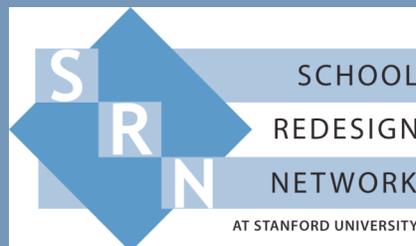


# **Analysis and Recommendations for Alternatives to the Washington Assessment of Student Learning**

**Linda Darling-Hammond  
Laura McCloskey  
Ray Pecheone  
School Redesign Network  
Stanford University School of Education  
September, 2006**



## **Table of Contents**

Tables and Figures .....	3
Abstract .....	4
Introduction.....	6
The Washington Exit Exam Policy and Context .....	7
The National Context .....	10
Rationales for Alternatives.....	15
Evaluation of Options.....	19
Alternate or Substitute tests .....	19
Grade Point Average.....	24
Collection of Evidence.....	27
Weighted Multiple Measure Approach.....	33
Conclusions .....	38
References .....	40

## Tables and Figures

- Table 1:** Overall Passage Rates on the WASL – p. 6
- Table 2:** Pass Rates on the WASL, by Subgroup – p. 7
- Table 3:** Alternate Paths for General Education Students to Obtain a Diploma, 2006 - p. 10
- Table 4:** Exit Exam Alternatives Used by States – p. 11
- Table 5:** Elements of Universally Designed Assessments – p. 16
- Table 6:** Key Definitions: Alternate Test, Substitute Test – p. 19
- Table 7:** Corresponding Scores for Florida’s Exit Exam and Substitute Tests – p. 21
- Table 8:** Potential Difficulties, Benefits of using Alternate or Substitute Tests – p. 24
- Table 9:** Key Definitions: Grade Point Average. GPA Cohort Comparison – p. 27
- Table 10:** Potential Difficulties, Benefits of using GPAs – p. 28
- Table 11:** Key Definitions: Collection of Evidence, Portfolio – p. 29
- Table 12:** Potential Difficulties, Benefits of Using Collection of Evidence – p. 34
- Table 13:** Key Definitions: Multiple Measures Approaches to Assessment, Weighted Multiple Measures System of Assessment – p. 35
- Table 14:** Potential Benefits and Difficulties of a Weighted Multiple Measures System – p. 36
- Figure 1:** Graduation rates in states that require exit exams (without alternatives) prior to 2001 – p. 12
- Figure 2:** Graduation Rates in States Requiring Multiple Measures Approaches to Graduation Prior to 2001 Compared to Test-Only States – p. 18
- Figure 3:** Arizona’s formula for calculating a pupil’s augmentation - p. 26

## Abstract

Washington's provision of alternatives for its exit exam, the Washington Assessment of Student Learning (WASL), represents a solid policy pursued by most other states with exit examinations. Psychometric standards, supported by the three leading national professional research organizations concerned with testing and measurement, also encourage the use of more than one measure to assess student learning to enhance the validity of the assessment system. The use of multiple measures can also improve teaching and encourage attention to learning standards that include more challenging skills like research, problem-solving, and extended communication not covered on standardized tests. The use of alternatives may also encourage more students to stay in school rather than being persuaded they cannot succeed, and may reduce the incentives for schools to push out students who do not initially test well, including those with special education and language needs.

In this paper, we examine the pros and cons of different approaches to alternatives in light of Washington's current policy and context, the lessons that have emerged from other states, and the research on different forms of assessment. We examine the three measures recently proposed by the legislature in Senate Bill 6475 as alternatives for students who have failed the 10<sup>th</sup> grade WASL twice. These include:

- 1) obtaining a score on the mathematics components of the PSAT, ACT, or SAT that is deemed equivalent to passing the WASL;
- 2) comparing students' grades with the grade of other students who took the same courses and who met the standard on the WASL;
- 3) evaluating a collection of work samples (also known as a collection of evidence) with specific collections designed for students in programs leading to a national or state industry certificate.

