

# How Does the Hawaii High School Assessment Measure Up?

A Comparison of the  
2005 Grade 10 Hawaii State  
Assessment in Reading and  
Mathematics with High School  
Graduation Exams from Other States

### **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 more than half the states benchmark their academic standards, tests and accountability systems against the best examples in the United States and around the world. Achieve also serves as a significant national voice for quality in standards-based education reform and regularly convenes governors, CEOs and other influential leaders at National Education Summits to sustain support for higher standards and achievement for all of America's schoolchildren.

Copyright © 2006 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.

# How Does the Hawaii High School Assessment Measure Up?

A Comparison of the  
2005 Grade 10 Hawaii State  
Assessment in Reading and  
Mathematics with High School  
Graduation Exams from Other States

## Table of Contents

<b>Executive Summary</b>	<b>1</b>
<b>I. Background</b>	<b>6</b>
<b>II. How does Hawaii’s policy regarding its high school assessment compare with that of other states?</b>	<b>9</b>
Table 1: State policy context for high school assessments (9)	
<b>III. How do Hawaii’s tests compare with the graduation exams from other states?</b>	<b>10</b>
<b>Reading</b>	<b>10</b>
<i>Test Features</i>	
Table 2: Distribution of items and points on reading section of seven state tests (10)	
<i>Item Type</i>	
Chart 1: Percentage distribution of points by item type (11)	
<i>Content of Reading Passages and Questions</i>	
Chart 2: Distribution of points by reading passage genre (12)	
Chart 3: Percentage of points within reading passage genres (13)	
Chart 4: Genre of reading passages: Hawaii, NAEP and other states (14)	
Chart 5: Reading point distribution by item content (15)	
Table 3: Reading point distribution within item content categories (15)	
<i>Cognitive Demand of Reading Passages and Questions</i>	
Chart 6: Average reading passage demand (16)	
Table 4: Reading passage demand distribution (17)	
Chart 7: Inference cognitive demand levels (18)	
Table 5: Distribution of points by low and high cognitive demand (19)	
Chart 8: Distribution of points by cognitive demand, Hawaii versus six-state average (19)	
<i>Rigor of the Assessment</i>	
Chart 9: Content on reading tests on ACT scale, Hawaii versus six-state average (21)	
Table 6: Content on reading tests on ACT scale (21)	
Chart 10: Content on reading tests on ACT scale, Hawaii versus Ohio and Florida (22)	
Chart 11: Reading rigor levels (23)	
Chart 12: Reading rigor profile, Hawaii versus six-state average (24)	
<i>Summary of Findings</i> (24)	
<b>Mathematics</b>	<b>25</b>
<i>Test Features</i>	
Table 7: Distribution of item points on mathematics assessments (25)	
Chart 13: Distribution of points by item type (26)	
<i>Content of Items</i>	
Chart 14: Distribution of points by content strand (27)	
Table 8: Distribution of points by strand (27)	
Chart 15: Distribution of algebra items by category (29)	
Table 9: Distribution of points by content: Geometry/measurement (30)	
<i>Rigor of the Assessment</i>	
Chart 16: Average IGP (32)	
<i>Cognitive Demand</i>	
Table 10: Distribution of points by level of cognitive demand (33)	
Chart 17: Distribution of points by level of cognitive demand (34)	
Chart 18: Average cognitive demand by strand, Hawaii versus six-state average (35)	
<i>Summary of Findings</i> (35)	

## Table of Contents *(continued)*

<b>IV. How do the performance levels on Hawaii’s grade 10 assessments in reading and mathematics compare with those of other states?</b>	<b>36</b>
<i>Methodology</i>	<b>36</b>
<i>Performance Levels on the Reading Test</i>	<b>36</b>
Chart 19: Difficulty of average passing scenario at “Meets Proficiency” cut score (ACT scale) (37)	
<i>Performance Levels on the Mathematics Test</i>	<b>38</b>
Chart 20: Average difficulty of mathematical content at “Meets Proficiency” cut score (International Grade Placement Scale) (38)	
<b>V. Conclusion</b>	<b>42</b>
<b>Appendix: Summary of Methodology</b>	<b>45</b>



## Executive Summary

In 2004, Achieve launched a study to compare the graduation exams in six states — Florida, Maryland, Massachusetts, New Jersey, Ohio and Texas. The goal of this study was 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 resulting report — *Do High School Graduation Exams Measure Up?* — compared the content and rigor of the exams and the cut scores students need to achieve to pass the tests. Across the states, we found that the tests do indeed set a floor for students that can be responsibly defended as a graduation requirement, but they do not effectively tap the higher-level skills that truly constitute “readiness” for college and work.

In 2005, Achieve was asked by the Hawaii Department of Education to compare its 2005 grade 10 Hawaii State Assessment in reading and mathematics with the six states’ exams employing the same methodology used in the initial study. Although Hawaii does not require students to pass its grade 10 assessments to graduate, such a comparison is nonetheless a useful exercise. Because the states that participated in the larger study together enroll nearly a quarter of the nation’s high school students, they provide a solid reference point for Hawaii as it works to strengthen its grade 10 assessments over time.

Specifically, the state asked Achieve to:

- Analyze the content of the grade 10 assessments in reading and mathematics;
- Determine how well the assessments measure the skills necessary to succeed in college and work;
- Analyze what it takes to pass the grade 10 assessments; and
- Compare the content and rigor of the assessments to those of other states.

### Findings for Hawaii

After a detailed analysis of Hawaii’s grade 10 assessments in reading and mathematics, Achieve found significant differences between the Hawaii tests and those of other states. While the reading test is generally less demanding than those of other states, the mathematics test contains considerably more challenging content than tests from other states. Nonetheless, neither assessment is overly rigorous. Indeed, Hawaii — like other states — will need to develop

assessments for use after grade 10 to help ensure that graduating seniors are on track for success in postsecondary institutions and today's workplace.

Achieve's major findings are as follows:

### **Reading**

- **Hawaii's reading test puts a premium on comprehension of informational text, which is exactly what colleges and employers say is essential for success in college-level courses across the curriculum and in the workplace.** The reading test underscores the importance of students being able to comprehend informational text by selecting passages that are mostly exposition (writing contained in textbooks, articles, reports and manuals), rather than literary prose and poetry, and developing test questions that probe students' ability to decipher informational text. Hawaii's emphasis is appropriate. As research by Achieve's American Diploma Project (ADP) has demonstrated, whether students are bound for college or the workplace, they must be able to understand informational text.
- **The passages on the Hawaii reading test tend to be less challenging than the reading passages on most of the other tests we analyzed, making the Hawaii test among the least rigorous.** Hawaii's reading test overall is among the least rigorous of the tests we analyzed, mainly because the reading passages on the test are of relatively low cognitive complexity, generally representative of upper middle school and early high school level reading. Moreover, while Hawaii's test questions are more cognitively demanding on average than those in most other states, the questions fall short in not requiring students to analyze text beyond a superficial level, a skill critical for success in college and today's work environment.
- **Hawaii's "Meets Proficiency" score in reading is comparable to the average score of the other states' passing levels.** As was true of the other states in Achieve's prior study, Hawaii students can pass the reading test with knowledge and skills that ACT considers more appropriate for the test it gives to 8th and 9th graders than for its college admissions test.

### **Mathematics**

- **Hawaii's test in mathematics is well balanced and contains rigorous content. The test gives more emphasis to geometry, algebra and data items than to number items, as is appropriate for a grade 10 test.** Moreover, the content of the test is more rigorous than all but one of the states analyzed in Achieve's earlier study due to Hawaii's relatively large proportion of "Advanced Algebra" items — bringing the test closer to the demands of college and work. This is particularly important in light of mounting evi-

dence that indicates that Algebra II is fast replacing Algebra I as the gatekeeper course for success in college and the high-skills workplace.

- **Hawaii’s “Meets Proficiency” score in mathematics requires students to know slightly more challenging content than students who scored at the passing level on the other tests Achieve analyzed.** While the content demand on the Hawaii assessment in mathematics is higher than those in other states, the test items themselves are less cognitively demanding on average than those of the six other states in the study, particularly for items assessing number and geometry content. Thus, the test as a whole does not present unreasonable expectations. In fact, from an international perspective, to pass the mathematics test, Hawaii students need to successfully answer questions that, on average, cover material students in most other countries study by grade 8.

### Recommendations for Improvement

Achieve’s analysis of six states’ graduation exams indicated that states must continue to raise the bar on their exit exams over time.

These findings hold true for Hawaii, and we recommend that the state:

- **Raise the overall rigor of the grade 10 reading test.** Hawaii should increase the complexity of the reading passages on its assessment. Some passages on the reading test should represent the level of demand typical of instructional materials written for a late high school reading level to raise the ceiling on the test and signal the level of text students need to comprehend to be on track for attainment in postsecondary education and the new economy. In addition, Hawaii should add items that measure the highest level of cognitive demand, requiring students to more deeply analyze text.
- **Phase in higher cut scores on the reading test over time.** In addition to increasing the cognitive demand of its passages and/or items, Hawaii can raise the rigor of its reading test over time by raising the score required for passing. Texas is using this approach with its new graduation exam. This strategy only works if a test has enough range in what it measures, so that a higher score actually reflects more advanced knowledge and skills. If a higher cut score simply means that students must answer more of the same kinds of items correctly, rather than items tapping more advanced concepts and skills, it is not very meaningful to raise the cut score.
- **Raise the level of performance demand of the mathematics items.** Although the content on the math test is challenging, the items themselves tend to make low-level demands in terms of performance. Hawaii should raise the

cognitive demand of its items over time by increasing the proportion of items that require complex problem-solving skills, problem formulation and advanced reasoning.

As it continues to raise expectations, Hawaii must continue to invest in improving teaching and learning by implementing exemplary instructional materials, by releasing its tests, as Massachusetts does, or by posting assessment exemplars on its Web site. Hawaii also should develop diagnostic, formative assessments for classroom use; align professional development with its tests; and provide extra support to struggling students. Doing so will undoubtedly result in Hawaii's producing an ever-increasing number of proficient students, well prepared for the rigors of postsecondary education and the realities of a global economy.

### **Beyond the Grade 10 Assessment**

Tests administered in the 10th grade cannot fully capture the range of content that students study in high school. Over time, Hawaii will need to go beyond its grade 10 test to develop a more comprehensive set of assessments that measure the full set of knowledge and skills that indicate readiness for college and work. One possible approach is for Hawaii to develop end-of-course tests for subjects such as Algebra II or upper-level English that are beyond the scope of its 10th grade tests. Such tests could be factored into course grades or included on high school transcripts, and they would provide valuable information that postsecondary institutions and employers could use in making admissions, placement or hiring decisions.

Finally, as critical as assessments are, they cannot measure everything that matters to a young person's education. The ability to make effective oral arguments and conduct research projects are considered essential skills by employers and postsecondary educators alike, but these skills are not well assessed by a paper-and-pencil test. To ensure that these important skills are measured, Hawaii will need to work with local districts to establish a systematic method for evaluating them across the state.



Achieve undertook its original 2004 study to provide educators, policymakers and the public with a clearer picture of what high school graduation exams measure and how difficult they are to pass. Although Hawaii does not require students to pass its grade 10 assessment to graduate, its assessment, like those of the other six states, sets a reasonable floor for students. Therefore, Achieve strongly encourages Hawaii not to lower its expectations. Rather, Hawaii should stay the course and ratchet up the level of demand of these assessments over time, and it should extend the high school assessment system to measure college- and work-ready skills.

To help the state accomplish this, Achieve encourages Hawaii to consider joining its ADP Network, a group of 22 states that have pledged themselves to a policy agenda in support of truly preparing students for success in college and work by the time they graduate from high school. Each state has committed to the following four actions.

- Aligning high school standards with the knowledge and skills required for success in college and work;
- Requiring all high school graduates to take challenging courses that actually prepare them for life after high school;
- Streamlining the assessment system so that the tests students take in high school also can serve as readiness tests for college and work; and
- Holding high schools accountable for graduating students who are ready for college or careers, and holding postsecondary institutions accountable for students' success once enrolled.

Although the Network has been in existence for just over a year, Achieve has already seen evidence of substantial progress on the part of participating states.

## I. Background

In June 2004, Achieve published an analysis of the graduation exams in six states — Florida, Maryland, Massachusetts, New Jersey, Ohio and Texas. The study — *Do High School Graduation Exams Measure Up?* — compared the content and rigor of the tests, as well as the scores that students needed to pass those tests.

The Hawaii Department of Education asked Achieve to undertake a study that would compare the 2005 grade 10 Hawaii State Assessment in reading and mathematics to the six state graduation exams based on the methodology employed in the larger study. It is important to note that Achieve's analysis is not an alignment study of how closely Hawaii's grade 10 assessment measures its standards. Achieve's alignment studies take somewhat different criteria into account and are based on a different methodology. Rather, this analysis is intended to explore the content and rigor of the tests in comparison to other tests and against common external benchmarks.

### Why Achieve Initiated the Study of Graduation Exams

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. More rigorous than an earlier generation of minimum competency tests initiated in the 1980s, these tests are a significant part of the decade-long movement to raise standards and improve achievement in the United States. They also have become a lightning rod for public debate.

The attention exit exams have received is understandable and deserved. They are the most public example of states' holding students directly accountable for reaching higher standards. For the most part, however, the public debate over high school exit exams has gone on without vital information about how high a hurdle they actually present to high school students.

Achieve launched its 2004 study to provide educators, policymakers and the public with a clearer picture of what high school graduation exams measure and how difficult they are to pass. The states that participated in the study together enroll nearly a quarter of the nation's high school students, making them an ideal point of comparison for Hawaii.

Achieve's methodology builds from describing the attributes and dimensions of individual test items, to grouping the items in meaningful categories, to identifying patterns and making comparisons among states. To ensure consistency in the way items are described, Achieve develops coding schemes

for each dimension and trains expert reviewers in their use. To ensure reliability, two reviewers independently code each individual item and reconcile any differences in judgment before assigning final characterizations. Categorizing items allows us to construct an overall representation of each state test and to make cross-state comparisons.

In general, the dimensions Achieve examines help unpack the content (what students need to know) and the level of performances (what students are expected to do with their knowledge) for each assessment. For instance, in terms of mathematics content, it is important to determine the kind of algebra being assessed — the proportion of algebra items that target prealgebra topics as opposed to topics typically addressed in Algebra I or Algebra II courses. In reading, for example, we are concerned with how much of each test is dedicated to assessing informational text as opposed to literary topics.

In analyzing content, Achieve uses independently devised benchmarks, particularly in estimating the grade level of particular content. In mathematics, we use an international scale created as part of the Third International Mathematics and Science Study (TIMSS, which is now known as the Trends in International Mathematics and Science Study). In English, we use a scale adapted from a scale devised by ACT, Inc., to describe questions on its college preparatory and admissions tests.

