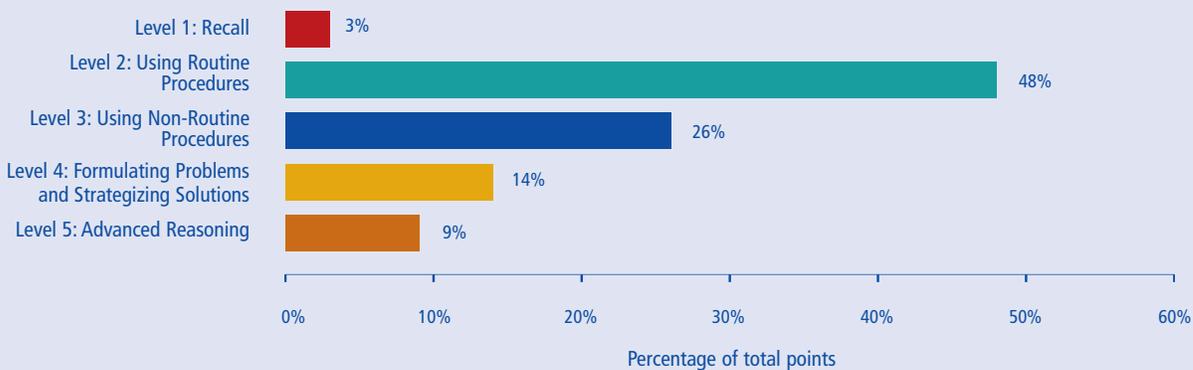
⁹It is important to note that this is an international, not domestic, grade index so that the 8th grade level on the IGP index may not equate to the typical 8th grade curriculum in the United States. In fact, TIMSS data suggest that students in many other countries are expected to master arithmetic concepts by the end of 6th grade and advance to the fundamentals of algebra and geometry in the 7th and 8th grades, which is earlier than in many states.

Most Test Questions Have a Relatively Low Level of Cognitive Demand, but Some States Aim Higher

The content measured by the test items tells an important part of the story, but a more complete understanding of what these tests measure requires an examination of the cognitive demand of the questions as well. Are students asked to apply routine procedures to mathematical problems? For example, does the question simply ask students to multiply two fractions to arrive at the answer? Or is the question framed in such a way that it requires students to first develop a more complex mathematical model to solve the problem? For example, students may need to create cost functions based on monthly rate plus cost per minute as the basis for comparing the cost of multiple cell phone plans.

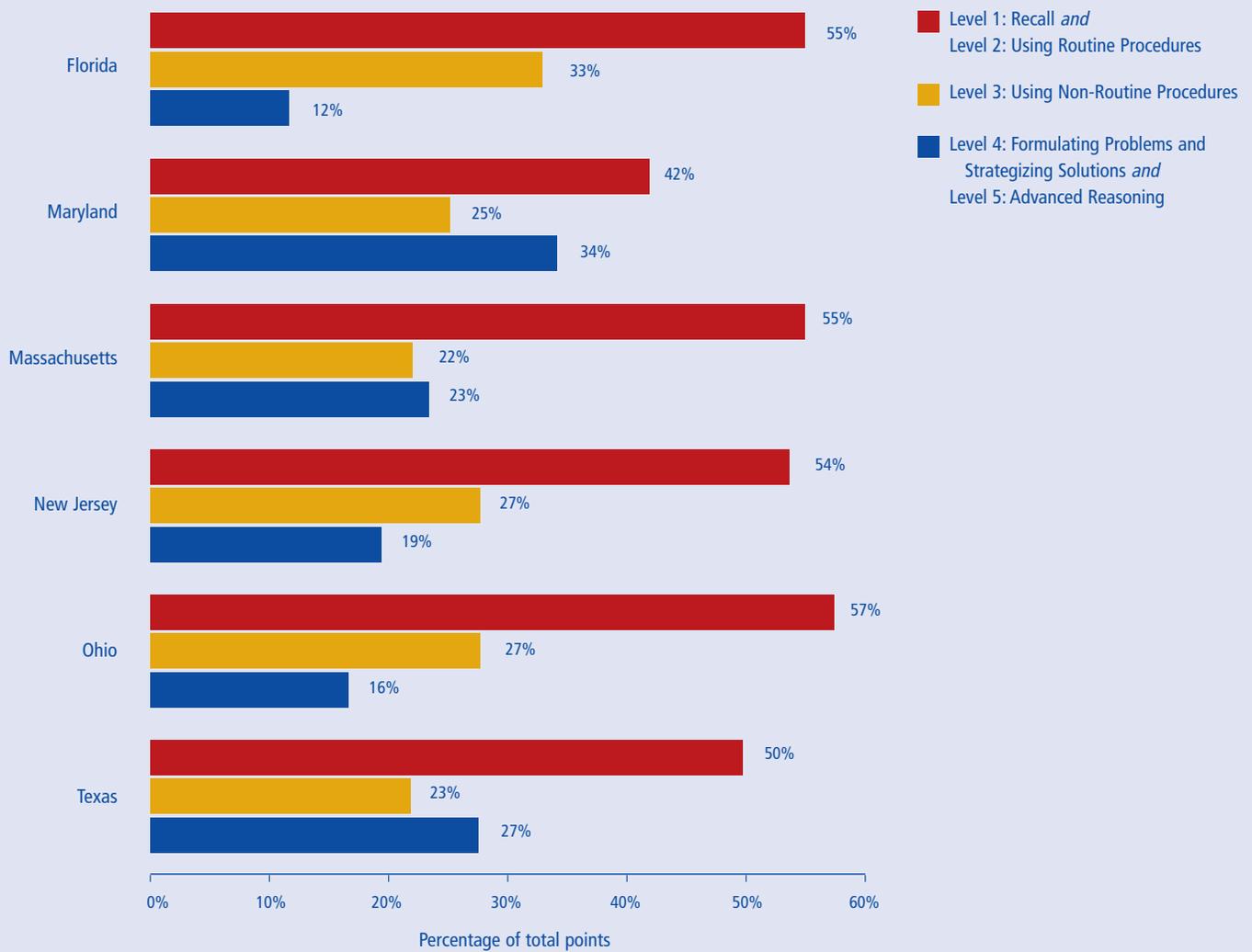
A majority of the points on the tests across the six states were associated with questions at the lower end of the cognitive continuum. On a five-point scale of rigor, with one being lowest and five highest, more than half of the questions across the tests fall at the lowest two levels. Forty-eight percent of points across the state math tests are devoted to Level 2 items — items that require students to use routine procedures and tools to solve math problems. These are only slightly more demanding than simple recall and recognition items. About a quarter of the points across all of the tests are attributed to items that require more advanced mathematical skills (Levels 4 and 5).

Cognitive Demand of Math Tests Across States



There are notable differences in the cognitive demand of the test questions across the states as well. For example, more than a third of the questions on the Maryland end-of-course tests tap Level 4 and 5 skills, while only 12 percent of Florida's and 16 percent of Ohio's test questions aim at that level.

Mathematics Cognitive Demand Levels by State



MEASURING COLLEGE AND WORKPLACE READINESS IN MATH

Whether they are interpreting graphs in an introductory economics course or determining the volume of a fertilizer mix on a chemical plant's production line, high school graduates face real-world demands for math knowledge and skills in college and in "good" jobs — those that can support a family well above poverty and offer benefits and career advancement.

As part of the American Diploma Project, Achieve worked for two years in partnership with The Education Trust, the Thomas B. Fordham Foundation and five states to define concretely the math that graduates must know to succeed in the classroom or on the job. When compared to the college and workplace readiness benchmarks that emerged from that research involving hundreds of front-line managers and faculty members, high school graduation tests do a good job measuring some of the necessary knowledge and skills (computing with fractions and decimals, for example). But none of the tests adequately measures the full range of the benchmarks.