We conclude that all three of these alternatives hold promise for inclusion in a multiple measures system and propose some refinements to the current proposals. Our recommendations include:

- Allowing students to use substitute tests in all of the fields tested by the WASL and, in addition to the SAT or ACT, considering as potential options assessments that would encourage students to undertake more rigorous coursework and challenging tasks, such as the Advanced Placement (AP), International Baccalaureate (IB), and Cambridge Advanced International Certificate of Education (AICE) assessments, and more diagnostic assessments like the Measures of Academic Progress (MAP) assessments. The use of these substitutes routinely, as in many states, rather than only for those who fail the exam, should be considered.

- Evaluating the design of the current assessments to see how well they meet principles for universal test design and how well they assess the knowledge and skills of English language learners, and, if necessary, creating a linguistically-modified version of the test that increases its validity for assessing English language learners.
- Considering student GPAs in standards-based coursework as an adjunct to the examination score, as some states do, rather than as an alternative measure only for those who have failed the exam.
- Ensuring that a collection of evidence approach included in the final assessment system uses tasks that are part of students' coursework and are well-specified, connected to the standards and course syllabus, and commonly scored, so that they are supported by teaching and can be reliably evaluated. This alternative should ideally be pursued from the beginning of the students' high school career and perhaps should be available to all students, not just those who fail, as a way of strengthening teaching and learning.

These measures should be made available to students as quickly as possible and ideally, as in some states, to all students so that the graduation decision is based on multiple forms of evidence. Those that are reserved as alternatives only for students who have failed the test should be available after one attempt. Especially in the case of a collection of evidence approach, implementation of the alternative should begin early in students' career both to ensure a high-quality process and to encourage students to work hard in school and to remain in school rather than dropping out.

## Introduction

In 1999, Washington State administered the Washington Assessment of Student Learning (WASL) for the first time to its tenth graders. This 10<sup>th</sup> grade test is intended for use by 2008 as an exit examination that students must pass in order to graduate. From the initiation of the policy, there was a provision that alternative assessments would be used to evaluate the learning of students who fail the exit exam, in recognition of the fact that a single test is not a valid basis for making a high-stakes decision like graduation.

At this point, the state is considering what kinds of alternatives to incorporate into its exit exam policy and how to use them. The importance of these decisions has been made clear by the high rate of failure on the 10th grade WASL, especially in mathematics. For example, only half of students passed the mathematics test in 2005-2006, and an even smaller number would be eligible to graduate based on passage of all three tests that are required. To address the need for alternative measures, the Washington legislature last year instituted three alternatives to passing the WASL, which include a GPA cohort comparison, use of SAT/ACT mathematics scores, and evaluation of a collection of evidence reflecting the student's work.

As we discuss below, Washington's support of alternative pathways to a diploma for students not passing the WASL makes sense on both psychometric and educational grounds. The three national professional research organizations concerned with testing and measurement recognize the importance of using multiple measures of assessment for any important decision. Research suggests a policy involving multiple measures is likely to improve pedagogy and support the teaching of analytical skills and performance abilities overlooked by test-based curricula. The use of alternatives may also encourage more students to stay in school rather than being persuaded they cannot succeed, and may reduce the incentives for schools to push out students who do not initially test well, including those with special education and language needs.

In this paper, we examine the pros and cons of different approaches to alternatives in light of Washington's current policy and context, the lessons that have emerged from other states, and the research on different forms of assessment. We conclude that all three measures proposed by the legislature hold promise for inclusion in a multiple measures system, and we argue that they could be configured in a broader and more integrated fashion than the current legislation envisions. In particular, we urge consideration of approaches like those used in many other countries that incorporate evidence from student work that is developed in line with standards.

## Washington's Exit Exam

By 2008, all students in Washington will need to demonstrate academic proficiency in order to receive a Certificate of Academic Achievement that enables them to graduate from high school. By law, the primary pathway for demonstrating proficiency is to receive a passing grade on the reading, writing, and mathematics subtests of the Washington Assessment of Student Learning (WASL) that test students' skills using multiple choice, short answer, and essay questions. By 2010, students must also pass the science portion of the WASL. Between the time the legislature initiated the tests in 1993 and administered the WASL to the first tenth graders in 1999, the authors of the WASL created a test that includes more written tasks and tasks that ask students to explain their thinking than do most traditional standardized tests (Chan, 2004; Shaw, 2005).