Judging the complexity of the performance or cognitive demand of each item is as important as characterizing its content. In mathematics, for example, students can be provided a formula and simply required to plug in appropriate values, or they can be required to reason their way through a problem and solve it by selecting the applicable formula from a chart that is provided. The cognitive demand of reading items also can vary across a wide range: At one end of the spectrum, students can be asked to apply a relatively low-level skill, such as locating information in a text, while at the other end, they can be expected to perform a far more intellectually demanding task, such as making generalizations by synthesizing information across different passages.

Reading presents a unique situation since test questions usually are based on one or more passages on the test, and both passages and questions can run the gamut from easy and straightforward to difficult and complex. In the end, it is the interplay of the items with the passages on which they are based that establishes the rigor of a reading test. To address this critical dynamic, Achieve developed a Reading Rigor Index (RRI), which is fully explained in the appendix.

Finally, Achieve's exit exam analysis investigated what it takes for students to pass each state test and how those expectations compare across states. Achieve and experts from Michigan State University devised a statistical approach to allow cut scores from different states' tests to be compared. Mathematics tests were compared on the TIMSS scale, and reading tests on a scale adapted from ACT's skills hierarchy scale. Using this methodology, Achieve was able to identify those items that students who scored at the cut score were likely to answer correctly and to determine the knowledge and skills encompassed by those items. This procedure helped us show how challenging each state test was to pass, relative to the other state tests included in the study.

Since completing its original analysis of the six state exit exams, Achieve has continued to refine the dimensions it uses to characterize mathematics items, as well as reading items and passages. More information about the methodology used in this analysis appears in the appendix.

## II. How does Hawaii's policy regarding its high school assessment compare with that of other states?

Hawaii and the six states that participated in Achieve's study of graduation exams have made different policy choices about the stakes they attach to the tests, the timing of their tests, the schedule of implementation and the subjects tested.

First and foremost, Hawaii, unlike the other six states, does not require students to pass its high school assessment in reading and mathematics to earn a high school diploma. Hawaii, like Florida, Massachusetts and Ohio, gives its tests for the first time to 10th graders. New Jersey and Texas give their exit exams in the 11th grade, while Maryland has created end-of-course exams, with the English exam given as early as the end of 9th grade. These states are also at different points in the rollout of the assessments. In Florida, Massachusetts and New Jersey, the tests already count for high school students, while in Maryland, Ohio and Texas, the tests will count in the future. Finally, states also test different subject areas (see Table 1).

**Table 1: State policy context for high school assessments**

HAWAII	Florida	Maryland	Massachusetts	New Jersey	Ohio	Texas
<b>TEST</b>						
Hawaii Grade 10 Assessment in Reading and Mathematics	Florida Comprehensive Assessment Test	Maryland High School Assessments	Massachusetts Comprehensive Assessment System	High School Proficiency Assessment	Ohio Graduation Tests	Texas Assessment of Knowledge and Skills
<b>GRADE FIRST GIVEN</b>						
10th	10th	End of Course	10th	11th	10th	11th
<b>YEAR FIRST GIVEN</b>						
2002	1998	2001	1998	2002	2003	2003
<b>REPLACED ANOTHER EXIT TEST</b>						
Yes	Yes	Yes	No	Yes	Yes	Yes
<b>SUBJECTS TESTED</b>						
Reading; writing; mathematics	Reading; mathematics	English I; algebra/data analysis; biology; government	English language arts; mathematics	English language arts/literacy; mathematics	Reading; writing; mathematics; science; social studies	English language arts; mathematics; science; social studies
<b>FIRST GRADUATING CLASS REQUIRED TO PASS</b>						
Not required	2003	2009	2003	2003	2007	2005
<b>OPPORTUNITIES FOR STUDENTS WHO HAVE NOT PASSED TO RETAKE TESTS</b>						
Not applicable	Yes	Yes	Yes	Yes	Yes	Yes
<b>OTHER POLICIES RELATED TO STAKES</b>						
No tests are required for graduation.	Students are permitted to substitute results on SAT or ACT to meet graduation requirements.	Students can fail any one subject assessment and still meet requirements by earning a high combined score across all four assessments.	Appeals process uses statistical comparison of GPAs in subject area courses of passing and non-passing students.	State currently provides alternative, performance-based assessment given and scored locally, which will be phased out by 2011.	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.

### III. How do Hawaii’s tests compare with the graduation exams from other states?

#### Reading

##### TEST FEATURES

The breakdown of questions, in terms of format, on Hawaii’s reading assessment, compared to that of the states participating in Achieve’s exit exam study, is shown in the table below.

Table 2: Distribution of items and points on reading section of seven state tests\*

State	Total questions	Number of points
HAWAII	40 multiple choice 5 constructed response	40 12
Florida	44 multiple choice 8 constructed response	44 27
Maryland	20 multiple choice 2 constructed response*	20 8
Massachusetts	34 multiple choice 4 constructed response*	34 16
New Jersey	20 multiple choice 4 constructed response*	20 16
Ohio	31 multiple choice 7 constructed response	31 20
Texas	28 multiple choice 3 constructed response*	28 9

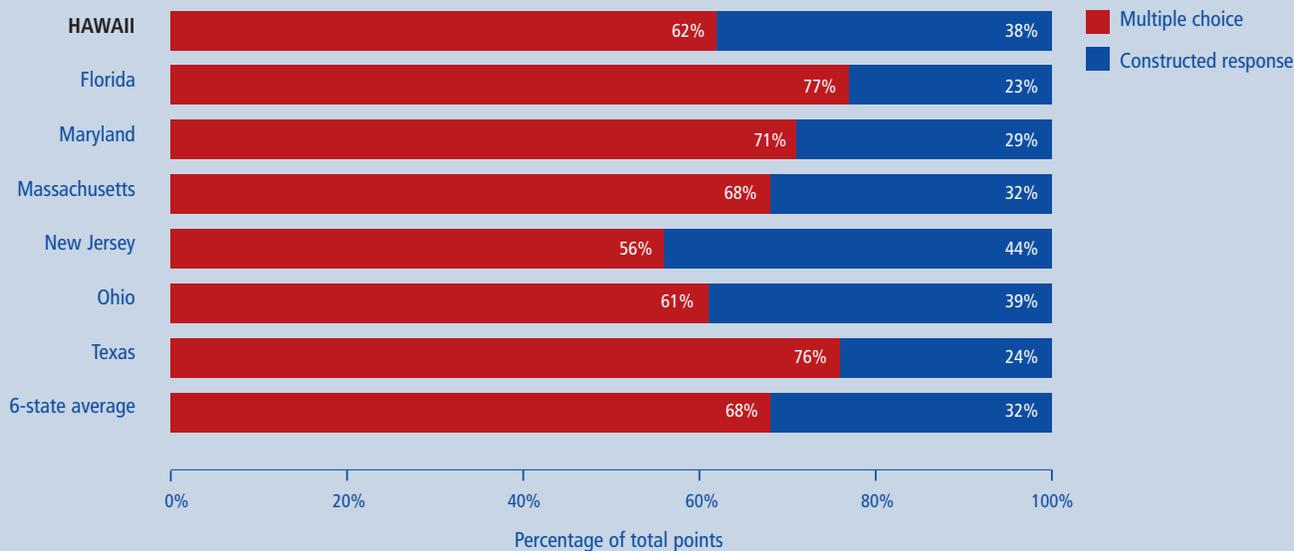
\*In Achieve’s 2004 study, four states — Maryland, Massachusetts, New Jersey and Texas — included the direct assessment of writing as part of their exit exams and combined the reading and writing items into a single score. To ensure comparability in this study, we base our comparison of Hawaii’s reading assessment on only the reading portions of the six states’ tests.

##### ITEM TYPE

***Hawaii allots a greater proportion of its reading assessment to constructed-response items than do most of the states in Achieve’s study.***

Hawaii’s high school assessment in reading awards a higher percentage of its points (38 percent) to constructed-response items than other states do on average (32 percent) and than any other state does except Ohio (39 percent) (see Chart 1). The inclusion of a high proportion of constructed-response items is a strength of Hawaii’s reading assessment, since the state uses the format appropriately and to good advantage in assessing more complex knowledge and skills. As indicated above, Maryland, Massachusetts, New Jersey and Texas include a direct assessment of writing as part of their graduation tests, while Hawaii administers a separate writing test. Since the publication of Achieve’s initial study of graduation exams, Ohio has introduced a separate writing test.

Chart 1: Percentage distribution of points by item type



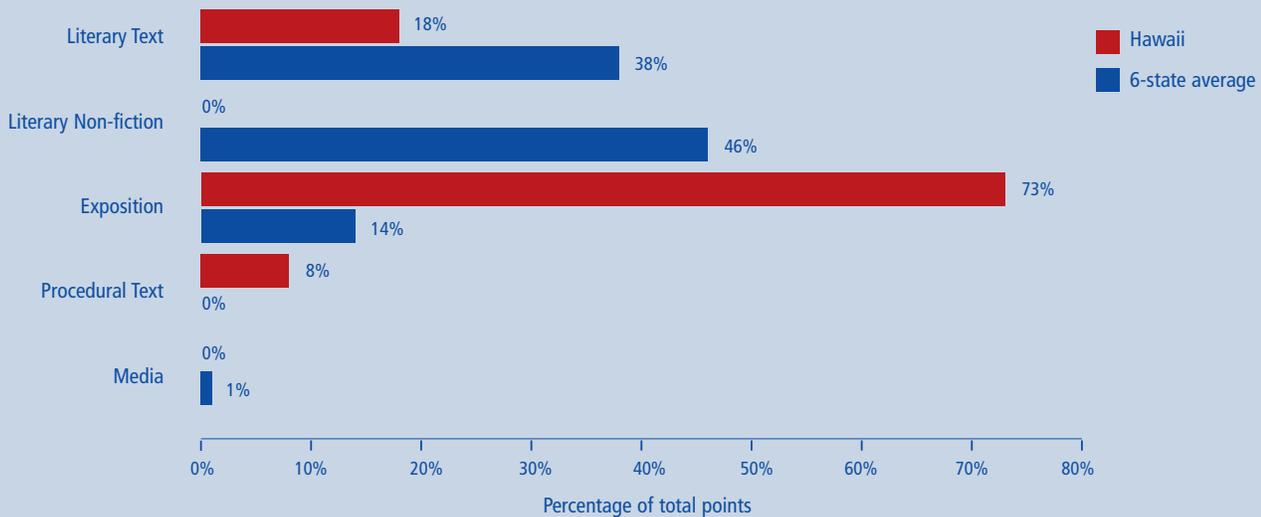
## CONTENT OF READING PASSAGES AND QUESTIONS

### *Hawaii's reading test consists mainly of informational passages.*

Achieve's ADP found that both college professors and employers stress the necessity of high school graduates being able to comprehend a wide range of informational materials, such as periodicals, memoranda, manuals, technical reports, and intricate charts and graphs. States select reading passages from a variety of genres in constructing their reading tests, and there is no set pattern that states follow. The following genres are included on one or more of the tests Achieve analyzed: literary text (short story/novel excerpt, poetry or drama), literary non-fiction (essay, autobiography/biography or literary speech), exposition (news story or textbook/informational article), procedural text/document (directions or manual) and media (photograph or advertisement). None of the tests, including Hawaii's, contains passages with graphic elements, such as charts or diagrams, and none contains passages that the National Assessment of Educational Progress (NAEP) characterizes as "argumentation or persuasive."

In sharp contrast to what postsecondary institutions and workplaces say is required for success, the six states' graduation tests that Achieve examined dedicated most of their passages and most of their test points to literary text and literary non-fiction. That is not the case with Hawaii: The state wisely prioritizes understanding expository text. In awarding 73 percent of its test points to items assessing students' comprehension of exposition, Hawaii is far above the average (14 percent) of the other six states. Moreover, Hawaii is the only state to include a procedural (a "how to") passage on its assessment. Following instructions to perform specific tasks, answer questions or solve problems, such as troubleshooting the failure of an appliance, is an important skill for students to develop.

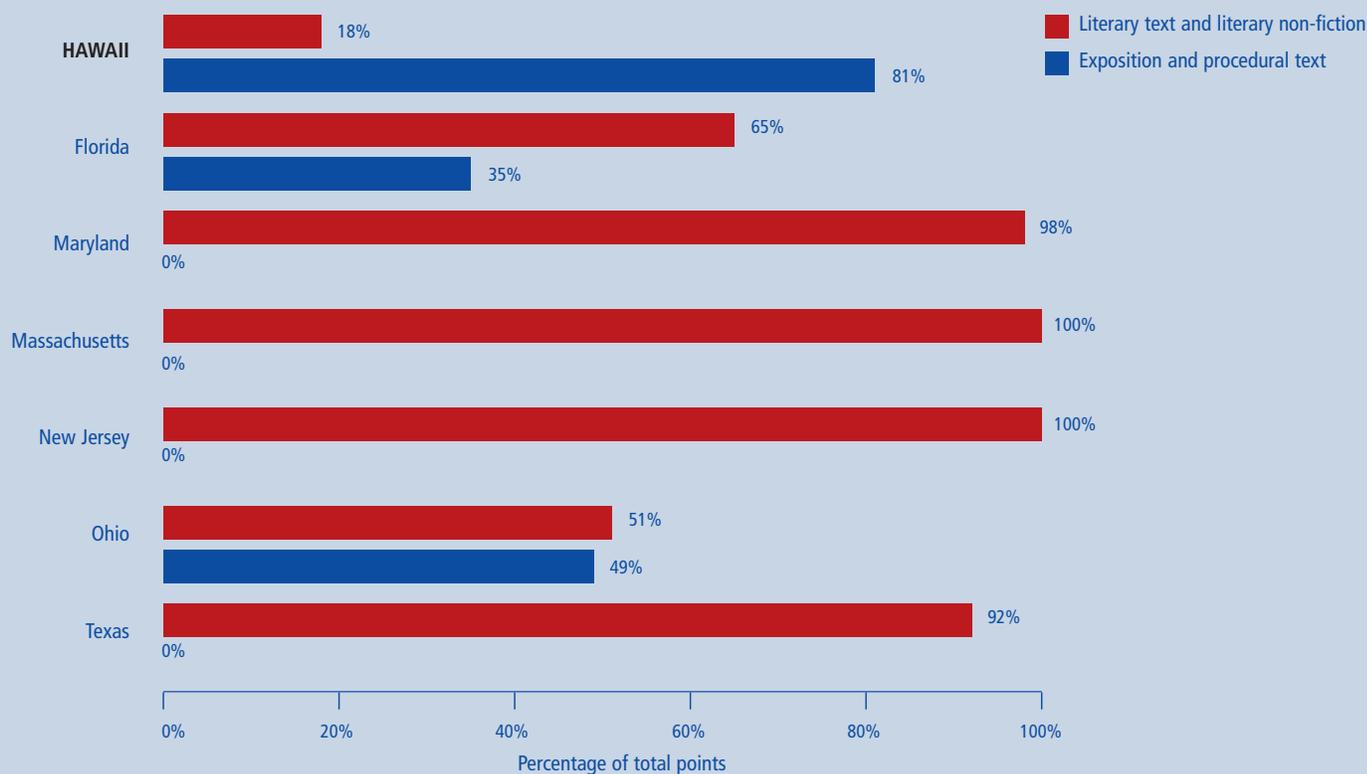
Chart 2: Distribution of points by reading passage genre



*\*Achieve has revised the genre classifications used in its 2004 study to more closely reflect those adopted by NAEP in its 2009 Reading Framework. We include media, although NAEP does not, to characterize state choices of genre as accurately as possible.*

Four of the six states in the exit exam study — Maryland, Massachusetts, New Jersey and Texas — focused their tests on literary text and literary non-fiction passages, as the distributions in Chart 3 indicate. Ohio and Florida dedicated a significant proportion of their points to exposition and procedural text, but not to the degree that Hawaii does.