- Skills and knowledge typically taught in Algebra II are critical whether students are going to college or aspire to highly paid, professional or well-paid, white-collar jobs. Only 15 percent of the points devoted to algebra across these six states' graduation exams — and only 5 percent of the points on the tests overall — are awarded to questions measuring Algebra II content.
- While two-dimensional geometry and measurement are covered well on these tests, three-dimensional concepts are covered only minimally. Only 14 percent of the geometry points across the tests — and 5 percent of the points on the tests overall — are devoted to concepts such as the properties and characteristics of solids, volume and surface area. Employers, in particular, stress the importance of these skills in industries such as construction and manufacturing.
- Although basic number concepts — fractions, decimals and percents — are assessed thoroughly on these exit exams, few of the questions go beyond that to assess concepts that are essential to understanding more advanced math and science. Only 2 percent of the points across the state tests target ratio and proportionality — concepts that are critical for making sense of fractions and variables and that are used in constructing graphs and interpreting maps, blueprints and schematic drawings. Minimal attention also is given to integers, rational numbers, real numbers, powers and roots.
- To succeed in many of today's jobs, graduates will need to analyze, interpret and describe data quickly and to create visual representations of data — charts, graphs and diagrams — to help convey a point. Nearly one-fifth of the points across the tests are devoted to these concepts.

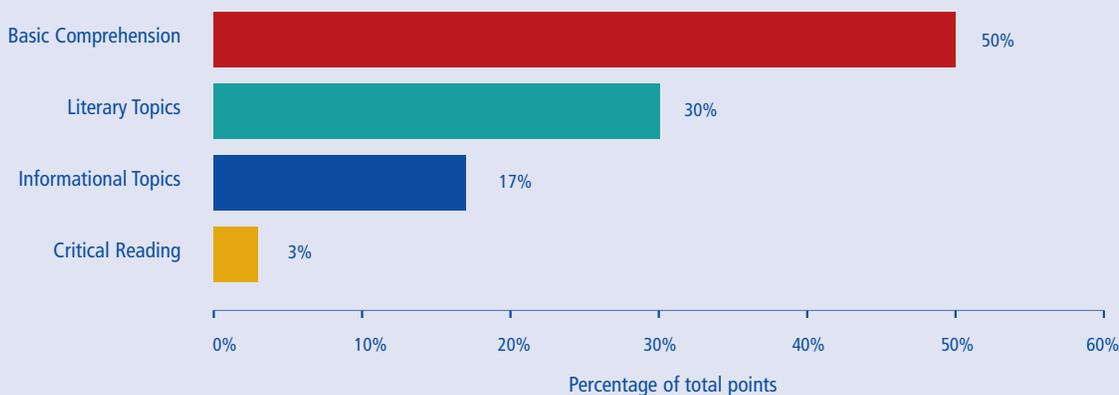
What Do the English Language Arts Exams Measure?

Four of the six states in the study have both reading and writing components of their exit exams — Maryland, Massachusetts, New Jersey and Texas. Ohio and Florida do not currently have a writing component, but officials in both states report that they are planning to add writing to graduation requirements in the future. For this aspect of the study, Achieve separated the analyses of the reading and writing parts of the exams. In the discussion of what it takes to pass the tests, reading and writing are combined.

Tests Emphasize Basic Reading Comprehension, but They Are Divided in Treatment of Informational Topics

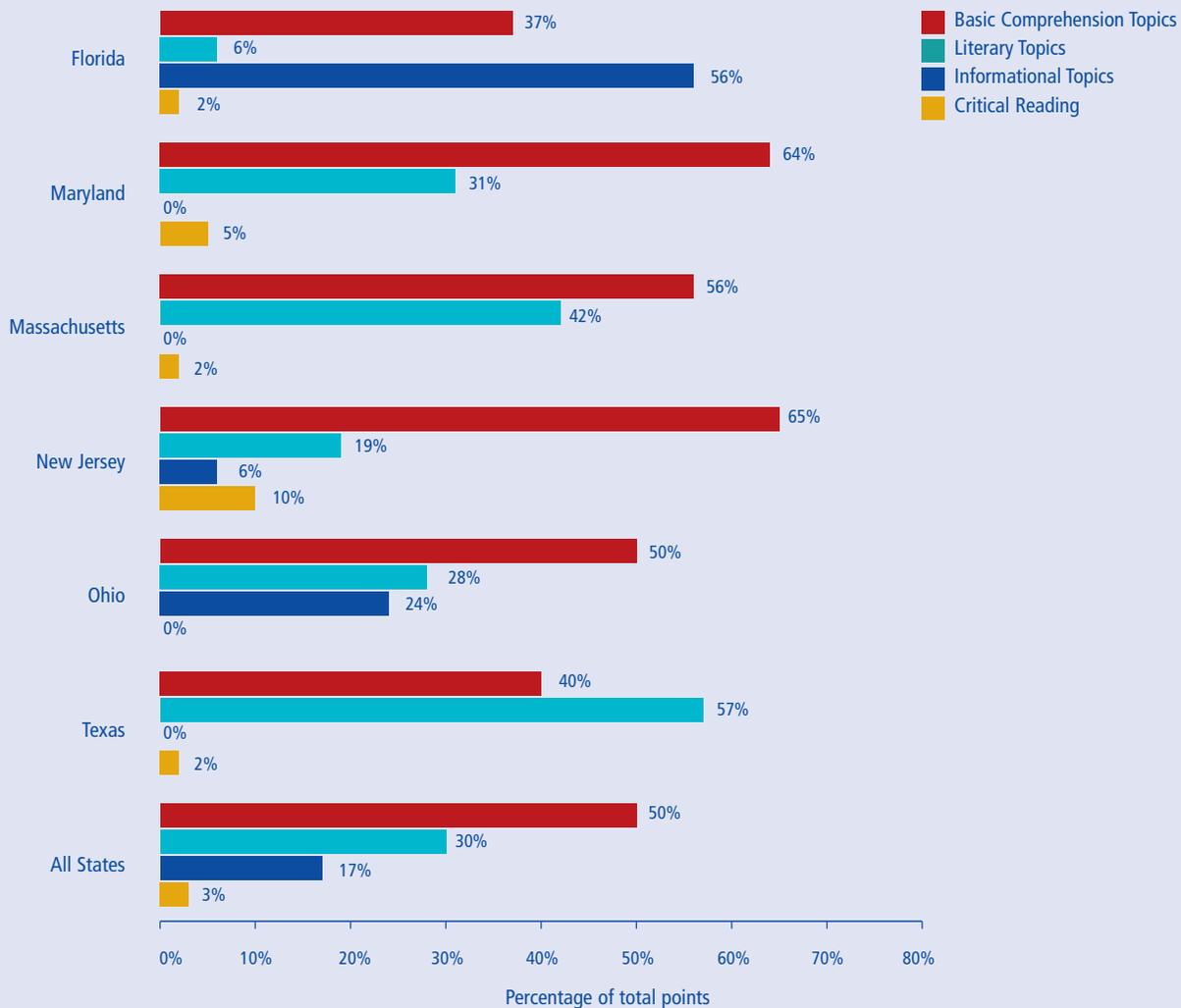
Achieve's analysis suggests that the state tests have one thing in common: They pay greater attention to basic topics and less attention to more advanced topics. Fifty percent of the total points on the assessments across the six states are devoted to basic reading comprehension topics, such as vocabulary, general comprehension, and main idea or theme of a reading passage. Only 3 percent of the points are associated with more advanced critical-reading skills — including discerning fact from opinion and faulty from logical reasoning — yet these are skills that college faculty and front-line managers in a variety of industries agree are essential to success in higher education or on the job.

Content on Reading Tests Across States



NOTE: Totals may not equal 100 percent due to rounding.