After eight years of piloting the tenth grade WASL, combined pass rates continue to be extremely low (barely 50% overall). Although they have climbed over the last 8 years, pass rates for 10<sup>th</sup> graders are only 51% on the math test overall, and considerably lower for students with disabilities and those who are English language learners. (See tables 1 and 2.) At this point, it is not clear the extent to which these results are due to problems with the content validity of the examination and the extent to which they are due to instructional shortcomings. It is likely that both factors play a role in these outcomes. Since Washington's students score well above the national average on all of the NAEP subject matter tests – and were ranked 7<sup>th</sup> in the nation in 8<sup>th</sup> grade mathematics in 2005, it does not seem likely that the low pass rates on the WASL fully reflect students' proficiencies in this field.

**Table 1: Overall Passage Rates on the WASL**

Year	Reading	Math	Writing	Science
1998-99	51.4%	33.0%	41.1%	
1999-00	59.8%	35.0%	31.7%	
2000-01	62.4%	38.9%	46.9%	
2001-02	59.2%	37.3%	54.3%	
2002-03	60.0%	39.4%	60.5%	31.8%
2003-04	64.6%	44.0%	65.4%	32.3%
2004-05	72.9%	47.5%	65.2%	35.8%
2005-06	81.9%	51.0%	79.7%	34.9%

*(Data from the Office of Superintendent of Public Instruction, "Washington State Report Card.")*

Although there are some supports in place for students who need special considerations, there are still large disparities in scores among subgroups. Special education students may use accommodations suggested by the IEP team when they take the exam; they may also take alternative tests such as the Washington Alternative Assessment System – Developmentally Appropriate WASL, but this pathway only allows them to receive a Certificate of Individual Achievement (CIA), which will be noted on the students' transcripts. Students with Limited English Proficiency (LEP) are allowed accommodations on the WASL such as native language dictionaries when suggested by a decision-

making team of education, administrators and parents (Office of Superintendent of Public Education, October 2005).

Even with these supports, Special Education students and students classified as limited English speakers passed the math test at rates of only about 12% and the reading and writing tests at rates of only about 30 to 40%. African American, Native American, and Hispanic students also had considerably lower scores than Asian and Caucasian students. All groups had low scores in science (only 35% passed overall), an area that will become part of the exit exam in 2010.

**Table 2: Pass Rates on the WASL, by Subgroup**

Year	Reading	Math	Writing	Science
<b>Students with Disabilities</b>				
2002-03	11.9%	4.0%	11.6%	3.6%
2003-04	15.3%	5.5%	15.0%	3.8%
2004-05	22.5%	6.2%	14.3%	3.7%
2005-06	41.4%	11.6%	38.2%	5.5%
<b>Limited English Proficient Students</b>				
2002-03	11.7%	8.1%	10.8%	2.8%
2003-04	16.8%	9.7%	17.3%	2.6%
2004-05	28.7%	11.9%	19.7%	4.2%
2005-06	35.5%	12.8%	32.8%	4.4%
<b>Caucasian Students</b>				
2002-03	65.1%	44.0%	65.5%	36.3%
2003-04	69.6%	49.2%	69.7%	37.4%
2004-05	77.0%	52.4%	69.2%	40.5%
2005-06	86.4%	56.4%	83.9%	39.5%
<b>African American Students</b>				
2002-03	37.1%	14.2%	39.3%	9.2%
2003-04	42.8%	16.0%	48.9%	9.2%
2004-05	53.7%	20.4%	47.9%	12.1%
2005-06	66.0%	23.1%	65.3%	12.0%
<b>Hispanic Students</b>				
2002-03	34.6%	16.2%	34.3%	11.1%
2003-04	41.3%	19.6%	42.5%	11.1%
2004-05	53.1%	23.9%	43.7%	14.2%
2005-06	62.5%	25.4%	59.9%	14.0%
<b>Asian Students</b>				
2002-03	64.2%	46.8%	66.1%	32.7%
2003-04	70.5%	52.2%	73.4%	33.6%
2004-05	78.8%	56.9%	72.9%	41.6%
2005-06	84.5%	59.7%	84.4%	40.2%
<b>American Indian Students</b>				
2002-03	42.5%	21.9%	41.1%	15.9%
2003-04	46.4%	23.3%	46.8%	14.9%
2004-05	55.8%	26.9%	45.0%	17.9%
2005-06	67.7%	30.1%	65.6%	18.1%

(Data from the Office of Superintendent of Public Instruction, "Washington State Report Card.")