Chart 3: Percentage of points within reading passage genres

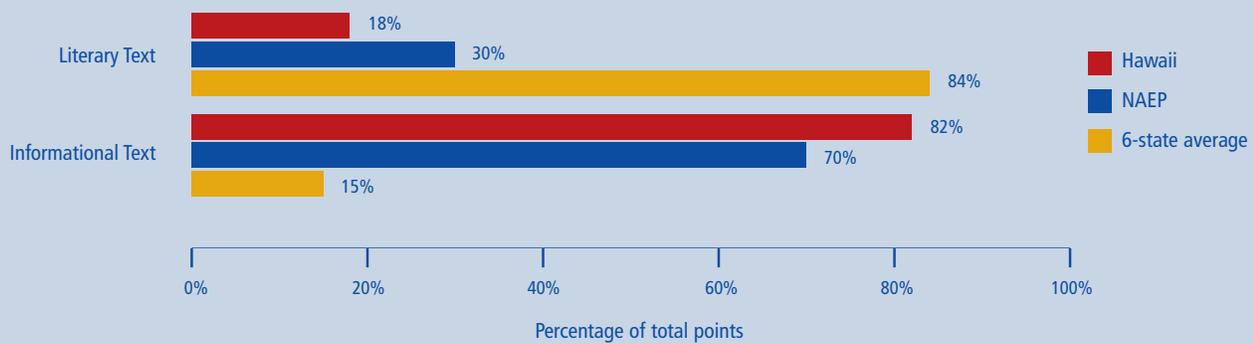


NOTE: Totals may not equal 100 percent due to rounding and to the exclusion of a third genre group, Media and Graphics, which appears only on the Maryland and Texas tests.

The 2009 NAEP Reading Framework reinforces what college professors and employers advocate. It requires 70 percent of its reading passages to be informational — a category that includes procedural text — and just 30 percent of passages to be literary. The upward shift from the previous framework’s requirement of 60 percent informational passages sends a clear signal to states regarding the kinds of skills students will need for success in postsecondary education and work. Hawaii’s reading test, in fact, exceeds NAEP’s stipulation by more than 10 percent, while all other states in the exit exam study are well

below the guideline. With 80 percent of its reading passages devoted to exposition and procedural text, Hawaii’s focus is promising, but the state should continue to include narrative passages, often a dominant genre in English classrooms, to target literary comprehension as well as the understanding of informational text.

**Chart 4: Genre of reading passages: Hawaii, NAEP and other states**

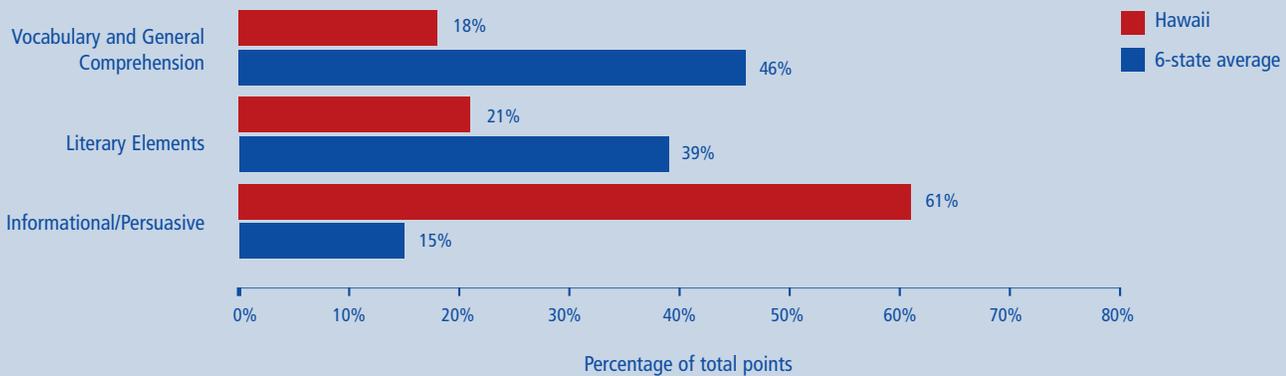


NOTE: Totals may not equal 100 percent due to rounding.

***Hawaii’s test questions emphasize comprehension of informational text, whereas the other states’ tests tend to focus on general comprehension, vocabulary and literary elements.***

While the process of selecting test passages and items is iterative, a state’s choice of genres for its reading passages clearly influences the content of its test questions. In view of Hawaii’s emphasis on expository text and its inclusion of a procedural passage, it is not surprising that the state allocates the majority of its test points to assessing students’ understanding of informational content. Rather than concentrating on vocabulary and comprehension in general — as the six states on average do — Hawaii’s test questions zero in on proficiency in comprehending informational text in particular. This finding is not characteristic of the other states. On average, the six states’ assessments allot 46 percent of total test points to fundamental reading comprehension topics (e.g., general comprehension of a word, phrase or paragraph and understanding the main idea of a selection) and just 15 percent to understanding informational and persuasive topics. Hawaii’s assessment follows a markedly different pattern in allotting just 18 percent of its test points to general comprehension and vocabulary topics, while dedicating 61 percent to comprehension of informational content.

**Chart 5: Reading point distribution by item content**



Hawaii’s awarding just 18 percent of its points to items that assess vocabulary and general comprehension differs markedly from the other states, where the range extends from 32 percent (Texas) to 61 percent (New Jersey). In keeping with its choice of expository text for the bulk of the reading passages on its assessment, Hawaii gives less weight to assessing literary elements (21 percent of test points) than five of the comparison states give; only Florida assigns even less weight (8 percent) to literary elements. In dedicating 61 percent of its points to assessing informational content, Hawaii towers above every state other than Florida.

**Table 3: Reading point distribution within item content categories**

Category	HI	FL	MD	MA	NJ	OH	TX
Vocabulary and General Comprehension	18%	35%	52%	50%	61%	45%	32%
Literary Elements	21%	8%	46%	50%	33%	31%	65%
Information/Persuasive	61%	58%	2%	0%	6%	24%	3%

Hawaii’s emphasis on informational text is prescient and in general accord with both Achieve’s ADP findings and NAEP’s latest recommendations. Having made that statement, we offer a note of caution: ADP devotes one set of its English benchmarks to literature, and the NAEP Reading Framework (2009) requires that 30 percent of its passages be literary text (i.e., fiction, literary non-fiction and poetry). In addition, Hawaii’s own high school standards in language arts label standards 1–4 as “Reading and Literature” and include a number of performance indicators calling

for students to read, understand and analyze literature. Hawaii should fine-tune the balance of its passages and test questions to reflect the national guidelines more closely and strengthen the alignment with its own standards.

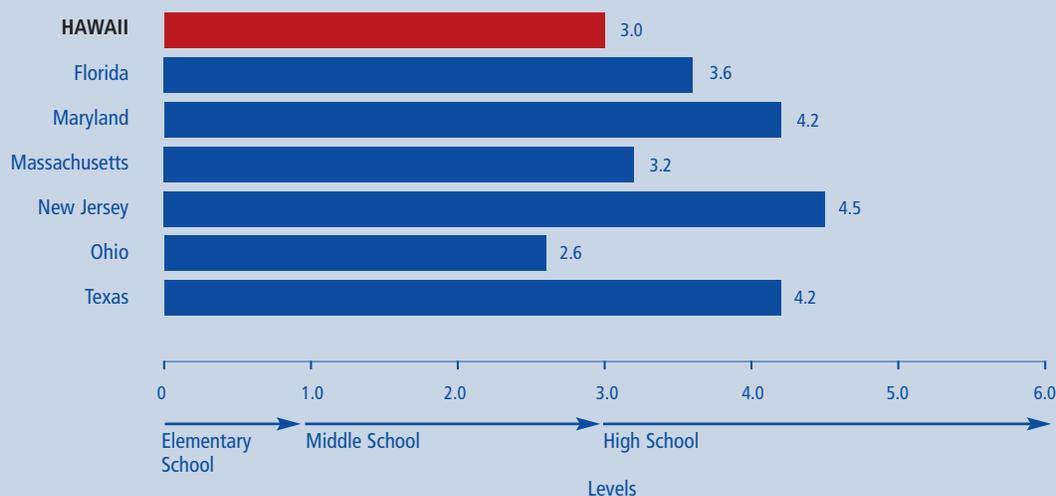
### COGNITIVE DEMAND OF READING PASSAGES AND QUESTIONS

*Hawaii's reading passages are less demanding than those found on the other tests Achieve examined. The state's reading assessment does not include any passages that represent the level of demand typical of instructional materials written at a late high school level.*

To judge the complexity of reading passages, Achieve's reading experts created a six-point scale describing texts from relatively simple to quite complex. 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 topic, and the number of concepts introduced in the passage. Level 1 represents upper-elementary reading, Level 2 and Level 3 represent middle school reading, Level 4 represents early-stage high school reading, and Level 5 and Level 6 represent later-stage high school reading.

The average demand of Hawaii's reading passages clusters at, and rarely extends beyond, Level 3. The reading passages employed by Texas and Maryland have an average demand at Level 4, and New Jersey's demand stretches toward Level 5.

Chart 6: Average reading passage demand



A closer look at the breakdown of selected states' passages by level of demand is instructive. As noted earlier, Hawaii's reading assessment most resembles Ohio's and Florida's tests in the attention paid to informational text, as opposed to literary text. However, Ohio's and Florida's tests contain reading *passages* with a higher level of demand than Hawaii's. The demand of Hawaii's reading passages centers at Level 3, while Ohio's extends to Level 4 and Florida is centered at Level 4. The remaining states' passages tend to focus on the upper levels of the reading demand they present to students. While it is fair to note that New Jersey's and Texas' tests are administered in grade 11, one would expect to see more passages on Hawaii's assessment that target Level 4, and some passages that target Level 5 and Level 6.

**Table 4: Reading passage demand distribution**

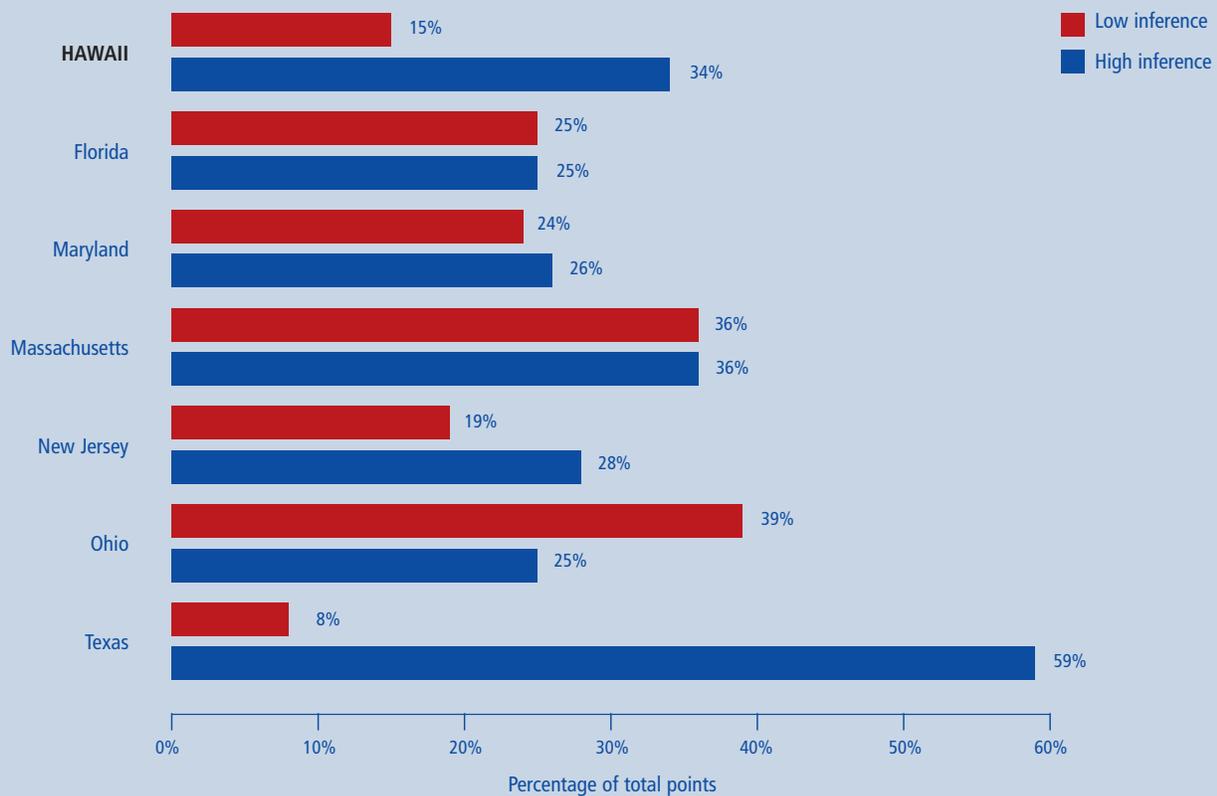
Level	HAWAII	Florida	Ohio	6-state average
1	0%	0%	14%	5%
2	4%	0%	37%	10%
3	94%	37%	25%	29%
4	0%	63%	24%	33%
5	1%	0%	0%	12%
6	0%	0%	0%	11%

***The questions on Hawaii's reading assessment are generally challenging, requiring students to go beyond the level of fundamental comprehension to making complex inferences and generating explanations.***

Since completing its analysis of state graduation tests in 2004, Achieve found that a finer delineation of the cognitive demand of reading questions could be had by splitting the category of inference into two categories — low versus high. The resulting five categories are Literal Recall, Low Inference, High Inference, Construct and Analyze.