Content on Reading Tests by State



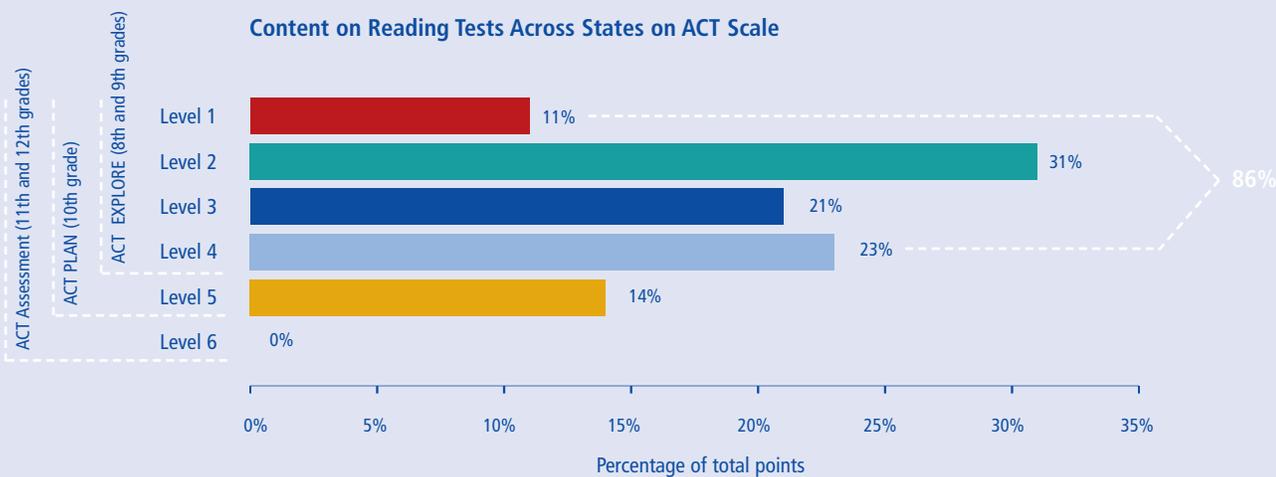
NOTE: Totals may not equal 100 percent due to rounding.

Perhaps most striking is the variable treatment of informational and literary topics across the six states. Literary topics include elements such as plot, setting and character in a short story or piece of literature. Informational topics target use of evidence and other features in texts such as biographical sketches, magazine articles or how-to manuals. Typically, both types of reading are included in state English language arts standards, so we had expected to see a more even-handed emphasis on the tests. Only Ohio divides its points almost evenly between the two. Maryland, Massachusetts and Texas target literary topics to the exclusion of informational topics. Their reading tests devote from almost a third to more than a half of their points to literary topics; the remainder are allocated principally to basic

comprehension topics. (Note: Massachusetts and Maryland do include informational passages on their tests, but they do not award reading points to informational topics, concentrating instead on the basic comprehension of those passages.) In stark contrast to those states, Florida targets the majority of its reading points (56 percent) to informational topics and only 6 percent to literary topics. It is worth pointing out that on the 12th grade NAEP assessment, 60 percent of the reading passages are informational, and 40 percent are literary.

Most Reading Questions Target Skills Found on ACT's 8th and 9th Grade Test

To gauge the approximate grade level of the content on the state exit exams in English language arts, Achieve used an index based on one created by ACT, Inc., to guide the development of assessments given to students as early as the 8th grade. Although ACT is most widely known for its test that college-bound students take in the 11th and 12th grades, it also has developed two additional tests, one that students take in 8th or 9th grade and another that they can take in 10th grade. These assessments provide more formative information on students' development leading up to the college admissions test.



Topics	ACT EXPLORE (8th and 9th grades)	ACT PLAN (10th grade)	ACT (11th and 12th grades)	All States in Study	Florida	Maryland	Massachusetts	New Jersey	Ohio	Texas
Level 1	10–20%	5–15%	5–15%	11%	10%	7%	24%	3%	16%	3%
Level 2	20–30%			31%	42%	21%	24%	22%	51%	11%
Level 3	30–40%	20–30%	10–20%	21%	19%	11%	20%	19%	27%	24%
Level 4	15–25%	20–30%	20–30%	23%	19%	32%	26%	25%	6%	38%
Level 5	0%	25–35%	25–35%	14%	8%	29%	6%	31%	0%	24%
Level 6	0%	0%	20–30%	0%	2%	0%	0%	0%	0%	0%

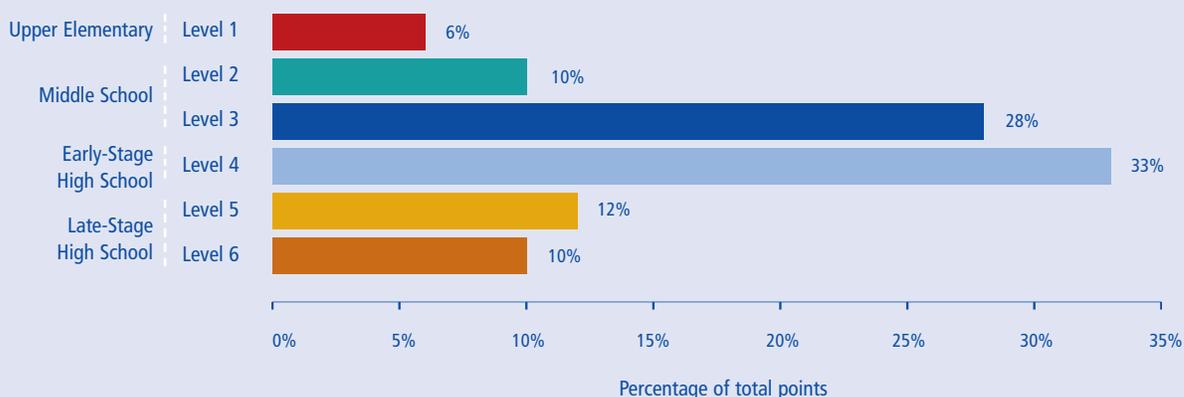
ACT has established six levels to differentiate the content and skills that are measured on its reading tests: Levels 1 through 4 cover skills found on ACT’s EXPLORE test given in the 8th and 9th grades; ACT’s PLAN test, which is given in the 10th grade, includes test questions from Levels 1 through 5; and the ACT Assessment — which students take in the 11th and 12th grades, and which colleges use in admissions, course placement and guidance decisions — incorporates questions in Levels 1 through 6.

Achieve experts coded each item on all six state tests using ACT’s six levels of demand. They then compared the levels on the state tests with the levels found on the ACT tests. When looked at in the aggregate, the level of demand on the six tests most closely resembled that of the ACT 8th and 9th grade EXPLORE test — the vast majority of points (86 percent) are tied to the less demanding content that is emphasized on the EXPLORE test. The Texas, New Jersey and Maryland tests, however, are more demanding and line up more closely with PLAN, the ACT 10th grade test. None of the state tests approaches the level of demand of the ACT college admissions test.

Reading Passage Selection Is a Strength of Tests

To help judge the complexity of reading passages on these tests, Achieve’s reading experts created a six-point scale describing texts from the relatively simple to the very complex. To develop the scale, they considered a large collection of reading passages, including not only those on the six state tests, but also samples from NAEP and college entrance exams such as the SAT and ACT. The levels are based on such characteristics as the specialization of the vocabulary, the predictability of text structures or organization, the complexity of the syntax, the level of abstractness, the familiarity of the

Reading Passage Demand Across States



NOTE: Totals may not equal 100 percent due to rounding.

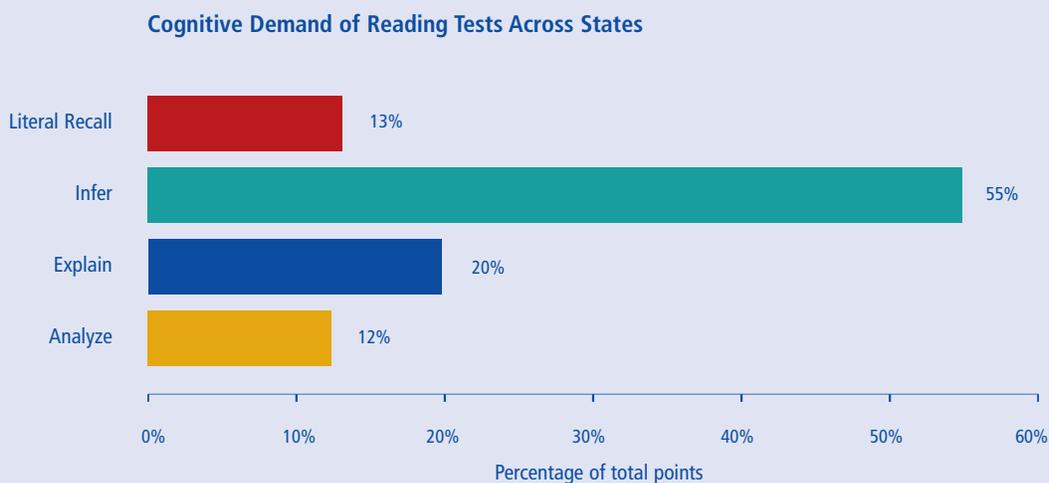
topic and the number of concepts introduced in the passage. Level 1 represents an upper-elementary reading level, Levels 2 and 3 represent middle school-level reading, Level 4 represents early-stage high school reading, and Levels 5 and 6 represent late-stage high school reading.