Among the 25 states that have enacted exit exams, no state has successfully implemented a graduation requirement based on a test with pass rates this low. Because a high school diploma is a gateway to most employment and even to service in the military, these fail rates – if they are not significantly mitigated by other pathways to demonstrate knowledge and skills – would cut off access to productive futures for most of Washington’s young people. Most states’ initial pass rates on exit exams are well over 70%, and these rates climb when students retake the exam (Center on Education Policy, 2006). Other states with high failure rates – though lower than Washington’s – have postponed, modified, or abandoned the exam.

For example, after complaints about the narrowing of the curriculum accompanying implementation of its exit exam and concerns about disparate fail rates, Wisconsin created a system of local performance assessments to be scored using a state scoring system. These were deemed so successful, the state repealed its exit exam requirement entirely. In Arizona, low pass rates led to two postponements of the date for implementing the exit exam, so the legislature voted in 2005 to provide an alternative route for students to boost their scores as much as 25% by using class grades (FairTest Examiner, May 2006; FairTest Examiner, Winter 2005; FairTest Examiner, Summer 2005). More recently, when Utah found 16% of seniors in the class of 2006 not passing all three sections of the exit exam by January, the state board of education decided to give diplomas to students who met all other graduation requirements and took the exit exam three times but did not pass. State diplomas will carry information about whether the student passed the exam (CEP, 2006).

Washington’s legislature has reacted to low passage rates on the WASL by including several kinds of evidence for graduation. In March 2006, the Washington legislature passed Senate Bill 6475, authorizing the development and implementation of alternative pathways for students who do not pass one or more sections of the WASL after at least two attempts. The three alternatives instituted by Senate Bill 6475 include:

- 1) obtaining a score on the mathematics components of the PSAT, ACT, or SAT that is deemed equivalent to passing the WASL;
- 2) comparing the students’ grades with the grade of other students who took the same courses and who met the standard on the WASL;
- 3) evaluating a collection of work samples (also known as a collection of evidence) with specific collections designed for students in programs leading to a national or state industry certificate.

## The National Context

Washington is not alone in its consideration of alternatives to its exit exam. Out of the 25 states that have enacted legislation to require an exit exam, 17 have adopted one or more alternative assessments for students who do not pass that test. (See Table 3). In addition, a number of states use an exit exam as part of their graduation decision along with other local measures, generally requiring that the test be considered but not be the sole basis for denying a diploma. Many states introduced alternatives in response to low and/or disparate pass rates, mounting legal pressures, and political backlash. The Center for Education Policy noted in 2003:

Public resistance to mandatory exit exams mounted as diplomas were withheld from thousands of students and as high initial failure rates set off alarms in states scheduled to begin withholding diplomas in the next few years. More evidence also emerged about impacts and costs of exit exams, making clearer to states — if they didn't know it already — that exit exams are no cheap or easy fix for education reform (Center on Education Policy, 2003).

Concerns raised about the use of exit exams include reduced graduation rates, especially for African American and Latino students, English language learners, and students with disabilities; reduced incentives for struggling students to stay in school rather than drop out or pursue a GED; narrowing of the curriculum and neglect of higher order performance skills where limited measures are used; and invalid judgments about student learning from reliance on a single set of test measures, a practice discouraged by professional testing experts (Center on Education Policy, 2004).

Washington state has the advantage of the lessons that can be learned from other states' experiences with exit exams and their use of alternatives. The Center on Education Policy (CEP, 2006), has completed annual reports about exit exams since 2002, documenting the experiences of other states. Table 4 provides a listing compiled by the CEP of the range of alternatives being used by states across the country. These include performance tasks or portfolios; alternative tests (such as the ACT, SAT, or even the military entrance examination); evaluation of courses, grades, attendance, teacher recommendations, and work samples; and the combined evaluation of local assessments with the state test. (See also, Darling-Hammond, Rustique-Forrester, & Pecheone, 2005). Several more states have a mandate to provide alternatives, but have not yet decided what they will be.

In addition to these general provisions, nearly all of the states that require exit exams offer alternative measures and sources of evidence for students with disabilities or for English language learners to receive a regular state diploma.