Inference items require students to deduce a meaning that is not explicitly stated in the text. Low inference items require students to make simple deductions — for example, identifying the main idea of an uncomplicated piece of text. High inference items are more cognitively taxing, requiring students to make more subtle deductions — for example, identifying the theme of a complex literary narrative. Achieve applied the distinction between low and high inference to the six states that participated in the original study, re-analyzing the 2004 data for the six states to obtain the results shown in Chart 7.

Chart 7: Inference cognitive demand levels



One effect of refining the category of inference is that it helps reveal a special strength of Hawaii’s reading assessment. To be specific, in comparing the cognitive demand of Hawaii’s questions to each of the six states studied, we find that Hawaii allocates a greater proportion of its test points to the high inference category than do most of the states.

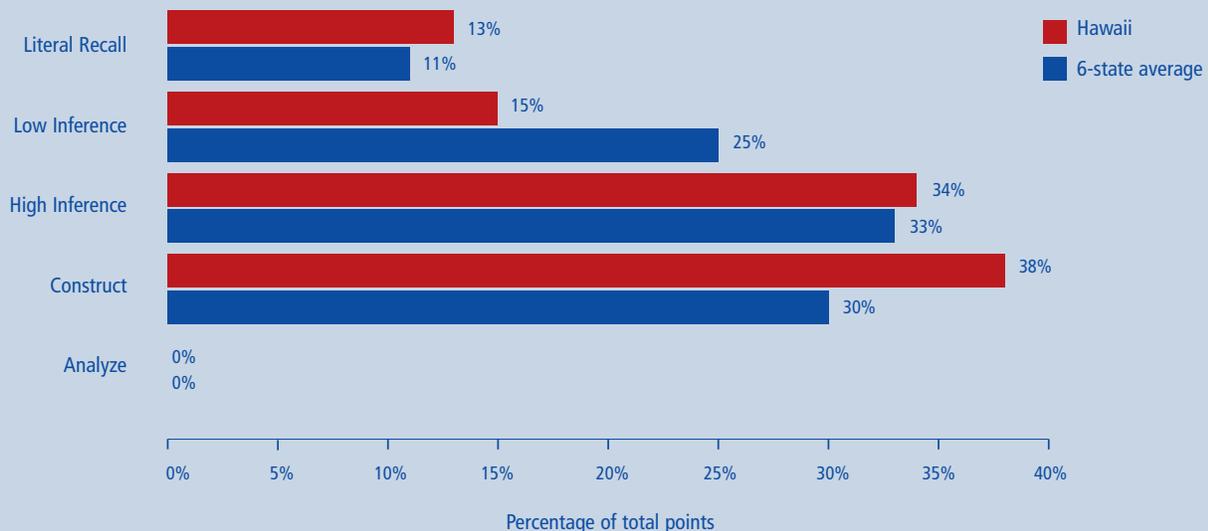
However, to fully appreciate how the cognitive demand of the items on Hawaii's reading test compares with that of the other states, it is important to examine the states' overall distribution of item points in the three most cognitively demanding categories — High Inference, Construct and Analyze. Hawaii's reading test allocates a total of 72 percent of item points to these upper categories, while states, on average, allocate 64 percent. Only two states — Texas at 86 percent and New Jersey at 75 percent — allocate a greater percentage of test points to the most cognitively demanding categories, and only New Jersey includes questions at the Analyze level.

**Table 5: Distribution of points by low and high cognitive demand**

Level of cognitive demand	HI	FL	MD	MA	NJ	OH	TX	6-state average
Literal Recall and Low Inference	28%	52%	42%	36%	25%	47%	14%	36%
High Inference, Construct, Analyze	72%	48%	58%	64%	75%	53%	86%	64%

Hawaii's lack of analytical questions may, in part, stem from the state's having fewer literary passages on its reading assessment than the other states studied. Achieve has found that on large-scale state assessments, items calling for analysis tend to address narrative reading passages more frequently than informational reading passages. However, this need not be the case. College preparatory tests, such as the ACT Assessment and SAT Reasoning Test, include items at this higher level of cognitive demand for both narrative and informational texts.

**Chart 8: Distribution of points by cognitive demand, Hawaii versus six-state average**



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

## RIGOR OF THE ASSESSMENT

*As was true of the other states studied, most questions on Hawaii's reading test target skills meant to be taught and learned by grade 8 or 9.*

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 originally created by ACT, Inc., to characterize its series of assessments. The index is a composite scale that takes into account the content of a test, the cognitive demand of reading passages and the cognitive demand of questions. ACT established six levels to differentiate the knowledge and skills that are measured on its reading assessments: Levels 1 through 4 cover knowledge and skills found on ACT's EXPLORE test, which is given in the 8th or 9th grade; ACT's PLAN test, which is given in the 10th grade, includes test items from Level 1 through Level 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 items from Level 1 through Level 6.

As is clear from Table 6, none of the states' tests, including Hawaii's, approaches the level of demand that the ACT says is characteristic of its college admissions test. On the contrary, the vast majority of points (84 percent) across the six states link to ACT Level 1 through Level 4, meaning the level of demand across the six tests most closely resembles that of the ACT EXPLORE test — which is given to students in 8th and 9th grades. This finding holds true for Hawaii's reading test as well; its profile reflects that of the ACT's EXPLORE test.

A close look at the structure of Hawaii's test shows it has a smaller percentage of points at Level 5 and Level 6 (1 percent) than the average of the six states (14 percent) but a significantly higher percentage of Level 4 items than any of the other six states. The latter finding stems mainly from the fact that ACT assigns higher levels of demand to questions about expository text and that genre is featured in Hawaii's test. None of Hawaii's questions, however, qualify as Level 5 because the related reading passages — as indicated by Hawaii's profile on Achieve's six-point scale of passage complexity — lack the complexity required to generate more cognitively demanding questions.

Chart 9: Content on reading tests on ACT scale, Hawaii versus six-state average

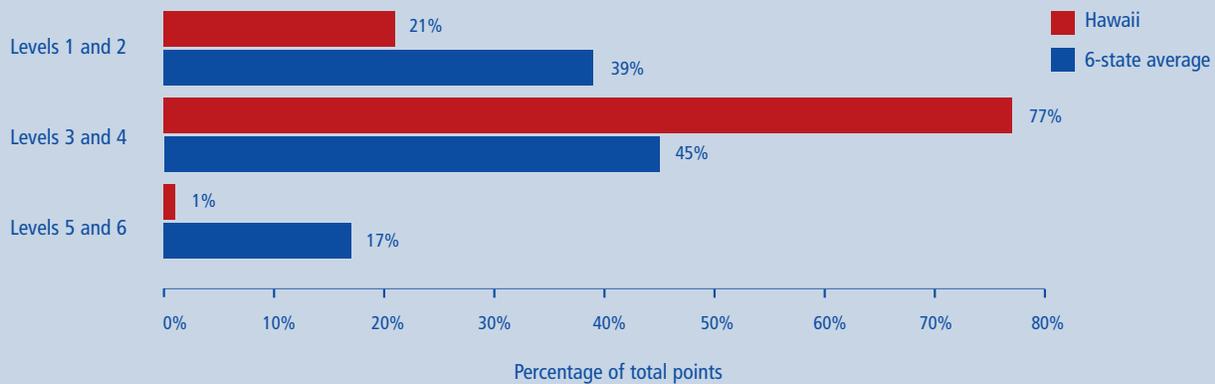


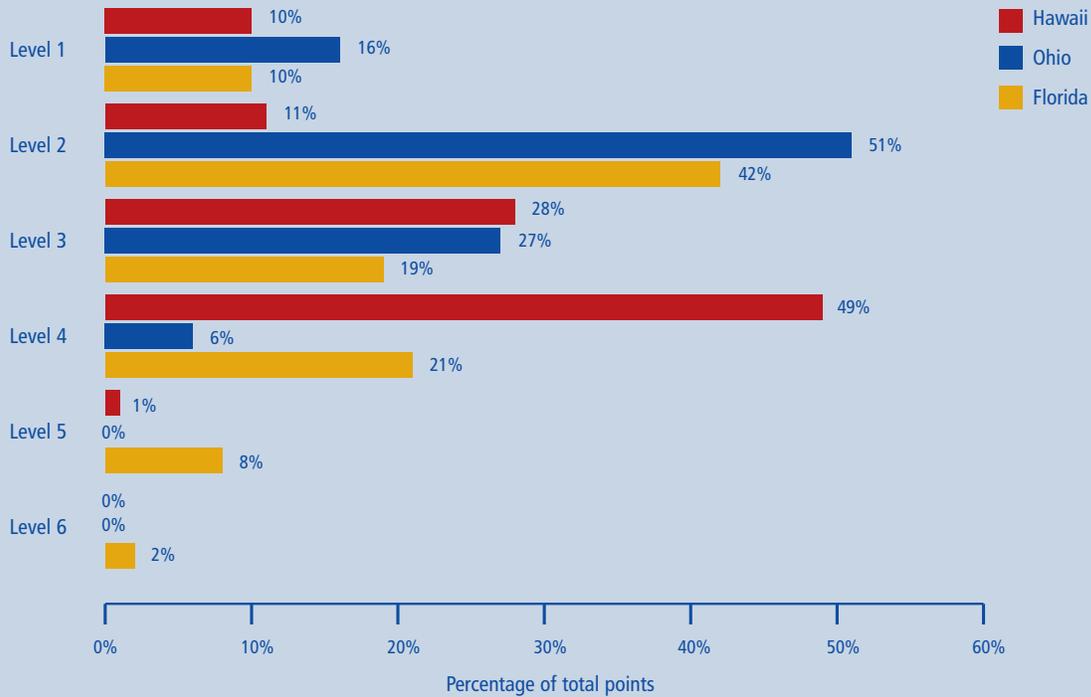
Table 6: Content on reading tests on ACT scale

Level	ACT EXPLORE (8th and 9th grades)	ACT PLAN (10th grade)	ACT Assessment (11th and 12th grades)	HI	FL	MD	MA	NJ	OH	TX	6-state average
1	10–20%	5–15%	5–15%	10%	10%	24%	7%	3%	16%	3%	10%
2	20–30%			11%	42%	24%	21%	22%	51%	11%	29%
3	30–40%	20–30%	10–20%	28%	19%	20%	11%	19%	27%	24%	20%
4	15–25%	20–30%	20–30%	49%	21%	26%	32%	25%	6%	38%	25%
5	0%	25–35%	25–35%	1%	8%	6%	29%	31%	0%	24%	16%
6	0%	0%	20–30%	0%	2%	0%	0%	0%	0%	0%	0%

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

As noted earlier, of the six comparison states, Hawaii’s assessment most closely resembles those of Ohio and Florida in the emphasis given to informational text. However, Hawaii’s assessment ends up with a higher level of cognitive demand because it contains more challenging questions than do Ohio’s and Florida’s tests. This trait helps offset Hawaii’s relatively undemanding reading passages.

Chart 10: Content on reading tests on ACT scale, Hawaii versus Ohio and Florida



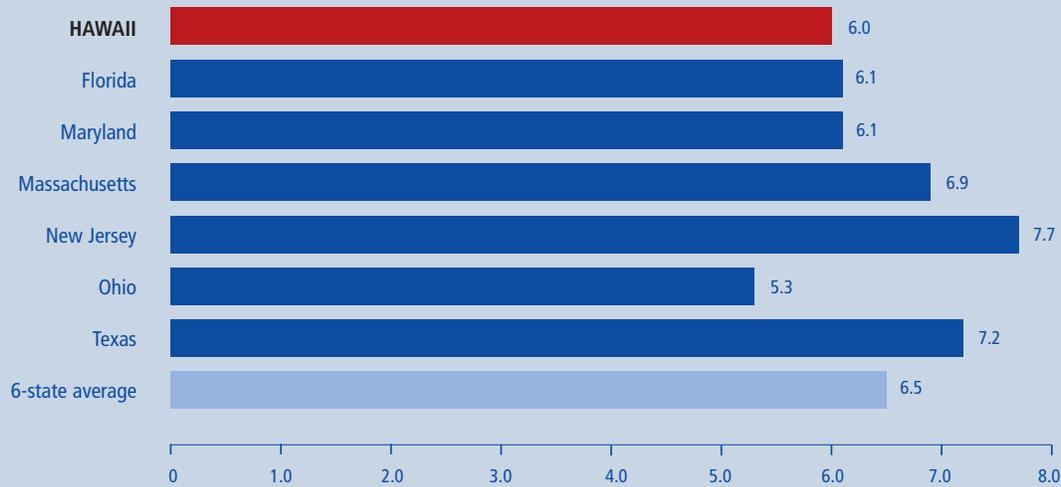
NOTE: Totals may not equal 100 percent due to rounding.

***The overall rigor of Hawaii’s reading test is below that of most of the other six states.***

The difficulty of a reading test is determined both by the complexity of the reading passages and by the cognitive demand of the questions about those passages. To capture this important interaction, 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. (Note: Cut scores are not factored into the RRI. See appendix for more information on the RRI.)

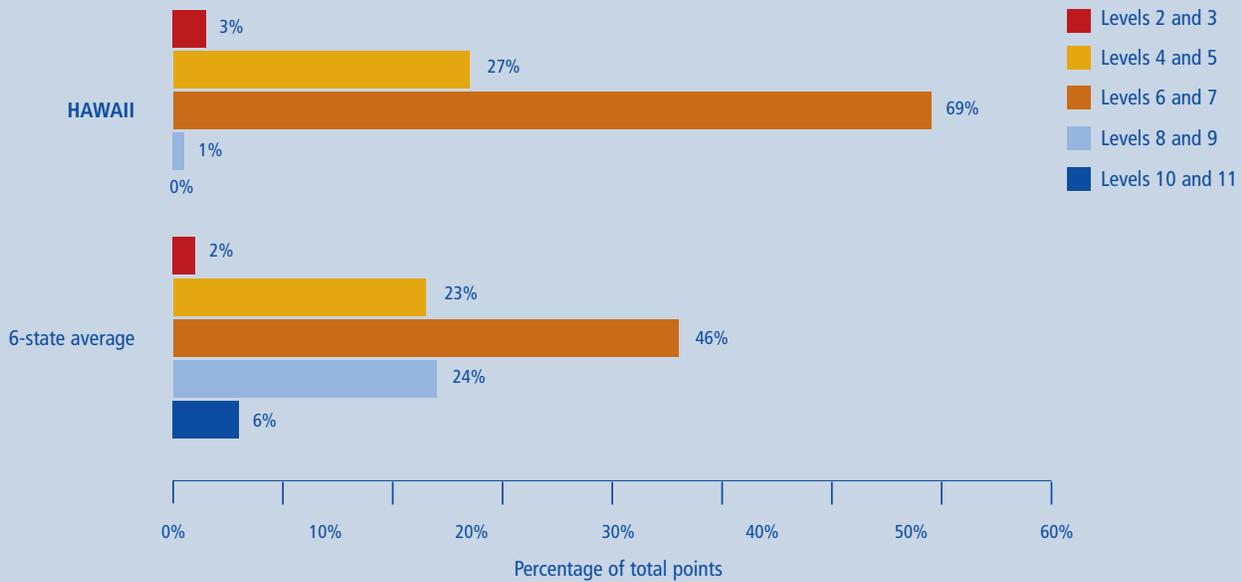
On this interactive scale, Hawaii, with an average index of 6.0, falls a bit below the six-state average of 6.5. The Hawaii, Florida and Maryland tests are roughly equivalent in terms of reading rigor, and all three tests are more rigorous than Ohio’s test. The New Jersey and Texas tests are the most rigorous, followed by Massachusetts’ test. It is worth noting that the two most rigorous tests — Texas and New Jersey — are given in the 11th grade, whereas the rest are 10th grade tests, except for Maryland’s, which is an end-of-course test.