While some differences exist from state to state, the majority of reading passages across the states fall in the middle of the range — Levels 3 and 4 — presenting, on the whole, passages that are very accessible to high school students. These passages are neither overly simplistic nor too sophisticated. Literary texts at these levels make moderate use of figurative language, convey messages with some subtlety and may feature somewhat complex interactions among characters. Informational texts at these middle-range levels present detailed supporting data, may include some context-dependent words and are written in a somewhat more demanding style than are passages at lower levels of reading.

Only a small percentage of the passages fell at the lower levels of demand, which is appropriate given that these tests are high school graduation exams. A slightly higher percentage of passages were judged to be at the highest levels of demand, levels that students should be expected to reach later in high school. Texas and New Jersey have the highest average reading demand, followed by Massachusetts.

Most Reading Questions Have a Relatively Low Level of Cognitive Demand

An analysis of the cognitive demand of the questions on the reading tests revealed a similar situation to that of math: A majority of the points across the state tests are devoted to questions that tap lower-level skills. Sixty-eight



percent of the points on the tests focus on what Achieve considers basic reading comprehension skills — literal recall (13 percent) and inference (55 percent). Twenty percent of the points are associated with questions requiring students to provide details to support their answers — to “explain” — and only 12 percent of the total reading points across the six states focus on analysis, which is regarded as the most demanding performance and is exhibited by expert readers.

New Jersey, Texas, Massachusetts Reading Passages, Questions Are Most Rigorous

The difficulty of a reading test is determined not only by the cognitive demand of the questions but also by the complexity of the reading passages the questions address. To capture this important interplay, Achieve developed a Reading Rigor Index (RRI) that combines the cognitive challenge level of an item with the difficulty level of the passage that the item targets (see “Appendix” on page 36 for a more detailed explanation). The index uses a nine-point scale to classify individual reading questions and entire reading tests. The low end of the scale reflects less demanding questions about less complex reading passages; the higher end reflects challenging questions about complex texts.

Although the majority of the reading questions across the six states fall in the middle of RRI, some states’ tests — most notably New Jersey’s — score higher on the scale, suggesting a more challenging reading assessment. The Texas and Massachusetts tests also rank high on this scale relative to the other states.

Reading Rigor Levels	Florida	Maryland	Massachusetts	New Jersey	Ohio	Texas
1 (low)	0%	0%	0%	0%	0%	0%
2	0%	0%	8%	0%	10%	2.5%
3	14%	15%	4%	3%	24%	2.5%
4	38%	19%	18%	33%	33%	5%
5	33%	44%	26%	14%	21%	22%
6	15%	7%	20%	3%	12%	35%
7	0%	15%	16%	22%	0%	20%
8	0%	0%	8%	11%	0%	13%
9 (high)	0%	0%	0%	14%	0%	0%

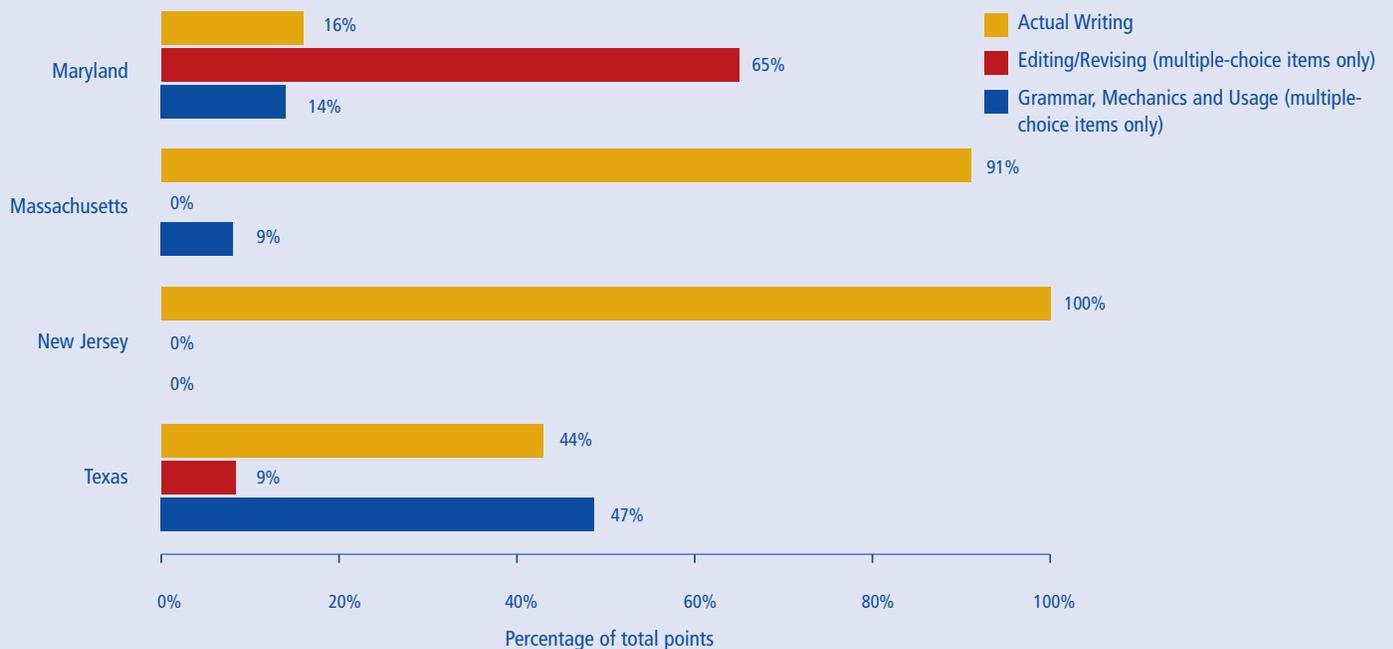
States Use Different Approaches to Test Writing

Four states taking part in the study — Maryland, Massachusetts, New Jersey and Texas — currently assess writing in some form on their exit exams.¹⁰ Florida currently gives a 10th grade writing test and plans to include this essay exam in its graduation requirements. Ohio also plans to develop a writing section for its graduation test.

Maryland relies most heavily on indirect writing measures (multiple-choice items that assess grammar, punctuation, and editing and revision skills). New Jersey and Massachusetts, on the other hand, focus most of their writing assessments on on-demand writing. Students are asked to write essays, and they are scored based on the quality of the writing. Texas has a more balanced approach.

New Jersey is the only state that relies exclusively on actual writing. Texas dedicates slightly more than half of its points — 56 percent — to multiple-choice items that assess grammar, punctuation and sentence revision, whereas Massachusetts assigns only 9 percent to such multiple-choice language items. The remaining 91 percent of its points are devoted to actual writing. Maryland dedicates the largest proportion of points to multiple-choice questions that measure editing and revision skills.

Content on Writing Tests by State



NOTE: Totals may not equal 100 percent due to rounding. Also, the Maryland writing exam includes two items that assess dictionary skills, which were not regarded as writing items.

¹⁰ Each English language arts test reviewed included open-ended items in which students had to write their responses to questions about a reading passage. Such items, however, were scored by the states only for reading comprehension, not for the writing skills displayed by the test takers. In this study, Achieve regarded as "writing" only those items or tasks that were intended to assess compositional skills.

MEASURING COLLEGE AND WORKPLACE READINESS IN ENGLISH

Whether they are preparing a “toolbox talk” to convey important safety rules to a construction crew or writing an expository essay for a first-year philosophy class, high school graduates face real-world demands for English knowledge and skills in college and in “good” jobs — those that can support a family well above poverty and offer benefits and career advancement.

As part of the American Diploma Project, Achieve worked for two years in partnership with The Education Trust, the Thomas B. Fordham Foundation and five states to define concretely the English that graduates must know to succeed in the classroom or on the job. When compared to the college and workplace readiness benchmarks that emerged from that research involving hundreds of front-line managers and faculty members, high school graduation tests do a good job measuring some of the necessary knowledge and skills (recognizing the main idea of a text, for example). But none of the tests adequately measures the full range of the benchmarks.