**Table 3: Alternate Paths for General Education Students to Obtain a Diploma, 2006 †**

State	Alternative Methods of Evaluation, Waivers, or Appeals					Accommodations For General Education Students	Alternative Diploma or Certificate
	Alternate assessment	Substitute test	Course grades	Classroom evidence	Criteria-based		
Alabama		✓					
Alaska							
Arizona			✓	✓	✓		
California							
Florida		✓					✓
Georgia				✓	✓		✓
Idaho		✓					
Indiana			✓	✓	✓		
Louisiana							
Maryland	✓	✓					
Massachusetts					✓		
Minnesota						✓	
Mississippi*			✓	✓	✓		
Nevada							✓
New Jersey**	✓						
New Mexico							✓
New York		✓	✓				
North Carolina		✓					✓
Ohio					✓	✓	
Oklahoma***							
South Carolina****							✓
Tennessee							✓
Texas							
Virginia		✓					✓
Washington		✓	✓	✓			

†From Center on Education Policy August 2006 report, "State High School Exit Exams: A Challenging Year."

Table reads: General education students in Florida, under circumstances described in that state's profile, may use a substitute test as an alternative method of evaluation to obtain a diploma. In addition, Florida provides a certificate of completion under specific conditions for students who do not meet the requirements for a regular diploma.

\*Mississippi uses an appeals process whereby students may submit course grades, classroom evidence, or other criteria as evidence that they have mastered the subject being tested (Mississippi Department of Education, 2001).

\*\*New Jersey's Special Review Assessment is being phased out for incoming freshman. As of July 2006, the state had not yet determined either an alternate appeals process or a timeline for this phase-out.

\*\*\*Oklahoma has not yet determined which specific alternative methods it will use.

\*\*\*\*Although South Carolina did not address the issue of certificates in its survey response, the state does give a certificate of attendance to students who do not pass its exit exam, according to correspondence between CEP and a state official.

Source: Center on Education Policy exit exam survey of state departments of education, June 2006, item 13.

Several states, in line with professional testing standards, require that the state high school examinations must be used only in conjunction with other performance measures, including local performance assessments, to make a graduation decision. These include Connecticut, Maine, Oregon, Pennsylvania, and Rhode Island. Connecticut, Maine, and Rhode Island use a combination of performance measures in a collection of evidence system that is similar to what has been proposed as a policy alternative for Washington. In these cases, the local district decides what evidence will be used and how it will be combined. However, the assessments must be shown to align with state standards evaluated on the state assessment. Some states use state tests or local performance assessments to offer as an endorsement on the diploma or place the scores on the transcript as information for colleges and employers, rather than as a requirement for graduation.