Chart 11: Reading rigor levels



Hawaii structures its reading test quite differently than other states do on average, as is evident when we compare the overall reading rigor profile of Hawaii's test with the average profile of the comparison states. Like the other states' tests, Hawaii's shows an increase in percentage of questions with higher demand in the interplay of its passages and questions at the lower levels of the index, but then it abruptly tops out at an index of seven. In fact, unlike the other states' tests, 99 percent of the items fall between Level 3 and Level 7 on the index. The remaining states on average show a more symmetrical pattern of passage and question interaction. This is due to the low level of demand of the reading passages on the Hawaii test, relative to those on the other states' tests.

Chart 12: Reading rigor profile, Hawaii versus six-state average



**SUMMARY OF FINDINGS**

In some ways, the Hawaii reading assessment is on par with the six state tests in Achieve’s original study. In terms of its approximate grade level (as measured by the ACT scale), Hawaii fares similarly to the six comparison states. All of the tests have a level of demand that most closely approximates ACT’s EXPLORE test, which is given to students in grades 8 and 9. And the Reading Rigor Index indicates that Hawaii’s test is about as cognitively demanding as Maryland’s and Florida’s tests and is more demanding than Ohio’s test (although it is less demanding than Texas’ and New Jersey’s tests).

Nonetheless, the overall rigor of the Hawaii State Assessment in reading is less challenging than the six states in Achieve’s original study of graduation tests, due to its relatively undemanding reading passages. In the end, it is the low demand of the passages and not the items themselves that reduces the rigor of Hawaii’s test.

## Mathematics

### TEST FEATURES

Hawaii’s test contains 77 items, more than any other test we examined. It is administered in four separate sections, totaling roughly three hours of actual testing time. This is slightly above the norm: Most other state tests we analyzed take two to two and a half hours. For the Hawaii test, Achieve examined 49 core items — worth a total of 75 points — that contribute to a student’s “Meets Proficiency” level score. Of the 28 items that Achieve did not examine, seven were field-test items, and 21 were SAT-9 items used to provide a norm-referenced score for each student. The breakdown in terms of item format of Hawaii’s mathematics assessment, as compared to that of the participating states in Achieve’s study, is shown in the table below. Achieve’s constructed-response item category includes both short-response and extended-response items. It is also important to note that Hawaii prohibits the use of calculators on its state assessment, unlike the six states in Achieve’s graduation test study, all of which permitted students to use calculators for all or part of their tests.

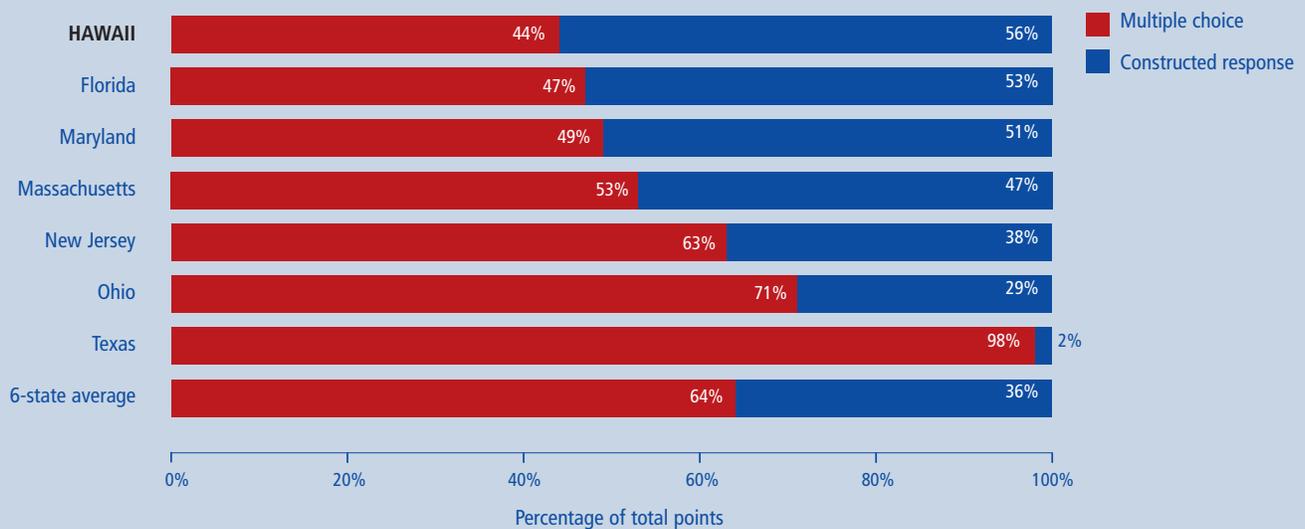
Table 7: Distribution of item points on mathematics assessments

State	Points	Total	Total time	Testing time	
HAWAII	33 multiple choice 42 constructed response	75	4 hours, 58 minutes	2 hours, 58 minutes	
Florida	28 multiple choice 32 constructed response	60	—	2 hours, 30 minutes	
Maryland	26 multiple choice 27 constructed response	53	3 hours	2 hours, 30 minutes	
Massachusetts	32 multiple choice 28 constructed response	60	untimed	2 hours (suggested)	
New Jersey	30 multiple choice 18 constructed response	48	2 hours, 26 minutes	2 hours	
Ohio	32 multiple choice 13 constructed response	45	—	2 hours, 30 minutes	
Texas	59 multiple choice 1 constructed response	60	untimed	untimed	

**Hawaii's high school assessment in mathematics allots more points to items in a constructed-response format than do other states' tests.**

On average, the states Achieve studied allotted 36 percent of the points on their graduation tests to items having a constructed-response format and 64 percent to those having a multiple-choice format. In contrast, Hawaii devotes a significantly larger proportion of test points to constructed-response items (56 percent) and a correspondingly smaller proportion to multiple-choice items. Only two states in Achieve's exit exam analysis approach Hawaii's distribution: Maryland assigns 51 percent and Florida assigns 53 percent of their test points to constructed-response items. Hawaii's decision to emphasize constructed-response items is a strength of its assessment. Generally speaking, the state used the item format as it is meant to be used, that is, for measuring critical thinking and reasoning skills and for solving multistep problems.

**Chart 13: Distribution of points by item type**



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

## CONTENT OF ITEMS

*Hawaii's assessment represents a good balance across the content domains; it gives more emphasis to algebra, geometry/measurement and data and less to number, as is appropriate for a high school test.*

To get a picture of the mathematics content state tests measure, Achieve categorizes the distribution of test points according to the discipline's four domains — number, algebra, geometry/measurement and data analysis. On average, the six states in our original study awarded the majority (66 percent) of the points students could earn to algebra and geometry/measurement (32 percent and 34 percent respectively), followed by data (19 percent) and number (15 percent). Hawaii's assessment follows a similar overall pattern in devoting 63 percent of the possible points to algebra and geometry/measurement. However, Hawaii awards slightly more points to algebra (39 percent) and slightly less to geometry/measurement (24 percent) than the other states do on average. In comparison to the average of the other states, Hawaii also allocates a larger proportion of points to data (29 percent as compared with 19 percent) and a smaller proportion of points to number content (8 percent as compared with 15 percent).

Chart 14: Distribution of points by content strand

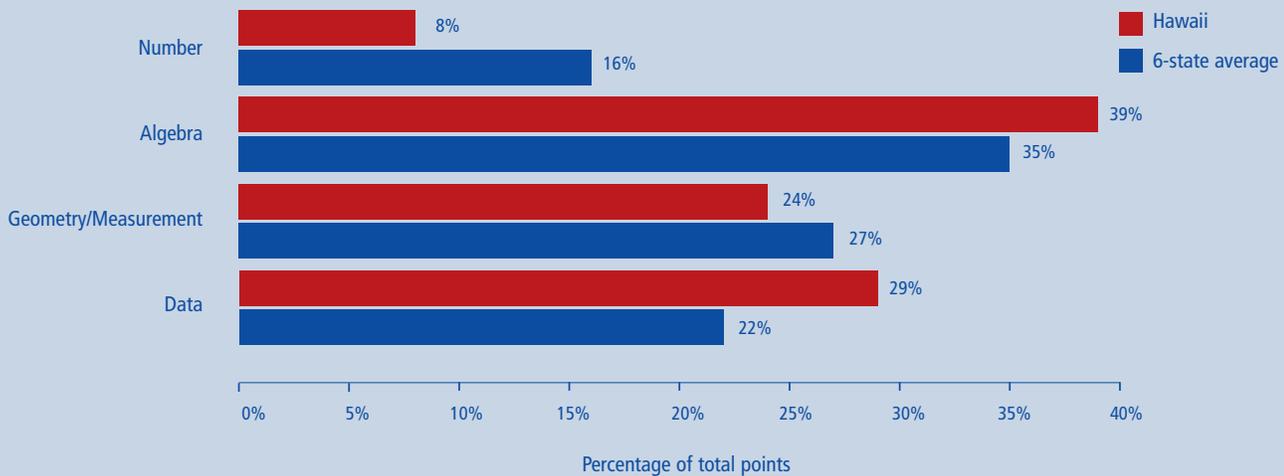


Table 8: Distribution of points by strand

Strand	HI	FL	MD	MA	NJ	OH	TX	6-state average
Number	8%	25%	8%	13%	27%	11%	10%	16%
Algebra	39%	25%	42%	40%	25%	27%	50%	35%
Geometry/Measurement	24%	35%	6%	27%	23%	40%	33%	27%
Data	29%	15%	45%	20%	25%	22%	7%	22%

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

In comparing the distribution of points by strand for each state, we find that Maryland's and Texas' tests stress algebra to an even greater degree than does Hawaii's test, although Maryland's emphasis on algebra is attributable to its being an end-of-course test for Algebra I and Data Analysis. It is also worth noting that Hawaii gives less emphasis to number (8 percent of its points) than any of the other states. Because number topics tend to be learned at earlier grade levels, Hawaii's lack of emphasis on this strand adds to the overall rigor of the assessment.

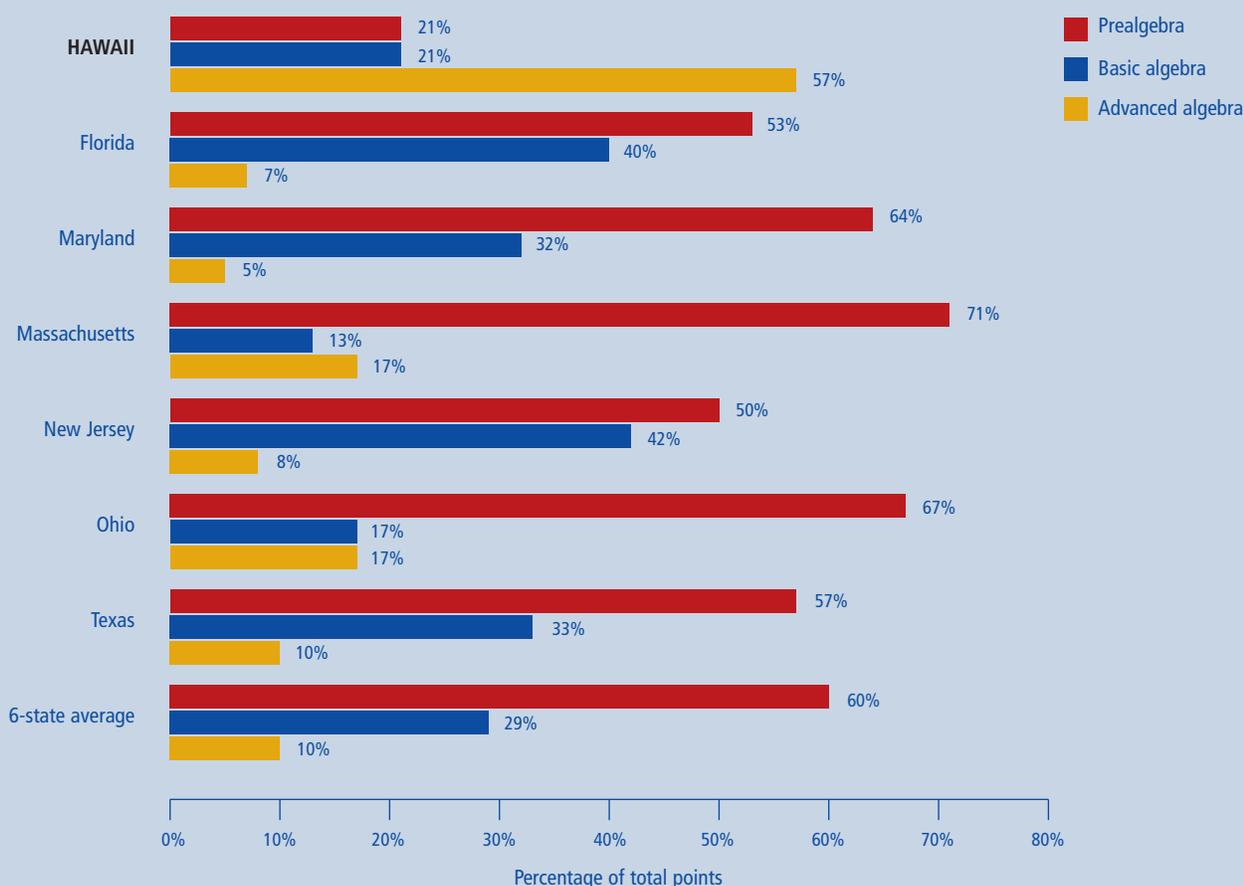
Hawaii is second only to Maryland in percentage of points devoted to the data strand, which is noteworthy given the importance of data in today's world. Maryland's emphasis is due to the fact that its test is specifically intended to assess algebra *and* data analysis, which is atypical of state graduation exams.