- These graduation exams do a good job measuring fundamental reading skills, such as recognizing the main idea in a text, identifying the theme and recognizing words in context. These skills are cited by both college professors and employers as prerequisites for success in their institutions.
- In today’s college classrooms and workplaces, graduates are required to comprehend, interpret and judge the quality of a wide range of “informational” materials, such as business correspondence, academic journal articles, financial statements, reference materials and newspaper editorials. However, these skills are not emphasized consistently across the exit exams. Only Florida and Ohio place real emphasis on informational reading, while the other four states each devote less than 10 percent of their reading points to this area.
- High school graduates today are increasingly expected to judge the credibility of sources, evaluate arguments, recognize faulty reasoning, and understand and convey complex information in the college classroom, in the workplace and as they exercise their rights as citizens. However, very few questions on state exit exams measure these critical reading skills.
- Strong writing skills have become an increasingly important commodity in the 21st century. States have taken different approaches to measuring writing — some do not currently include a writing component in their graduation tests; others award a significant percentage of writing points based on multiple-choice questions targeting grammar and editing skills; others measure writing with actual, on-demand writing tasks. The discipline and skill required to create, reshape and polish pieces of writing “on demand” prepares students for the real world, where they inevitably must be able to write quickly and clearly, whether in the workplace or in college classrooms.

How Difficult Is It To Pass the Tests?

Up until this point, the report has focused on what is measured on the state exit exams — the content, the difficulty level of the questions and the complexity of the reading passages. However, students taking these tests are not required to answer all of the questions correctly to pass. States establish “cut scores” that students need to achieve to pass the tests. To achieve those scores, students usually need to get only a certain number of items correct or earn a certain number of points. These cut scores define the level of achievement that students are ultimately held accountable for — they establish the “floor” of performance required to earn a high school diploma.

Comparative studies of where states set their cut scores are rare and difficult to do. They typically involve comparing the percentage of students passing each state’s test with the percentage of students passing a common test, such as NAEP. This methodology permits judgments about the relative difficulty of different tests, but it does not provide information on what students need to know to pass each test.

Achieve worked with Michigan State University to develop a new method for analyzing those scores and comparing them from state to state. The procedure, described more fully in “Appendix” on page 36, uses the IGP index in math and the ACT index in English language arts to compare what the typical student needs to do on these tests to pass. It is important to note that the analysis focused only on students scoring at the cut score, not on those who scored at higher levels on the tests. In lay terms, Achieve looked at what it takes to *just pass*, rather than to *comfortably pass* or “*ace*” the tests.¹¹

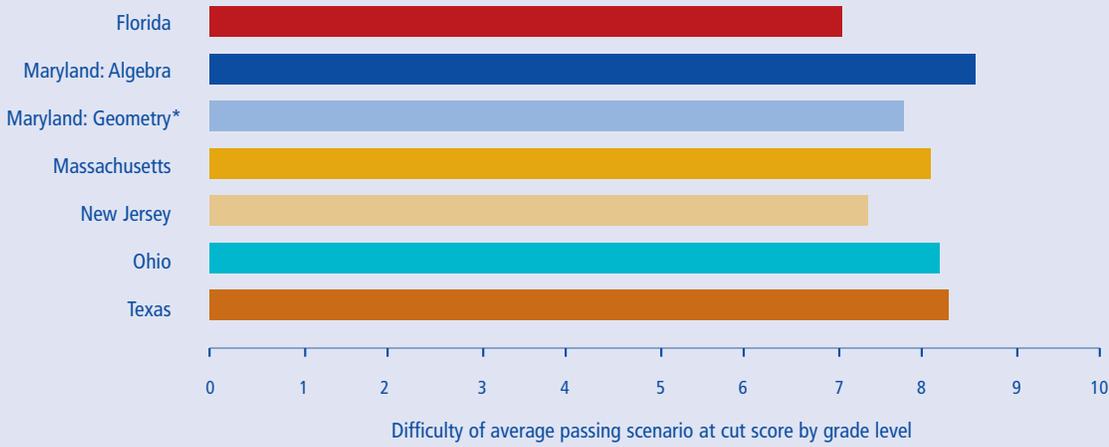
Students Can Pass the Math Tests with a Command of Content Typically Taught in 7th and 8th Grade around the World

Based on the IGP index, Achieve found that the level of math content knowledge students need to pass the state exit exams falls, on average, at the 7th or 8th grade level internationally. That is, the questions on the tests that students scoring at the cut score are likely to get correct measure concepts that students around the world focus on in the 7th and 8th grades.

When the tests were compared across the states, Maryland’s end-of-course algebra test appeared to be the most challenging one to pass in terms of content difficulty. The Texas, Ohio and Massachusetts tests followed, and these three resemble each other closely in topic challenge at the cut scores.

¹¹ The cut scores Achieve reviewed for Ohio were from the 2004 math test and the 2003 English language arts test (the 2004 English language arts test was not available in time for the analysis). The Ohio State Board of Education opted not to approve the 2003 passing score in English language arts but did set passing scores based on the 2004 tests in both subjects.

Difficulty of Passing Each State's Math Test Based on TIMSS International Grade Scale



*Not required for graduation

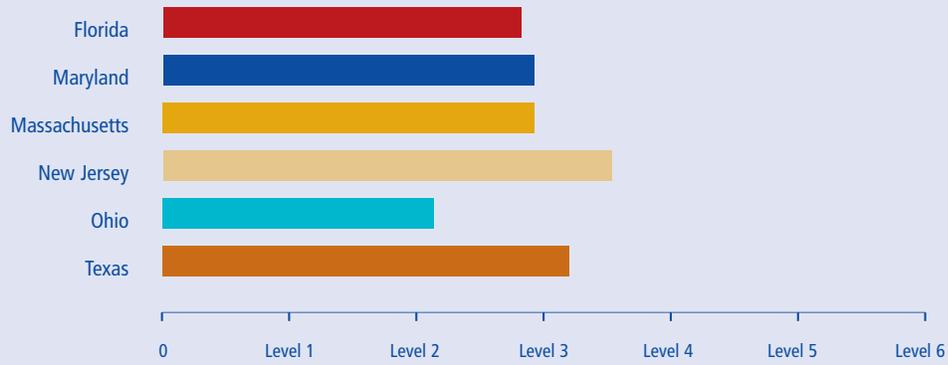
Students Can Pass the English Language Arts Tests with Skills ACT Expects of 8th and 9th Graders

Achieve used the ACT skills hierarchy to compare cut scores across the English language arts tests. Because cut scores on the English language arts tests that have reading and writing components are derived based on a combined reading and writing score, Achieve combined reading and writing for the cut score analysis. Achieve modified the ACT scale to allow for combining these two elements.

Across the state tests, the average ACT skill level of the typical set of test items that must be answered correctly to pass at the cut score is 2.9 out of a six-point scale. ACT assesses Levels 1–3 most heavily on the EXPLORE test, which is given to 8th and 9th graders. The PLAN test, given to 10th graders, focuses most heavily on Level 3–5 questions, while the ACT college admissions exam focuses on Levels 4–6. Given this frame, students scoring at the cut score on these state exit exams are generally being asked to perform at the level that ACT considers appropriate for 8th and 9th graders.

The New Jersey English language arts test appears to be the most challenging one to pass among the six, followed by Texas. That New Jersey's and Texas' exams are at the top of the list is not surprising given the relatively high level of content and cognitive demand in these English language arts tests. Massachusetts', Maryland's and Florida's tests present approximately the same challenge for passing, while reaching the cut score on the Ohio exam is less demanding.

Difficulty of Passing Each State's English Language Arts Test Based on ACT Scale



Difficulty of average passing scenario at cut score by ACT level

ACT EXPLORE (8th and 9th grades)

ACT PLAN (10th grade)

ACT Assessment (11th and 12th grades)

Recommendations for States

Based on our extensive analysis, Achieve has come to three main conclusions. *First*, it is perfectly reasonable to expect high school graduates to pass these tests — they are not overly demanding. States should neither lower the standards on these exit exams nor delay their implementation. *Second*, these exams will need to be strengthened over time to better measure the knowledge and skills high school graduates need to succeed in the real world. These improvements will need to be made gradually, so that as expectations rise, students are provided with the supports they need to succeed. *Third*, states should not rely exclusively on these tests to measure everything that matters in a young person’s education. Over time, states will need to develop a more comprehensive set of measures beyond on-demand graduation tests.