**Table 4: Exit Exam Alternatives Used by States**

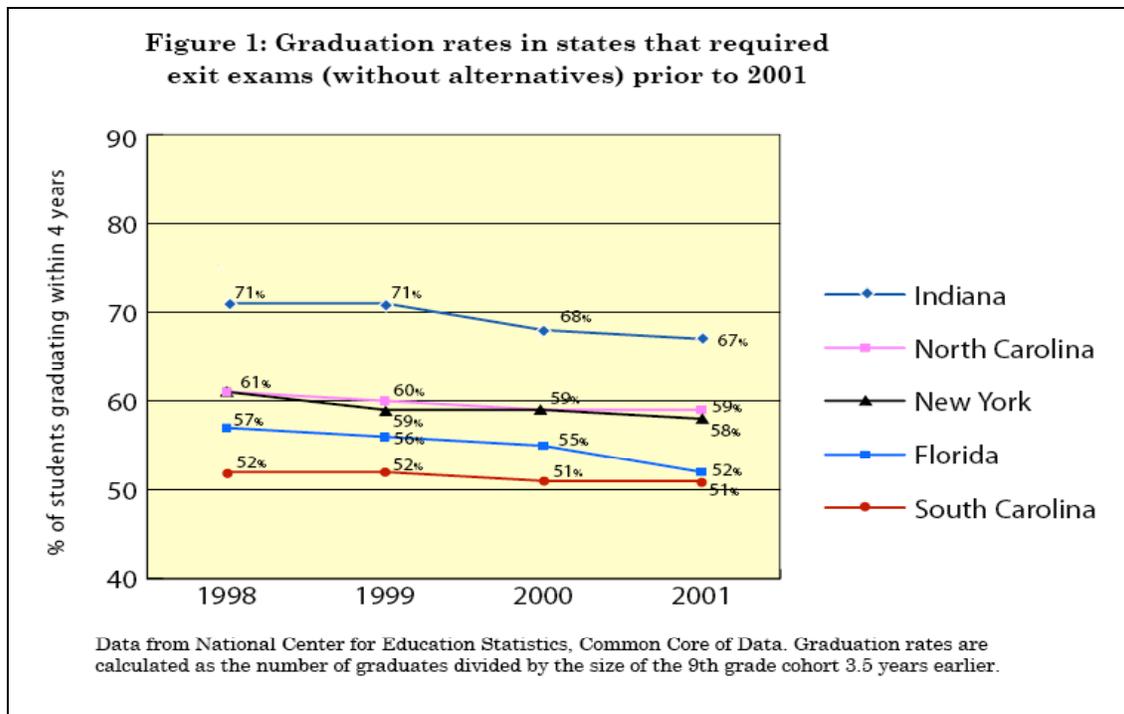
Alternatives	Definition	States Using Alternative
Alternate Assessment	An assessment – usually a performance assessment -- that is aligned with the state standards as an alternative to the exit exam.	MD, NJ use a state performance assessment CT, ME, and PA, among others, allow or require the use of local performance assessments
Substitute Test	The state uses a score on a standardized test, such as the SAT, ACT, AP, or IB as a substitute for its exit exam.	AL, FL, ID, MD, NY, NC, VA, WA
Course Grades	The state allows specific course grades or a GPA as an alternative to the exit exam,	AZ, IN, MS, NY, WA
Classroom Evidence	The state uses the students' coursework or a portfolio of students' coursework.	AZ, GA, IN, MS, WA
Criteria-based	The state uses a variety of criteria on various measures to decide whether a student has met the graduation standard.	AZ, GA, IN, MA, MS, OH
Collection of Evidence	A compilation of the measures listed above.	CT, ME, OR, RI

(Center on Education Policy, 2006; Darling Hammond et al., 2005)

The Center on Education Policy notes in its most recent report that the number of states adding exit exams to their graduation policies has stalled, while more states are adding alternatives for students not passing the test. The CEP report attributes this slow down to the many legal and political battles being fought over the exit exams as well the emergence of data illustrating the negative impact many states' exit exams have had on students (CEP, 2006). Figure 1 shows declining graduation rate trends in the five states that developed exit

examinations in the 1990s and used them, without alternative assessments or options, to withhold diplomas prior to 2001.<sup>1</sup>

Since 2001, other states have implemented exams with similar effects. For example, Massachusetts' exit exam was implemented in 2003, and graduation rates in Massachusetts declined from 76% in 1998 to 72% in 2003 (Massachusetts Department of Education, "Enrollment Data;" Massachusetts Department of Education, "Plans of High School Graduates;" Wheelock, 2003; Bernstein, 2004).<sup>2</sup> As in other states, declines were even more dramatic for African American and Latino students. For example, the proportion of African American students in Massachusetts graduating in 2003 was 60%, as compared to 71% in 2002, the year before the exit exam was instituted; Hispanic students' graduation rates dropped from 54% to 45%, while white students' graduation rates dropped less steeply, from 79% to 77%. Dropout rates in Massachusetts have also increased in the years since the exit exam became a requirement, and have been tied to the exit exam requirement in state department studies (CEP, 2006; Massachusetts Department of Education, October 2005; Massachusetts Department of Education, April 2006).



<sup>1</sup> States with exams that were part of a multiple measures system (i.e. are considered with other evidence about learning) are not represented in this figure, nor are states with exams that allowed a local waiver or option to the test (e.g. Alabama, Georgia, Idaho, Texas). States that enacted new exit exam policies that took effect after 2001 are not represented in this figure.

<sup>2</sup> These rates are calculated by dividing the number of graduates by the number of 9<sup>th</sup> graders 4 years earlier. Enrollment and graduation data are from the Massachusetts Department of Education.