***Hawaii's test includes more advanced algebra topics than any other state test in this study.***

Because algebra is a prerequisite for success in credit-bearing college mathematics courses and today's high-skills workplace, Achieve closely examined the specific algebra content being assessed on each state test. Across the six states that participated in our original study, we found that a majority of the algebra points students can earn are associated with less-demanding algebra topics. In fact, the six states, individually and on average, dedicated 50 percent or more of their algebra points to assessing prealgebra concepts that most students learn prior to high school. These include such basic skills as working with integers, rational numbers, patterns, representation, substitution, basic manipulation and simplification. The six states, on average, assign less than one-third of the total of their algebra points to concepts such as linear equations and basic relations and functions that are typically associated with basic algebra or Algebra I — a course commonly taken in the 9th grade or in many cases even earlier. Moreover, the states allocate an even smaller proportion of the algebra points (10 percent on average) to assessing advanced algebra concepts, such as non-linear functions, equations, inequalities, and work with real and complex numbers. These concepts are typically encountered in Algebra II and generally considered essential for success in credit-bearing college mathematics courses.

Compared with the six states that participated in Achieve’s original study, Hawaii assigns a significantly greater proportion of its items to assessing advanced algebra concepts. In fact, a majority of Hawaii’s algebra points (57 percent) come from items that assess advanced algebraic understandings, greatly exceeding the six-state average of just 10 percent. For example, the Hawaii assessment includes items that extend beyond linear functions and equations to include the non-linear — especially quadratics. Students also are called on to solve systems of linear equations and to display their understandings of real and complex number systems. These are also the kinds of topics that distinguish grade 12 NAEP from grade 8 NAEP.

Chart 15: Distribution of algebra items by category



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

**Hawaii gives the same weight to two-dimensional geometry as the other states do, on average.**

Achieve found that 54 percent of the geometry/measurement points on the state tests were associated with two-dimensional geometry and measurement, with the exception of New Jersey, which favored this area of study with 82 percent of its geometry points. Hawaii, on track with the other five states, allots 56 percent of its geometry points to two-dimensional geometry. States gave significantly less attention — 19 percent on average of geometry test points — to three-dimensional geometry that includes concepts such as volume and surface area. Geometry tends to be less hierarchical than algebra, so two-dimensional geometry is not necessarily less challenging than three-dimensional geometry. It is worth noting, however, that NAEP includes two-dimensional geometry and measurement on its 8th grade assessment, but it includes formal three-dimensional geometry on its 12th grade assessment, indicating that three-dimensional geometry is considered to be end-of-high school level content. Measurement, including such concepts as units and estimation of measurements, was not a major focus of any of the six graduation tests Achieve analyzed, nor was trigonometry. Hawaii, however, along with Ohio and Texas, includes a few items that assess knowledge of basic right-triangle trigonometry.

**Table 9: Distribution of points by content: Geometry/measurement**

Geometry area	HI	FL	MA	NJ	OH	TX	5-state average*
Congruence, Similarity, Transformations	11%	33%	19%	18%	28%	15%	23%
2D Geometry and Measurement	56%	48%	44%	82%	50%	50%	55%
3D Geometry and Measurement	28%	14%	38%	0%	11%	30%	19%
Basic Measurement	0%	5%	0%	0%	6%	0%	2%
Trigonometry	6%	0%	0%	0%	6%	5%	2%

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

\*Because the Maryland exam focuses on Algebra and Data Analysis, it includes only three questions within the realm of geometry. These items constitute too small a sample to merit inclusion in this comparison.

## RIGOR OF THE ASSESSMENT

*For the most part, Hawaii's assessment, like the other state tests Achieve examined, measures mathematics content that students in other countries study prior to high school.*

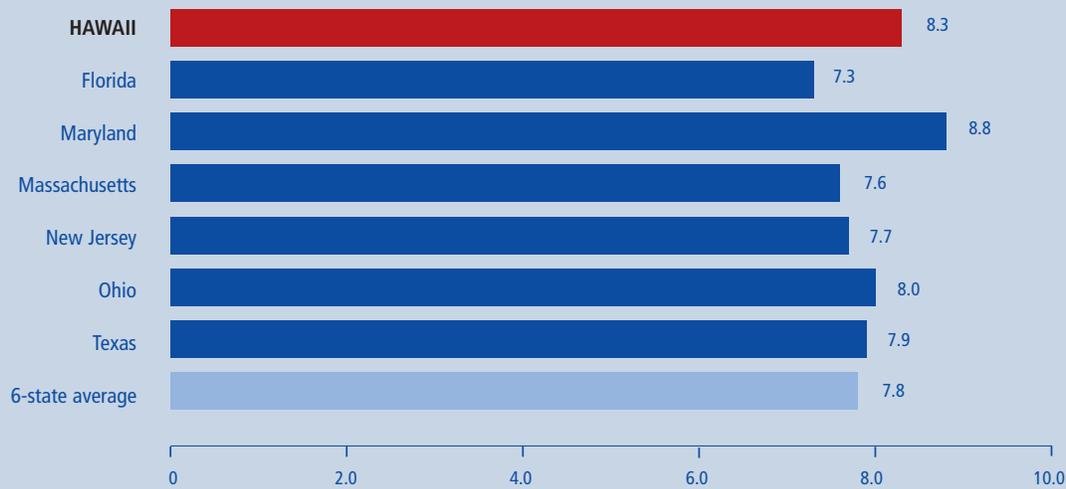
Because the performance of U.S. high school students in mathematics lags behind that of students in other industrialized countries, it is valuable to compare what is expected of students on these tests with expectations in other countries. In our exit exam study, Achieve had the advantage of looking at the mathematics exams through the lens of the International Grade Placement (IGP) index developed by Michigan State University as part of its ongoing work on the 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, since decimals and fractions tend to be taught at the 4th grade level internationally, this topic has an IGP rating of four. Right-triangle trigonometry, on the other hand, tends to be taught in the 9th grade around the world, so it receives an IGP rating of nine.

When applied to assessment items, the IGP describes content only. It is not intended to reflect cognitive or performance demands, nor item format — these are captured by another dimension of Achieve's methodology. When Achieve applied the IGP index to the six states' exit exams, we found that the average of the content measured on the tests is at the 8th grade level internationally. In other words, the material on the exams the six states are using as a requirement for high school graduation is, on average, considered middle school content in most other countries. While there was some variation across the states, no test had an average IGP rating higher than the 9th grade. The range of average IGP values across the six tests in the Achieve's study extended from a low of 7.3 for Florida to a high of 8.8 for Maryland. When compared with these six states, Hawaii's average IGP of 8.3 ranks second only to Maryland's end-of-course Algebra test (8.8).

Hawaii's test attributes more points to advanced algebra concepts than does Maryland, and on that basis we would expect a higher IGP for Hawaii than Maryland. However, Maryland's test is a test of Algebra and Data Analysis, and items having to do with data have a relatively high IGP rating because most countries teach data at higher grade levels than does the United States. This attribute of the IGP boosts Maryland's average IGP relative to the other states.

Chart 16: Average IGP



### COGNITIVE DEMAND

*The majority of points on Hawaii's mathematics test are attributable to items at the middle to lower end of the cognitive demand continuum.*

The content measured by mathematics 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 items as well. At issue is what students are actually required to do with the content. For example, does an item ask students to apply a routine procedure to a mathematical problem, or is the item framed in such a way that it requires students to first develop a more complex mathematical model to solve the problem? The scale Achieve has devised to measure cognitive demand is designed to capture the *processes* that students employ as they “do” mathematics.

In our original study, Achieve found that a majority of the points on the tests across the six states were associated with items that require students to employ processes at the lower end of the cognitive continuum. On a five-point scale of rigor, with one being the least demanding performance and five being the most demanding, slightly more than half the points across the

six tests Achieve studied were tied to the lowest two levels. The cognitive demand profile of Hawaii’s test varies somewhat from the average of the other state tests. Hawaii’s test allots a greater proportion (12 percent) of its points to items that require recall (Level 1) than does any other state test, with the average being just 3 percent. But in its allocation of 49 percent of its test points to Level 2 items — items that require students to use routine procedures and tools to solve mathematics problems — Hawaii squares with other states’ average allocation of 48 percent of points. Similarly, Hawaii meets the six-state average of 26 percent allocation of test points to Level 3 — using non-routine procedures. However, it falls out of line in terms of the points it awards to assessing advanced mathematical skills represented in Level 4 (formulating problems and strategizing solutions) and Level 5 (advanced reasoning), allocating a total of 14 percent of item points to these upper levels as compared with the state average of 22 percent.

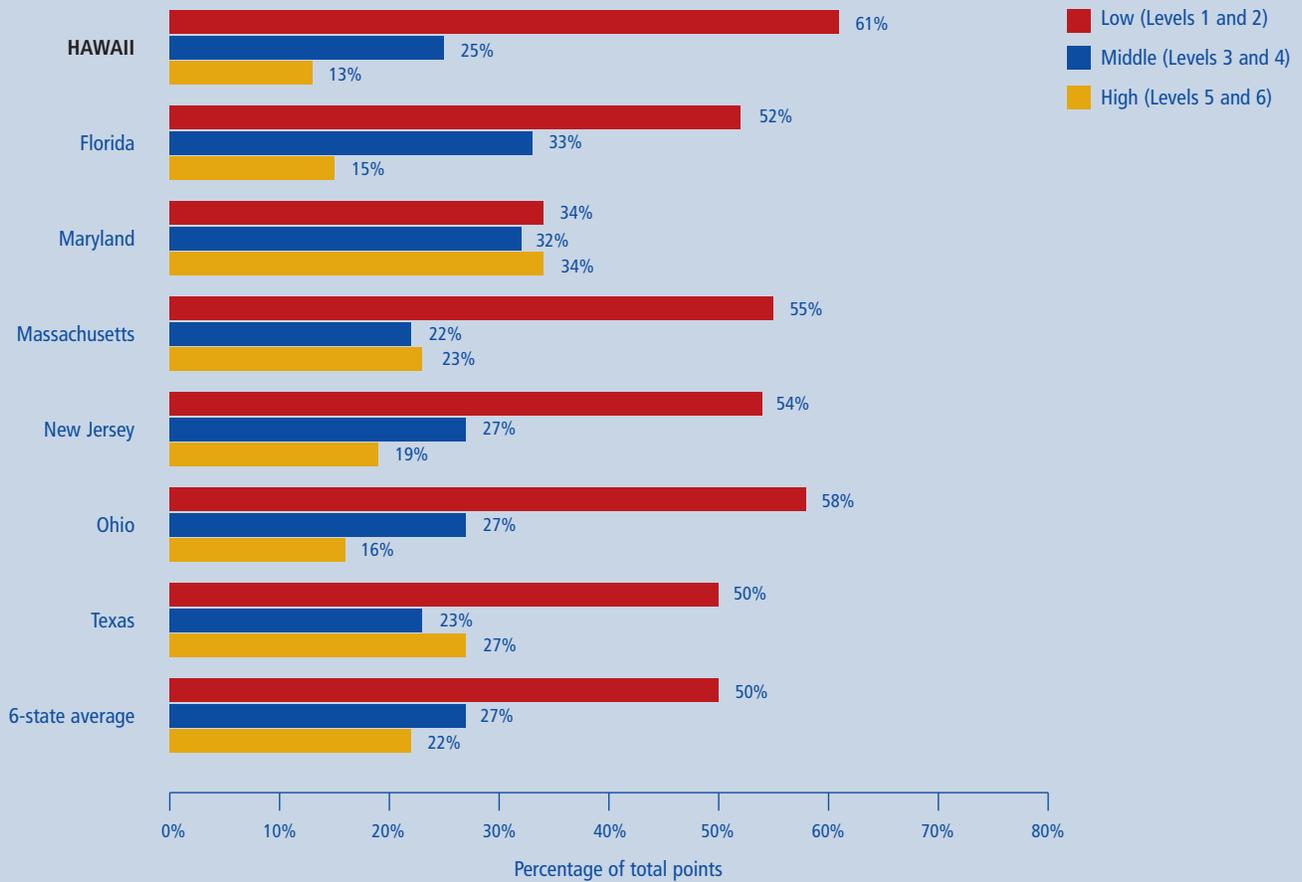
**Table 10: Distribution of points by level of cognitive demand**

Cognitive demand level	HI	FL	MD	MA	NJ	OH	TX	6-state average
1: Recall	12%	2%	4%	2%	8%	4%	2%	3%
2: Using Routine Procedures	49%	50%	30%	53%	46%	53%	48%	48%
3: Using Non-Routine Procedures	25%	33%	32%	22%	27%	27%	23%	26%
4: Formulating Problems and Strategizing Solutions	11%	8%	19%	8%	15%	16%	22%	14%
5: Advanced Reasoning	3%	7%	15%	15%	4%	0%	5%	8%

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

Clustering items in three larger categories of low, medium and high demand helps reveal the underlying structure of Hawaii’s math test. As noted above, only 22 percent of the points across all of the tests are attributed to items that require more advanced mathematical skills (Level 4 and Level 5). However, Hawaii’s distribution of points is a less challenging one than any of the other states, placing the least emphasis on the highest levels of cognitive demand. Only 11 percent of Hawaii’s test points are attributable to Level 4 items that require students to formulate a problem, to strategize or to critique a solution method. And only 3 percent of Hawaii’s points correspond to Level 5 items, which ask students to develop algorithms, generalizations, conjectures, justifications or proofs.

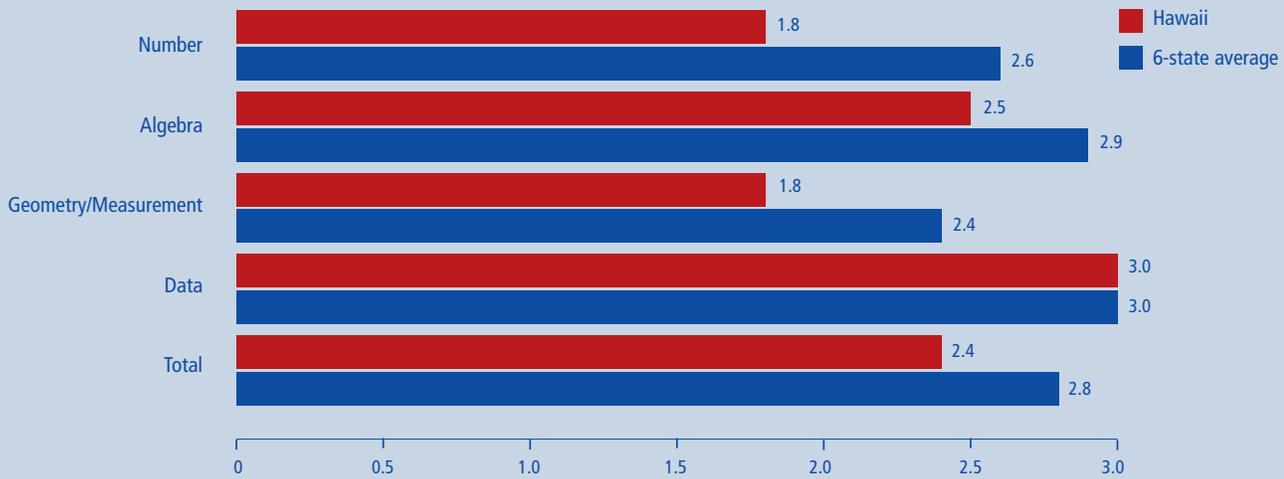
Chart 17: Distribution of points by level of cognitive demand



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

Hawaii’s overall pattern of allotting a high proportion of test points to items of low cognitive demand (recall and using routine procedures) holds true at the strand level. Except for data, Hawaii’s cognitive demand for the major strands is below the average of the other states, especially in the number and geometry strands.