Don’t Lower the Standards or Delay Implementation

None of the tests Achieve reviewed in this study presents unreasonable expectations for high school graduates. On the contrary, the tests cover material that most students study by early in their high school careers. Given where the bar is set, it is perfectly reasonable for states to require students to pass these exams to earn a high school diploma.

- **The questions on these exams reflect material most students study in middle school and early high school.** In math, the tests place a heavier emphasis on prealgebra and basic geometry and measurement concepts — content most students complete in middle school — than on concepts associated with Algebra I and high school geometry. In English language arts, the exams are a better measure of basic reading comprehension skills than the more advanced reading skills students will need to succeed in college and the new economy.
- **The “cut scores” required to pass the tests reflect modest expectations.** To pass the math tests, students in these states need to successfully answer questions that, on average, cover material students in most other countries study in 7th or 8th grade. To pass the English language arts tests, students need to successfully answer questions that ACT considers more appropriate for the test it gives to 8th and 9th graders than its college admissions test.
- **The tests measure only a fraction of the knowledge and skills that colleges and employers say are essential.** A high school diploma should signify that students are prepared for their next steps in life, whether that means continuing on to postsecondary education or pursuing a fulfilling

career. The Achieve analysis suggests that the exit exams in these six states measure some of the skills essential for college and workplace success, but a significant number of those skills go largely unmeasured. The skills that do get measured are staples in the real world. Students simply cannot succeed without them. But the large gap that exists between these tests and the real-world expectations of colleges and employers is clear evidence that these tests are not aiming too high; in fact, one could argue they are not aiming high enough.

In states where the exit exams are being debated, Achieve strongly encourages policymakers not to lower the standards or delay implementation. Massachusetts has proven that all students can reach these standards if they are given the opportunity and the support. Only 48 percent of 10th graders in that state passed the math portion of the exit exam when it was first given in 1998. Some called for the state to lower the bar or delay implementation, but instead state officials and local educators redoubled their efforts to strengthen the curriculum and provide academic supports. In 2003, 95 percent of students in the graduating class had passed the test. The lesson for other states is clear: The students can do it, if the adults stand firmly behind them.

Continue Improving the Tests Over Time

The exit exams reviewed in this study are considerably more challenging than the exams these states once used. In Ohio, for example, the new math test places greater emphasis on algebra than the previous proficiency test did. In Maryland, algebra and geometry were not even measured on the previous graduation test; the “functional test,” as it was called, was purely a test of computation skills. The story is similar in the other states; the tests were pitched at a low level, reflecting a time when higher education was an option for a select group of students and good jobs at decent wages were available to motivated, but unskilled, people.

These states — and others around the country — have raised the bar to keep pace with the new realities that high school graduates face as part of a broader effort to restore value to the high school diploma. This analysis shows that the bar needs to be raised even higher over time. Even as they work to help students reach the current standards, states will need to pave the way for higher standards and more challenging tests. Here are some of the most important improvements states can make to their exit exams over time.

- **Create more balanced reading tests that better reflect real-world demands.** States should redirect the large percentage of points on their tests currently dedicated to basic comprehension to other important reading topics. They also should work to strike a balance between literary and informational topics.
- **Require more writing.** States that do not currently have a writing component on their exit exams should build one. States that do have a writing component, but largely measure writing through multiple-choice questions focused on grammar or revision skills, should dedicate a higher percentage of their assessments to actual demonstrations of writing. In the real world, graduates will be asked to write papers and memos, and their success will depend on their ability to do that well. States should prepare them for that reality.
- **Emphasize more challenging content.** Over time, states should increase the rigor of the content on both the math and English language arts tests. In math, this will mean increasing the emphasis on content at the Algebra I and II levels, while decreasing the amount of prealgebra. It will also mean phasing in more challenging geometry and measurement concepts. In English language arts, states need to place more emphasis on such topics as critical and persuasive reading — hallmarks of expert readers — rather than on basic comprehension. They also need to increase the challenge of the reading passages to include more upper high school reading.
- **Ask more challenging questions.** States should work over time to ensure that a larger percentage of assessment items measure higher-level skills. In both reading and math, a majority of the test items tap basic skills — routine procedures in math and inference and literal recall in English language arts. These are important skills, to be sure, but they should not crowd out higher-level skills, such as mathematical reasoning and critical reading.
- **Phase in higher cut scores.** One way states can phase in more challenging exams over time is by raising the score required for passing. Texas is using this approach with its new graduation exam. Transparency is important here — Texas has made clear to teachers, students and parents when the bar will be raised and by how much, so that everyone knows what to expect and can prepare accordingly. It is also important to be sure that the exam is robust enough, and has enough range in what it measures, that a higher score actually reflects more advanced knowledge and skills.

On the other hand, if a higher cut score simply means that a student can answer more of the same kinds of questions correctly, rather than questions tapping advanced concepts and skills, it is not very meaningful to raise the cut score.

Given the large numbers of students in some of the states who are not passing the current tests, it may not be easy for states to raise the bar higher. This may be especially true given the pressures of the No Child Left Behind Act to get all students to the “proficient” level on existing assessments. In the long run, however, states must be able to assure students and their parents that a high school diploma has currency in the real world — being proficient must mean that students are prepared for their next steps in life.

Build More Comprehensive Assessment Systems

Even if states improve their exit exams over time, it is unlikely that they will be able to measure everything that matters in a young person’s education. This is due, in part, to the fact that states administer these exams before the 12th grade. Basic fairness requires that students have multiple opportunities to take high school exit exams, so it is reasonable for states to begin to administer the tests in the 10th or 11th grades. It is also reasonable that the tests measure only content that has been taught by the time the exam is first given. This may explain why current exit exams are not measuring everything that matters to colleges and employers.

Ultimately, however, it is important for 12th grade students to be able to do 12th grade work, not just pass a 10th or 11th grade test. Over time, states will need to develop a more comprehensive set of measures, including but not limited to on-demand graduation tests.

States can take a number of different approaches to addressing this assessment gap.

- **Add 12th grade assessments that are well aligned to college and workplace knowledge and skills.** Rather than requiring students to pass these tests, the scores might be factored into course grades or simply included on high school transcripts. This would provide valuable information that postsecondary institutions and employers could use in making admissions, placement or hiring decisions.

- **Add end-of-course exams for subjects such as Algebra II or upper-level English language arts that are beyond the range of the exit exams.** Once again, these assessments could be factored into course grades and considered by postsecondary institutions and employers when making hiring, admissions or placement decisions.
- **Administer augmented college placement or admissions exams to students in the 11th and 12th grades.** As colleges already rely on tests such as the ACT, SAT and a handful of placement exams, states may want to explore the possibility of working with the companies that develop those exams to augment them so that they better align with state standards. That way, testing can become streamlined — one exam can produce information on how well students are meeting state standards while also providing valuable information to colleges.
- **Set a higher cut score on exit exams for college and work readiness.** States that have enough range in their high school exit exams may want to set two cut scores on the test — one that signifies eligibility for high school graduation and another, higher score that signifies readiness for college and the workplace. This is the strategy Texas currently is employing on its new test; the Texas Assessment of Knowledge and Skills serves as both a graduation exam and a college placement test.
- **Do not rely exclusively on large-scale assessments.** As critical as they are, large-scale assessments cannot measure everything that high school graduates need to know and be able to do. Making effective oral arguments, for example, is considered an essential skill by both employers and postsecondary educators. Both groups also stress the importance of conducting significant research projects. These skills — important aspects of the American Diploma Project benchmarks — are very difficult to assess on a paper-and-pencil test and may be better measured at the local and school levels. If states believe that all students should be responsible for mastering these essential skills, they should work with local districts to develop ways to incorporate research projects and oral examinations into instructional programs and to establish rigorous, systematic criteria for evaluating them across the state.

Regardless of the combination of assessments they use, states should conduct periodic studies of the predictive validity of the tests, particularly with respect to indicators of success in postsecondary education, such as grades, persistence and degree attainment. Such studies will help ensure and improve the validity of the assessments and the appropriate use of the information they generate.



Achieve launched this study to help answer some basic questions about the expectations states are setting for their high school graduates through the use of exit exams. Do the tests reflect material that students should be familiar with by the time they complete high school? Is it reasonable to expect all students to pass these tests before they graduate? If they pass these tests, does it mean students are ready for their next steps in life?

The tests do indeed set a floor for students that states can responsibly defend as a graduation requirement. If states stay the course with these exams and make the necessary investments to improve teaching and learning, they undoubtedly will find that their students will rise to the challenge. When they do, states should continue to raise the floor to reflect the demands students will face in postsecondary education and the world of work.

Appendix: Summary of Methodology

To compare assessments, each assessment item was analyzed and coded according to a range of lenses designed to capture different characteristics of individual test items and the tests as a whole. Many of the criteria in English language arts and math are similar, although there are important differences that stem from the distinct natures of the disciplines. To ensure the reliability of the data, at least two experts trained in the use of the criteria coded each test. Those experts reconciled any differences in coding before the data were analyzed.

The following are summaries of the various criteria according to which assessments in the study were analyzed. For the complete descriptions of the criteria, please visit Achieve's Web site at www.achieve.org.

Content Expectations of the Item

Mathematics

This lens compares the content of state mathematics exams, using the Third International Mathematics and Science Study (TIMSS) Mathematics Framework adapted for use in this study by the U.S. TIMSS National Research Center at Michigan State University and Achieve experts. The framework provides a detailed, comprehensive taxonomy of mathematics content, organized at its most general levels according to the following major domains of mathematics:

- Number
- Algebra
- Geometry/Measurement
- Data

These domains are further broken down into smaller units to allow for finer-grained comparisons. For example, geometry content is divided into a variety of categories such as two-dimensional geometry and measurement; three-dimensional geometry and measurement; transformations, congruence and similarity; and trigonometry. The majority of these categories are subdivided even further to facilitate a high degree of content specificity in coding. Item coders for this study assigned up to three primary content codes to each test item. In many cases, the multiple content codes aligned with the same reporting category (e.g., geometry/measurement or algebra), but this was not always the case. Items that aligned with more than one reporting category were re-examined, and one primary code was identified.

English Language Arts

The lens for English language arts provides a comprehensive taxonomy of the domains of reading, writing and language skills, using content descriptions developed by the Council of Chief State School Officers (CCSSO) and adapted by Achieve experts for use in this study. The CCSSO framework was developed in collaboration with several states that are a part of the Surveys of Enacted Curriculum.

A taxonomy that included all the aspects of English language arts as described in state standards, and therefore targeted in state tests, was developed to describe as accurately as possible the content or topic that each item measured. The study also required a taxonomy that was as specific as possible, providing sufficient discrimination among the topics to yield a clear portrait of what each state was emphasizing in its assessment of English language arts.

The major reporting codes for reading are:

- Basic comprehension (includes word definitions, main idea, theme and purpose)
- Literary topics (includes figurative language, poetic techniques, plot and character)
- Informational topics (includes structure, evidence and technical elements)
- Critical reading (includes appeals to authority, reason and emotion; validity and significance of assertion or argument; style in relation to purpose; and development and application of critical criteria)

The reporting categories for writing are:

- Writing (All items included in this category were *direct writing assessments*, typically writing in response to a prompt that asks students to address a particular question or thesis in a narrative, expository or persuasive essay. Although all such assessments included attention to language conventions, either as part of a holistic scale or as a discrete rubric, all direct writing tasks were coded to this category only and not coded as well to editing and revising or to grammar, mechanics and usage.)
- Editing and revising (*Items coded to this category assessed the following topics through multiple-choice items*: editing for conventions; organizing for meaning; and revising for meaning, style and voice.)
- Grammar, mechanics and usage (*Items coded to this category assessed the following topics through multiple-choice items*: spelling, mechanics, punctuation, syntax and sentence structure, grammatical analysis, and language usage.)

Approximate Grade-Level Demand of Each Item

Mathematics

This lens provides a scale to rate the grade level of the mathematics content included in state exams according to the TIMSS International Grade Placement (IGP) index. The IGP index represents a kind of composite among the 40 TIMSS countries (other than the United States) to show when the curriculum focuses on different mathematics content — at what point the highest concentration of instruction on a topic occurs. Using their nation’s content standards document, education ministry officials and curriculum specialists in each TIMSS country identified the grade level at which a math topic is introduced into the curriculum, focused on and completed. The IGP index is a weighted average of those determinations. For example, a topic with an IGP of 8.7 is typically covered internationally toward the end of 8th grade. The content topics to which Achieve coded test items all have an IGP value associated with them. For items that spanned more than one category and were subsequently assigned a single code, the retained content code tended to be that with the highest IGP value.

The following are examples of the IGP ratings of various mathematics content.

CONTENT DESCRIPTION	IGP INDEX
Whole Number: Operations	2.5
Rounding and Significant Figures	4.7
Properties of Common and Decimal Fractions	5.6
Exponents, Roots and Radicals	7.5
Complex Numbers and Their Properties	10.7

English Language Arts

To approximate the grade level demand for English language arts items, Achieve adapted the ACT Standards for Transition (for English language arts and reading), which provide a hierarchy of skills in these topic areas by taking into account the performance and content of an item as well as the reading demand of the reading passage being assessed. ACT, Inc.’s Educational Planning and Assessment System encompasses three assessments administered during 8th and 9th grades, 10th grade, and 11th and 12th grades. The Standards for Transition form the basis of all three, with each successive test including more complex content and performances from the standards. The standards are divided into six levels:

- Levels 1 through 4 are assessed on the EXPLORE test (8th and 9th grades);
- Levels 1 through 5 are assessed on the PLAN test (10th grade); and
- Levels 1 through 6 are assessed on the ACT Assessment (11th and 12th grades).

The following is an example of the most advanced three levels of one standard from the Reading Standards for Transition.

STANDARD	Comparative Relationships
LEVEL 4	Have a sound grasp of relationships between people and ideas in uncomplicated passages. Identify clearly established relationships between characters and ideas in more challenging literary narratives.
LEVEL 5	Reveal an understanding of the dynamics between people and ideas in more challenging passages.
LEVEL 6	Make comparisons, conclusions, and generalizations that reveal a feeling for the subtleties in relationships between people and ideas in virtually any passage.

Cognitive Demand of Each Item

Mathematics

This lens provides a taxonomy of performance expectations (what students are expected to “do” with the mathematics content they know) based on a synthesis of the TIMSS Mathematics Framework and Achieve’s assessment-to-standards alignment work with states. The five-point scale provides information on the kind and complexity of performance required of students — ranging from simple recall of information to complex reasoning skills.

- Level 1 includes demonstrating basic knowledge or recall of a fact or property.
- Level 2 includes routine problem-solving that asks students to do such things as compute, graph, measure or apply a mathematical transformation.
- Level 3 includes estimating, comparing, classifying and using data to answer a question or requiring students to make decisions that go beyond a routine problem-solving activity.
- Level 4 includes asking students to formulate a problem or to strategize or critique a solution method.
- Level 5 includes asking students to develop algorithms, generalizations, conjectures, justifications or proofs.

Coders often assigned multiple performance codes to items. Sometimes primary performance codes for an item spanned two or more of the reporting levels. In cases such as this, each item was re-examined, and a decision rule was made to accept the highest performance level category as representing the performance expectation of that item.

English Language Arts

The cognitive demand lens for English language arts provides a taxonomy of performance expectations based on Achieve’s assessments-to-standards alignment protocol and CCSSO’s description of performances in its Survey of Enacted Curriculum. Four levels of reading cognitive complexity provide information on the kind and complexity of reasoning required of students, ranging from simple recall of information to complex reasoning skills.

- Level 1, under the heading “Literal Recall,” covers such skills as providing facts, terms and definitions; describing ideas; locating answers in a text; identifying relevant information; and identifying grammatical elements.
- Level 2, under the heading “Infer,” covers such skills as inferring from local data, inferring from global data, drawing conclusions, identifying purposes, identifying main ideas or theme, identifying organizational patterns, and predicting.
- Level 3, under the heading “Explain,” includes such skills as following directions, giving examples, summarizing information, checking consistency and recognizing relationships.
- Level 4, under the heading “Analyze,” covers such skills as categorizing; distinguishing fact from opinion; ordering, grouping, outlining and organizing ideas; comparing and contrasting ideas; and interpreting techniques.