Chart 18: Average cognitive demand by strand, Hawaii versus six-state average



### SUMMARY OF FINDINGS

The content of Hawaii’s grade 10 assessment in mathematics is more rigorous than all but one of the states analyzed due to its relatively high proportion of advanced algebra. Hawaii’s test also is well balanced, giving more emphasis to geometry, algebra and data and less to number concepts, as is appropriate for a high school test. However, the relatively low cognitive demand of Hawaii’s test items in mathematics reduces the overall challenge of the test. This is offset in part by the state’s substantial allotment of item points to robust constructed-response items. Achieve’s analysis indicates that Hawaiian students find items in this format to be more challenging to answer than multiple-choice items — even though they may not in fact involve higher-level cognitive skills. (See section, “What makes the Hawaii mathematics assessment so challenging for students?” for a discussion of the additional analysis of Hawaii’s mathematics test that Achieve conducted.)

## IV. How do the performance levels on Hawaii’s grade 10 assessments in reading and mathematics compare with those of other states?

The aim of a standards-based education system is for all students to acquire the knowledge and skills described by a state’s content standards. State assessments are the principal tool for measuring how well students have mastered that content. Up until this point, this report has focused on what is measured on Hawaii’s assessment in comparison to six states’ exit exams — the content, the difficulty level of the items 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 reach to pass the tests. These cut scores define the level of achievement that students are ultimately held accountable for — they establish the floor of performance expected of high school graduates. As such, these scores represent the level of mastery that a state deems satisfactory.

The Hawaii Department of Education asked Achieve to compare the “Meets Proficiency” level on its grade 10 assessments in reading and mathematics with the cut scores students had to reach to pass the equivalent tests in the six states that participated in Achieve’s 2004 study of graduation exams.

### Methodology

Comparative studies of where states set their cut scores are rare and difficult to conduct. 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 approach permits judgments about the relative difficulty of different tests, but it does not provide information on the knowledge and skills students need to pass each test.

Achieve, working with researchers from Michigan State University (MSU), developed a new procedure for comparing cut scores across state tests that focuses on the content of the test questions, thus giving states a broader comparative picture of their expectations for students. The procedure was first used in *Do Graduation Tests Measure Up?* — published in June 2004 — and has been replicated for the analysis of Hawaii’s assessment. Because the items on Hawaii’s assessment and the six other state tests have been coded according to common metrics discussed previously in the report (e.g., content and cognitive demand), it is possible to use these metrics to identify what a typical student passing the assessments is likely to know and be able to do.

### Performance Levels on the Reading Test

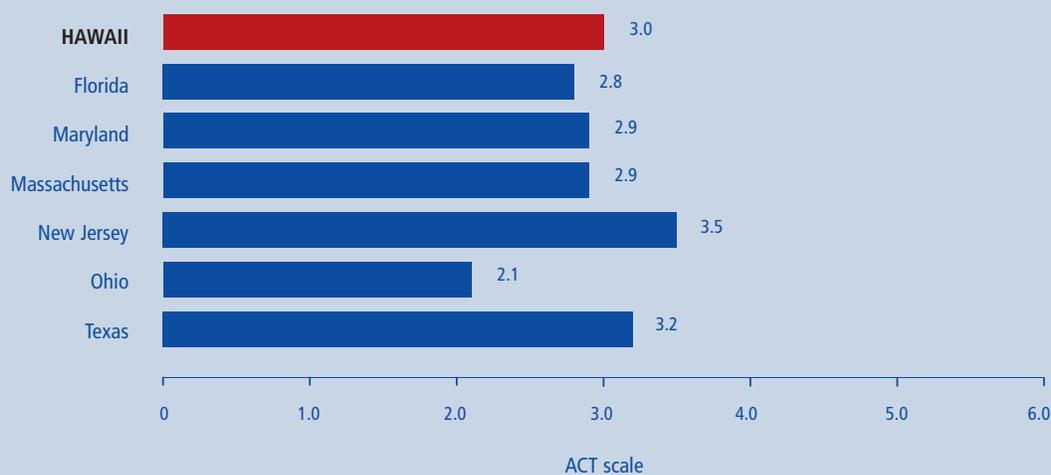
Achieve compared cut scores across the English language arts tests using the ACT skills hierarchy. As stated earlier, ACT indicates that Level 1 through

Level 3 are most heavily assessed on its EXPLORE test, which is given to 8th and 9th graders. Its PLAN test, given to 10th graders, focuses most heavily on Level 3 through Level 5 questions, while the college admissions exam — the ACT Assessment — focuses on Level 4 through Level 6.

Given this frame, Achieve found that the average ACT skill level at the passing score on the state exit exams in the original study ranged from 2.1 to 3.5. Thus, students scoring at the passing level are, generally speaking, being asked to perform below the level that ACT considers appropriate for 8th and 9th graders.

This finding holds true for Hawaii’s reading test. The average ACT skill level at the “Meets Proficiency” score (303) on Hawaii’s reading assessment is 2.9, exactly matching the average across the other six states. A comparison of Hawaii’s rating with the other states suggests that scoring “Meets Proficiency” on Hawaii’s reading assessment is about as challenging as passing Florida’s, Massachusetts’ and Maryland’s exit exams, and considerably more challenging than passing Ohio’s test. With average ACT skill levels of 3.5 and 3.2, respectively, the New Jersey and Texas tests appear to be the most challenging ones to pass among the seven tests, which is not surprising given the relatively high level of content and cognitive demand in these tests. (Note: Item format is not considered part of this scale.) It also is worth noting that New Jersey and Texas administer their tests in the 11th grade, whereas most of the other states, including Hawaii, administer their tests in 10th grade. The exception is Maryland, whose test is an end-of-course test that is administered at the end of the 9th grade.

Chart 19: Difficulty of average passing scenario at “Meets Proficiency” cut score (ACT scale)

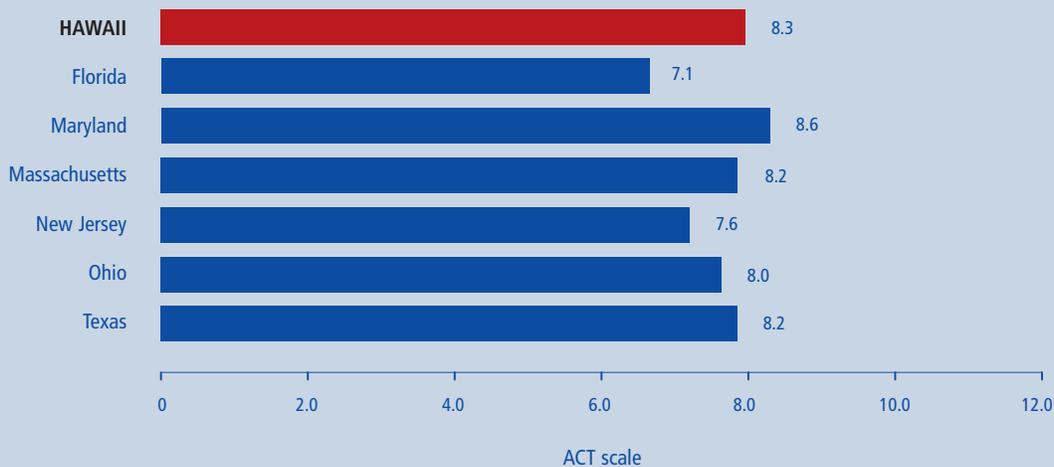


## Performance Levels on the Mathematics Test

As described earlier, Achieve used the IGP index to identify the average level of content measured on the tests. In our original study, we found that, on average, the tests from the six states measured mathematical content that tends to be focused at the 8th grade level internationally. The level of mathematics content knowledge students need *to pass* the state exit exams ranged from 7.1 to 8.6. That is, the questions on the tests that students scoring at the cut score are likely to get correct measure, on average, concepts that students around the world focus on in the 7th and 8th grades.

The average IGP score at Hawaii’s “Meets Proficiency” cut score (300) is 8.3. Essentially, this means that to pass Hawaii’s grade 10 mathematics assessment, students are required to know mathematics content that is taught, on average, in early 8th grade internationally. Hawaii’s score of 8.3 also indicates that passing its mathematics assessment offers a similar challenge passing in terms of its content difficulty as Ohio, Texas and Massachusetts. Hawaii’s assessment exceeds the content demand of New Jersey and Florida. Only Maryland’s end-of-course test in Algebra and Data Analysis received a higher IGP score than Hawaii. Hawaii’s high score is likely due to the emphasis the mathematics assessment gives to advanced algebra. As noted previously, the average IGP score for Maryland is elevated somewhat by data topics placing relatively high on the IGP scale because internationally these topics fall later in the grade curriculum sequence. Again, it is content — not cognitive demand or item format — that is the basis for the IGP index.

Chart 20: Average difficulty of mathematical content at “Meets Proficiency” cut score (International Grade Placement scale)



### *What makes the Hawaii mathematics assessment so challenging for students?*

Hawaii is rightly concerned that large numbers of students are not passing the state mathematics assessment. The percentage of students who attained the “Meets Proficiency” or “Exceeds Proficiency” cut scores in mathematics in 2005 are 18 percent and 2 percent, respectively. The passing rate, which includes both categories, has been relatively stable: 19 percent of Hawaii students passed the mathematics test in 2002, 18 percent passed in 2003, 21 percent passed in 2004 and 20 percent passed in 2005.

What is it that makes Hawaii’s mathematics assessment so challenging for students? One important factor is that Hawaii’s test, as indicated by its average IGP value, contains more demanding content than all but one state test. This is due to the fact that Hawaii allots more points to advanced algebra topics than do other states. In addition, Hawaii appropriately includes basic right-triangle trigonometry, as do only two of the other states. Hawaii also includes a substantial number of robust constructed-response items, which students seem to find more difficult. (Achieve’s analysis of Hawaii’s test data in mathematics revealed that students seem to have substantially less difficulty answering multiple-choice items.)

Nonetheless, Achieve did not find the Hawaii test to be too challenging. As in the other state tests Achieve analyzed, the majority of items on Hawaii’s test assess content that students study prior to and early in high school. In addition, Hawaii’s test items fall in the low range of overall cognitive demand, as compared to the six states in Achieve’s study. The majority of Hawaii’s test points are awarded to items that make minimal demands — “recall” (12 percent) or “use routine procedures” (49 percent) — or modest demands — “use non-routine procedures” (25 percent).

Are there other characteristics of the test that could account for the relatively low passing rates that are not easily captured by the criteria in this study? Achieve asked the Hawaii Department of Education to provide course-taking patterns, meaning the percentage of students who had completed or were enrolled in Algebra II at the time the test was administered in 2005 (i.e., in the spring of grade 10). Since the Department of Education does not yet have a data system in place capable of tracking the number of students enrolled in Algebra II and in what grade, it asked registrars to make an educated guess. Following are the grade-by-grade estimated percentages of students enrolled in Algebra II that the Department of Education shared with Achieve: Grade 9 is 1 percent, grade 10 is 29 percent, grade 11 is 50 percent and grade 12 is 20 percent. Given this data, one possible reason for the low number of Hawaiian high school students who score at the “Meets Proficiency” level on the grade 10 assessment is the contrast between the relatively high proportion of advanced algebra content on the assessment and the relatively low proportion

of students who had been taught the content by the time they took the test. It seems that the majority (70 percent) of Hawaii's students are not enrolled in Algebra II until grades 11 or 12, placing them at a significant disadvantage in terms of responding to items based on advanced algebra content. It is also important to keep in mind that even those students enrolled in Algebra II in grade 10 are taking the assessment before they have completed the coursework, as the test is administered in March or April. It is Achieve's understanding that the state already is acting to correct this preparation gap by revising its high school standards and ensuring that the new assessment will be tightly aligned to the revised standards.

Three other minor factors also may be contributing to low student performance:

1. A lack of motivation, as the test does not count for students;
2. A lack of sufficient exposure in coursework to constructed-response items, which present a significantly higher level of challenge to Hawaii's grade 10 students than do multiple-choice items; and
3. The fact that students are not allowed to use a calculator on any portion of the test, in contrast to most other states.

There is growing evidence from other states that high school students take standards and assessments more seriously when they know that their performance on those tests counts. For example, only 48 percent of 10th graders in Massachusetts passed the mathematics portion of the state's new graduation 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 a variety of academic supports. When the 10th graders from the class of 2003 took the test — the first group that had to pass it to graduate — the scores jumped up nearly 20 percentage points, suggesting that when it counts, students (and schools) put forth more effort. By spring 2003, 95 percent of students in the graduating class had passed the test.

A similar story played out in Virginia as it phased in new end-of-course exams for high school graduation. Only 40 percent of students passed the Algebra I exam when it was first given in 1998 (more students passed the reading and writing tests). By 2003, 78 percent passed the Algebra I test, and by the time the first class of high school seniors had to pass several of the end-of-course tests to graduate in spring 2004, all but 1 percent earned their diplomas.

The combination of low student motivation and the significant role that constructed-response items play in Hawaii's assessment may help explain the state's low passing rate in mathematics. As noted previously, Hawaii's mathematics assessment includes a significant number of constructed-response items (56 percent of total points), particularly when compared with other states (36 percent average across the other six states). Items in this kind of format tend to be challenging because they often require a substantial amount of reading and/or reasoning through multiple steps to solve a problem. Perhaps most significantly, constructed-response items do not provide the crutch of offering a set of responses, one of which is correct and sometimes verifiable by guessing and checking. Achieve verified that students, on average, got a lower percent of constructed-response items (30 percent) correct than multiple-choice items (43 percent). Constructed-response items are essential because they have the ability to measure more advanced skills and closely reflect the kind of tasks students will face in college courses and the workplace. However, if students are not used to routinely solving open-ended problems, their format can pose an additional challenge.

It is possible that — because they know the assessment doesn't count for graduation — students are not putting forth the necessary effort to complete constructed-response items. To test this hypothesis, MSU researchers looked at the student-response data in mathematics to see how students performed on each type of question. Their findings were revealing. Again, statistically significant differences were found by item type. On average, constructed-response items are more than twice as likely to be skipped by students (19 percent) as multiple-choice items (8 percent). We cannot know the mindset of these students at the time they took the test, but it is conceivable that they simply were unmotivated to complete the more demanding items because the test does not count for them.