Demand of Reading Passages on the English Language Arts Tests

Achieve analyzed the difficulty level of each reading passage according to a six-point scale ranging from straightforward text to more complex, challenging and abstract text. This scale was developed by noted reading experts who reviewed various characteristics of passages, such as level or specialization of vocabulary, predictability of structures or organization, complexity of syntax, level of abstractness, familiarity of the topic, and the number of concepts introduced in the passage. Generally speaking, Level 1 represents upper-elementary reading levels, Levels 2 and 3 represent middle school-level reading, Level 4 represents early-stage high school reading, and Levels 5 and 6 represent late-stage high school reading.

Categories for consideration of reading passage difficulty include:

- **Structure**
 - Narration
 - Description
 - Explanation
 - Instruction
 - Argumentation
- **Vocabulary**
 - Poetic
 - Idiomatic
 - Technical
 - Unusual/unfamiliar
- **Syntax/connectives**
 - Dialogue
 - Sentence structure
- **Characters/ideas**
- **Narrator/stance**
- **Theme/message/moral**
- **Literary effects**
 - Foreshadowing
 - Flashback
 - Irony
- **Familiarity**
 - Topic
 - Place
 - Time period

For examples of reading passages at all six levels, please visit Achieve’s Web site at www.achieve.org.

Reading Rigor Index

The Reading Rigor Index (RRI) is a method of determining how the cognitive demand of an item interacts with the level of a reading passage. For example, an item could require a low performance of a difficult passage, a high performance of an easy passage, a high performance of a difficult passage or a low performance of an easy passage. This interaction of level of cognitive demand and reading level contributes to the challenge of an item. Items also are accounted varying point values. An item attributed one point is weighted less than an item attributed two or more points.

The RRI score is obtained by adding the cognitive demand level and the reading demand level for each reading item on a test. The Cognitive Demand Scale ranges from a low of one to a high of four and the Reading Level Demand Scale from a low of one to a high of six. This makes nine Reading Rigor levels — Level 1 for items with the lowest possible score of two (an item with a cognitive demand of one and a reading level demand of one); Level 9 for the highest possible score of 10 (an item with a cognitive demand level of four and a reading demand level of six).

An item's point value determines the number of times the RRI score is counted to determine RRI percentages. An item worth two points will have its RRI score counted twice, an item worth three points will have its score counted three times and so on.

"Cut Score"

Each state determines the levels of proficiency its students must reach to pass the state's exit exam based on scaled scores. The difficulty in comparing performance levels and the cut scores that reveal these levels is that these scaled scores are unique to each state's exam and students. Without a comparison sample — giving different state exams to the same group of students or giving a common exam to students in all six states — no connections among these scaled score distributions exist. Consequently, aside from a subjective analysis of proficiency-level setting procedures, it has been impossible to determine objectively if the proficiency levels set by different states have similar meaning.

Achieve, working with researchers from Michigan State University, has developed a procedure to establish comparability of proficiency levels across states according to the different dimensions by which the assessments analyzed in this study have been coded. Because the assessments from the six states were coded item by item according to common metrics, it became possible to compare what passing the assessments exactly at the cut score would mean, state to state. Achieve chose, in this study, to look at the mathematics cut scores through the IGP index lens and the English language arts cut scores through the ACT index (both are described above).

States almost universally use Item Response Theory (IRT) models to scale assessment items and to estimate a scaled value for each student. The cut score is established in this metric. Consequently, the cut scores (the scores needed simply to pass, not reach any level of greater proficiency) and scaling information provided by the states were used to determine sets of correctly answered items — or passing "scenarios" — that allow students to reach the cut score and the likelihood that those scenarios would occur. When coupled with the IGP (for mathematics) or ACT (for English language arts) codings of the items, the process transforms the cut scores into the corresponding IGP or ACT metrics. Comparisons of states' cut scores are done in these metrics. Because of the large number of potential passing scenarios (2^n where n is the number of items or points on the test), only a random sample of 20,000 passing scenarios were used for the computation.

American Diploma Project

Achieve, The Education Trust and the Thomas B. Fordham Foundation joined forces in 2001 to establish the American Diploma Project (ADP) to help states align their high school graduation assessments (in reading, writing and mathematics) with the increasing demands of higher education and the new economy of the 21st century. Through extensive research involving K–12 teachers, postsecondary faculty from two- and four-year colleges and universities, and front-line managers and human resource professionals from high-growth industries, ADP established a set of benchmarks in mathematics and English that reflect what a high school graduate needs to know and be able to do to be successful in higher education or the workplace.

In addition to establishing the set of skills and knowledge that enable a high school graduate to succeed in college or the workplace, ADP gathered examples of the tasks and assignments these graduates would face in both these arenas, which show how these skills and knowledge are put to use. The benchmarks and samples of work were released earlier this year in the report *Ready or Not: Creating a High School Diploma that Counts*.

For this study, the assessment items were compared to the ADP benchmarks to analyze how well the assessments measure the skills and knowledge that colleges and employers say are essential to success.

The following are the main categories of skills and knowledge in mathematics and English in the ADP benchmarks. The full set of ADP benchmarks, as well as the samples of workplace tasks and postsecondary assignments and complete ADP methodology, can be downloaded from the Achieve Web site at www.achieve.org.

Mathematics

- Number Sense and Numerical Operations
- Algebra
- Geometry
- Data Interpretations, Statistics and Probability

English

- Language
- Communication
- Writing
- Research
- Logic
- Informational Text
- Media
- Literature

Acknowledgments

From conception to completion, this study was an ambitious undertaking. Achieve would like to thank the individuals and organizations whose contributions made this work possible.

This study would not have been possible without the cooperation of the state education officials in the six participating states. These states agreed to be part of the study and were very forthcoming with the data and other information Achieve needed to complete the analyses.

The Prudential Foundation generously provided financial support for the work carried out here. Prudential Chariman and CEO and Achieve board member Arthur F. Ryan, together with Mary O'Malley and Lata Reddy, have been steadfast supporters of Achieve's benchmarking work.

The U.S. TIMSS National Research Center at Michigan State University partnered with Achieve on this project and provided considerable analytic and technical expertise. In particular, we would like to thank William H. Schmidt, Richard T. Houang, Kathleen A. Wight and ZhiWen Zou for their substantial contributions.

The team of people at Achieve who conceptualized and carried out the work on this study deserve particular thanks. Matthew Gandal, executive vice president, provided overall leadership and direction for the study and led the writing team who prepared the final report. JoAnne Eresh and Kaye Forgione, Achieve senior associates, helped create the analytic tools and led the research and analysis work in English language arts and mathematics, respectively. Susan Pimentel, consultant to Achieve, helped to shape the analysis and draft the final report. Mara Clark, research associate, provided ongoing research and editorial assistance, and Joseph Garcia, director of public leadership, provided editorial advice and assistance. Special thanks are also due to Adam Kernan-Schloss, Kathy Ames and their colleagues at KSA-Plus Communications, Inc., for their editorial advice and their help designing this report.

Finally, we would like to thank the following individuals for their advice and assistance during the research phase of this study:

Arthur Applebee
*University at Albany,
State University of New
York*

Gail Burrill
*Michigan State
University*

Pat DeVito
*Assessment and
Evaluation Concepts,
Inc.*

Carol Jago
*Teacher, Santa Monica
High School/California
Reading and Literature
Project*

Sandra Murphy
*University of California,
Davis*

Steve Newman
*Northern Kentucky
University*

Charlie Peters
University of Michigan

Sharif Shakrani
*National Assessment
Governing Board*

Uri Treisman
*The University of Texas
at Austin*

Hung-Hsi Wu
*University of California,
Berkeley*

Gene Young
*Sam Houston State
University*

Richard Askey
University of Wisconsin

Sue Eddins
*Teacher, Illinois
Mathematics and
Science Academy*

Jerry Halpern
*Consultant/Retired
High School Teacher*

Sally Hampton
*National Center on
Education and the
Economy*

Linda Hunt
*Teacher, Chicago City
Schools*

Paul Sally
University of Chicago

Paul Sally III
*Teacher, Chicago City
Schools*

Harold Asturias
University of California

Mary Crovo
*National Assessment
Governing Board*

Eunice Greer
*American Institute for
Research*

Andy Porter
Vanderbilt University

Lynn Steen
St. Olaf College

Michael Cohen
President, Achieve, Inc.



Achieve, Inc.
1775 Eye Street, NW
Suite 410
Washington, DC 20006
www.achieve.org