In revising its assessment, Hawaii will need to ensure that the new test is fully aligned with the new standards and thus fair to students. At the same time, Hawaii will want to ensure its grade 10 assessments are rigorous (i.e., include the more demanding content and performances delineated in the revised standards). Indeed, Hawaii should be careful not to reduce the overall rigor of its assessment, nor reduce the proportion of constructed-response items, nor lower the cut score for "Meets Proficiency." In fact, Achieve's American Diploma Project found that states need to educate their high school graduates to achieve a far higher level of math proficiency than they currently are accomplishing, if students are to succeed in college and a workplace increasingly steeped in quantitative analysis.

## V. Conclusion

Achieve launched its original 2004 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?

Across the states, we found that the tests do indeed set a floor for students that can be responsibly defended as a graduation requirement. However, they do not effectively tap the higher-level skills that truly constitute “readiness” for college and work. Generally speaking, our findings for Hawaii are consistent with those of our previous study, with a few critical differences.

- In reading, Hawaii’s assessment is somewhat less rigorous than other tests in the study. Although the test includes challenging questions and emphasizes informational text, the reading passages are not as rigorous as in other states, which diminishes its overall rigor. The result is that the “Meets Proficiency” level of performance sets a standard that is roughly comparable to that of other states in Achieve’s study.
- In mathematics, Hawaii’s assessment includes more rigorous algebra content than the other tests we studied, making it a more effective measure of the knowledge and skills that are important for college and work. At the same time, however, the average cognitive demand of the items is relatively low compared to that of other tests. The result is that Hawaii’s “Meets Proficiency” level is not significantly different than that of four of the six participating states in Achieve’s study and does not, in our opinion, set an unreasonable standard for high school graduates.

### Recommendations for Improvement

As Achieve found in its original study of graduation exams in six states, the Hawaii tests set a reasonable floor of expectation that should be raised over time. As Hawaii moves forward, Achieve recommends that the state:

- **Raise the overall rigor of the grade 10 reading test.** Hawaii should increase the demand and complexity of the reading passages on its assessment. Some passages on the reading test should represent the level of demand typical of instructional materials written at a late high school reading level to raise the ceiling on the test and signal the level of text students need to comprehend to be on track for attainment in postsecondary education and the new economy. In addition, Hawaii should add items that tap the highest level of cognitive demand, requiring students to more deeply analyze text.

- **Phase in higher cut scores on the reading test over time.** In addition to increasing the cognitive demand of its passages and/or items, Hawaii can raise the rigor of its reading test over time by raising the score required for passing. Texas is using this approach with its new graduation exam. This strategy works only if a test has enough range in what it measures so that a higher score actually reflects more advanced knowledge and skills. If a higher cut score simply means that students must answer more of the same kinds of items correctly, rather than items tapping more advanced concepts and skills, it is not very meaningful to raise the cut score.
- **Raise the level of performance demand of the mathematics items.** Although the content on the math test is challenging, the items themselves tend to make low-level demands in terms of performance. Hawaii should raise the cognitive demand of its items by increasing the proportion of items that require complex problem-solving skills, problem formulation and advanced reasoning.
- **Build assessments of college and work readiness.** While the Hawaii grade 10 assessments set a reasonable floor for students, over time, Hawaii will need to go beyond its grade 10 test and develop a more comprehensive set of assessments that measure the full set of knowledge and skills that indicate readiness for college and work.

In addition, Hawaii will need to work with its local districts to establish a systematic method for evaluating important skills such as research and oral communication that are not easily assessed on a paper-and-pencil test.

### Additional Recommendations

As Hawaii raises its standards over time, it will face the simultaneous challenge of raising student achievement. To help meet this challenge, particularly in mathematics, Hawaii should:

- Ensure the tests are aligned with the state standards and vertically aligned with each other, and ensure students are exposed to content prior to taking the assessment.
- Conduct analyses to determine as precisely as possible why the percentage of students scoring at or above the “Meets Proficiency” level on the grade 8 test is low and provide targeted professional development to help teachers teach content and skills more effectively, including maximizing students’ exposure to constructed-response items. Shoring up middle school performance will help ensure the larger majority of students will reach the “Meets Proficiency” level on the grade 10 test.

- Publish the state compendium of sample items, scoring rubrics and sample students responses on the state Web site to more widely inform the public of state expectations.
- Identify and disseminate the most effective instructional materials available, taking full advantage of the distribution power of the Internet.

### **Achieve's ADP Network**

Achieve encourages Hawaii to consider joining its ADP Network, the group of 22 states that have pledged themselves to a policy agenda in support of preparing students for success in college and work by the time they graduate from high school. To close the expectations gap, the ADP Network states have committed to the following four actions.

- Aligning high school standards and assessments with the knowledge and skills required for success after high school;
- Requiring all high school graduates to take challenging courses that actually prepare them for life after high school;
- Streamlining the assessment system so that the tests students take in high school also can serve as readiness tests for college and work; and
- Holding high schools accountable for graduating students who are ready for college or careers and holding postsecondary institutions accountable for students' success once enrolled.

Although the Network has been in existence for just over a year, Achieve has already seen evidence of substantial progress on the part of participating states. Hawaii should take advantage of the opportunity to join in partnership with these other states so Hawaiian students will be ready to take on the challenges of living and working in a global economy, equipped with the knowledge and skills required for success.

## Appendix: Summary of Methodology

To compare assessments, each assessment item was analyzed and coded on the basis of distinguishing attributes to capture different characteristics of individual test items and the tests as a whole. Many of the criteria in reading and mathematics 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 further analyzed.

The following are summaries of the various criteria according to which assessments in the study were analyzed.

### Content of Items

#### *Mathematics*

To classify the content on state mathematics assessments, Achieve used the Third International Mathematics and Science Study (TIMSS) Mathematics Framework, adapted 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.

#### *Reading*

To identify the content on reading assessments for its original six-state study (*Do Graduation Tests Measure Up?*), Achieve adapted a comprehen-

sive listing of the domain of reading, originally developed by the Council of Chief State School Officers (CCSSO) in collaboration with several states for its Survey of Enacted Curriculum. The list was intended to fully encompass all topics addressed in reading classes from the primary to the secondary level.

Based on this framework, Achieve developed a taxonomy that included all the aspects of reading described in state standards — and therefore targeted on state tests — to describe as accurately as possible the content or topic that each item measured. The listing used in Achieve’s study has been revised to more clearly reflect the topics addressed by test items at the secondary levels. Because the list was originally developed to cover all grades from kindergarten through grade 12, some codes that were irrelevant for higher-level tests have been deleted. In addition, the listing has been reorganized to clarify the relationship of the elements.

In Achieve’s original study, the major reporting categories for reading were as follows:

- 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)

In this previous grouping, some of the elements overlapped. To streamline and clarify reporting, Achieve has regrouped some of the codes into more discrete categories. For example, argument and assertion, formerly under critical reading, are both aspects of persuasive texts and now are grouped under informational/persuasive elements. Codes also have been realigned into groupings that reflect all of the elements within them. For example, all the literary elements have been grouped together — narrative elements with the author’s craft elements. Additionally, some elements formerly included in the critical reading category were deleted because they combined references to both content and cognitive demand, such as determining the validity of an assertion.

This revision yields four different major categories of codes for reading:

1. Vocabulary (includes word definitions)
2. General comprehension (includes purpose and main idea)
3. Literary elements (includes figurative language, plot and character, theme, setting, and poetic language)
4. Informational/persuasive elements (includes organization and structure, assertions, evidence and technical elements)

## Approximate Grade-Level Demand of Items

### Mathematics

To approximate the grade-level demand of mathematics items, Achieve used the TIMSS International Grade Placement (IGP) index, developed by the U.S. TIMSS National Research Center at Michigan State University. 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 mathematics 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 topics.

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

## Reading

To approximate the grade-level demand of reading items, Achieve adapted the ACT Standards for Transition (for 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 related demand of the reading passage. 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 Items

### Mathematics

Achieve developed a taxonomy of performance expectations (i.e., 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. This taxonomy was categorized into five levels of cognitive demand of mathematics items. The

five levels describe the kind and complexity of performance required of test-takers — 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 making decisions that go beyond a routine problem-solving activity.
- Level 4 includes formulating a problem, as well as strategizing or critiquing 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.

### **Reading**

In its original study of six states' graduation exams, Achieve used a taxonomy of performance expectations derived from CCSSO's description of performances in its Survey of Enacted Curriculum and influenced by Achieve's assessments-to-standards alignment protocol. This taxonomy was categorized into four levels of cognitive demand of reading items. The four levels provided information on the kind and complexity of reasoning required of students, ranging from simple recall of information to complex reasoning skills. The former categories were as follows:

- Literal recall
- Infer
- Explain
- Analyze

Some revisions have been made to the former categories, informed by a recently revised taxonomy based on Bloom developed by Anderson and Krathwohl in 2001. The revised taxonomy retains much of Bloom’s 1956 model. However, the original Bloom taxonomy combined both content and performance, while the revised taxonomy separates the content from the performance. Several of the original codes in the former Achieve cognitive demand scale also included content topics as well as processes, making the codes for some items redundant, such as “identifying main ideas or theme,” where both main idea and theme are now coded with an appropriate content code.

The revised scale retains some of the same headings as the original, with some expansion that allows for a better discrimination among cognitive processes typically assessed in reading tests:

- Recall (includes locating and recognizing)
- Low inference (includes paraphrasing and generalizing)
- High inference (includes concluding, comparing and illustrating)
- Construct (includes summarizing and explaining)
- Analyze (includes discriminating and outlining)
- Evaluating (includes critiquing)
- Creating (includes designing and hypothesizing)

Categories six (evaluating) and seven (creating) are not characteristic of items found on large-scale, on-demand state tests and are, therefore, not included in the related data charts.

### **Demand of Reading Passages**

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, Level 2 and Level 3 represent middle school-level reading, Level 4 represents early-stage high school reading, and Level 5 and Level 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

## 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. This interaction of cognitive demand level and reading level contributes to the challenge of an item. For example, an item could require a low level of performance on a difficult passage, a high level of performance on an easy passage, a high level of performance on a difficult passage or a low level of performance on an easy passage. 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 five and the Reading Level Demand Scale from a low of one to a high of six, allowing for a total of 10 possible scores (from two to 11) that items can achieve on the RRI.

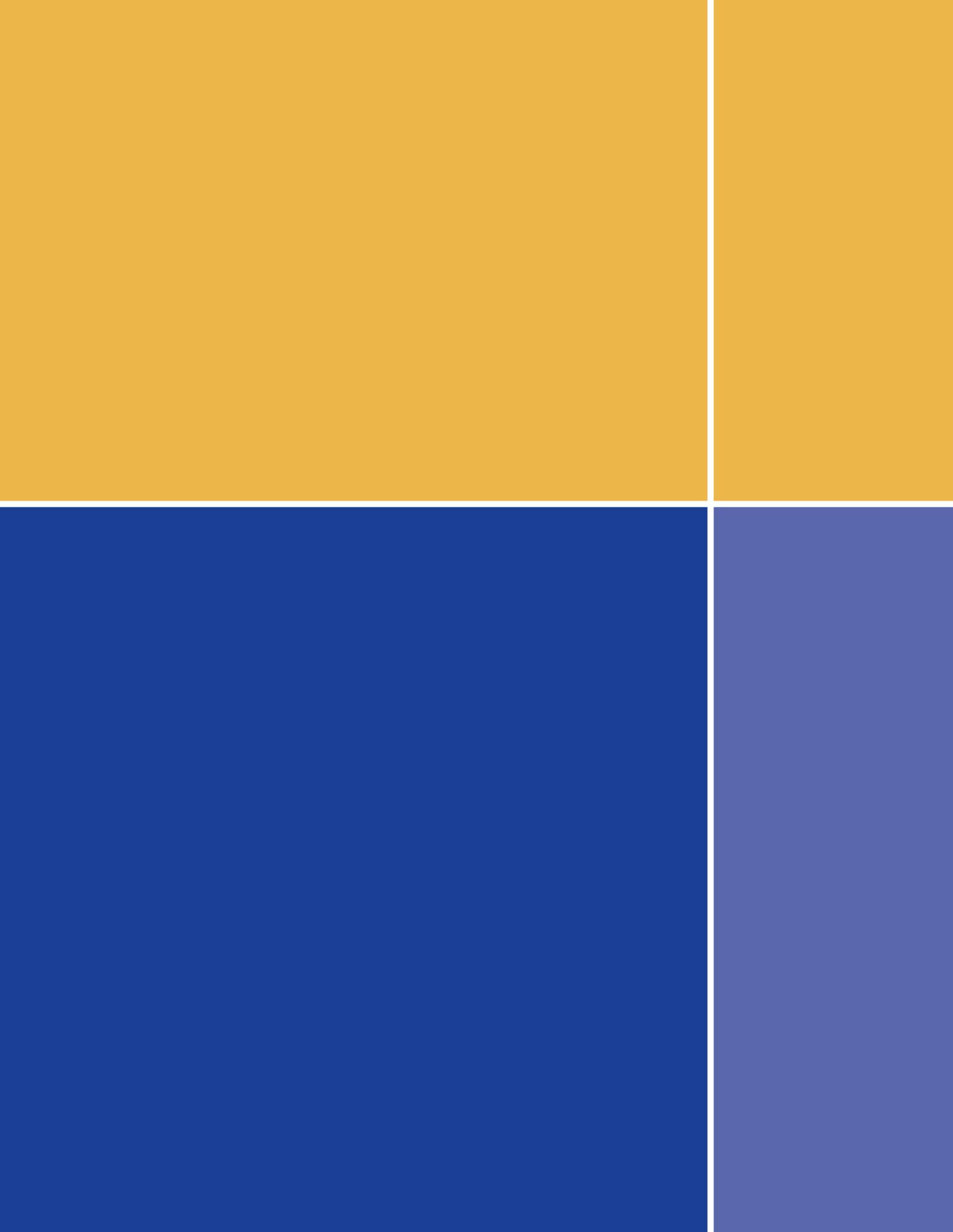
The average RRI score for each test is calculated by weighting each item according to its point value and averaging the result. For example, an item based on a Level 3 reading passage with a cognitive demand of two would have an RRI score of five. If the item were worth two points, as in a constructed-response item, the item would be given double weight in calculating the average RRI level of the test.

## Cut Scores

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, 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 in the original study 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 lens of the IGP index 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 was used for the computation.





Achieve, Inc.

Achieve, Inc.  
1775 Eye Street, NW  
Suite 410  
Washington, DC 20006  
[www.achieve.org](http://www.achieve.org)