Because of these concerns, four of the five states shown in Figure 1 above have introduced alternatives to their exit exams. In Florida and North Carolina, legislatures have recently required that non-test alternatives for demonstrating competency be developed. In Indiana, coursework and grades can be evaluated for graduation for those who fail the exit exam, and in New York, a variety of substitute tests can be used, as well as performance assessments in some schools.

Although students have always dropped out for a variety of personal and school-related reasons, analysts have identified several test-related reasons for these declines in graduation rates, including:

- Student discouragement after repeated unsuccessful efforts to pass the exams;
- Policies that encourage grade retention of students who do not pass exams at benchmark grade levels – or who are low-achieving in the grade level prior to the one in which the high-stakes exam results are reported – which has been found to substantially increase dropout rates;
- Failure of students to pass the exams, including students who have received lower quality education, students who do not demonstrate their learning well on on-demand tests, students with disabilities, and students who are limited English proficient.
- Incentives for schools to push out students who do poorly in school, when school ratings are contingent on the average pass rates of students. This occurs through encouragements to transfer to other schools or GED programs as well as exclusions for tardiness, attendance or behavior. Some schools have increased their test scores by creating barriers to the enrollment or continuation of low-achieving students (Advocates for Children, 2002; Haney, 2002; Heubert and Hauser, 1999; Jacob, 2001; Lilliard and DeCicca, 2001, Orfield et al., 2004; Roderick et al., 1999; Rumberger and Larson, 1998; Rustique-Forrester, In Press).

A number of studies have found that high school graduation exams increase dropout rates. A large-scale study using individual-level data from the National Educational Longitudinal Survey found that graduation tests increased the probability of dropping out among the lowest achieving students (Jacob, 2001). Another study published by the National Bureau of Economic Research found that students in states with relatively easy exit exams are roughly 4 percent more likely to drop out of high school than similar students in states with no exit exams. In states with relatively difficult exit exams, students are 5.5 percent more likely to drop out of high school. The study also found that exit exams

exaggerated achievement gaps between subgroups (Glen, 2006; Dee and Jacob, 2006).

A recent large-scale study found that, after controlling for students' demographic characteristics (e.g., race, family education and income, GPA and class rank), states requiring graduation exams had lower graduation rates and lower SAT scores than states not using exit exams. Individually, students from states requiring a graduation exam also performed more poorly on the SAT than did students from states not requiring an exam (Marchant and Paulson, 2005). One possible explanation for this is the narrowing of the curriculum that can occur with high-stakes testing.

For all of these reasons and others, states have confronted a variety of legal challenges regarding exit exams. These challenges have often led to policy changes. For example, the most recent legal challenge to the Massachusetts exit exam was settled out of court by the state, which agreed to implement a set of changes to the current graduation policy, including:

- Eliminating the requirement that students earn a minimum score of 216 on the state's exit exam test to qualify for a performance appeal, while maintaining the same substantive standards for evaluating that appeal.
- Taking steps to reduce the number of students who drop out of school, including setting a state standard for a "high school graduation rate" that school districts should strive to achieve.
- Taking additional steps to ensure that students with disabilities have access to the curriculum taught to all students.
- Developing guidance for school districts on improving classroom instruction for limited English proficient students
- Requiring school districts to provide students who do not pass the states exit exam with written notice of post-high school opportunities for learning the necessary material and retaking the test (Massachusetts Department of Education, 2006).

Legal, political, and legislative issues have caused many states to become much more cautious in regard to their exit exam policies, in most cases developing alternatives that include non-test options.

## **Rationales for Alternatives**

Numerous sources support the validity of using more than one measure to assess student learning, also known as a multiple measures approach. Professional testing standards emphasize that no test is sufficiently reliable and valid to be the sole source of important decisions. Instead, the standards urge that high-stakes decisions about such matters as student placement, promotion, or graduation should rely on several different kinds of evidence about student learning. The three national professional research organizations concerned with

testing and measurement state in Standard 13.7 of their joint publication *Standards for Educational and Psychological Testing*:

In educational settings, a decision or characterization that will have major impact on a student should not be made on the basis of a single test score. Other relevant information should be taken into account if it will enhance the overall validity of the decision (AERA, APA, and NCME, 1999, p. 146).

The standards suggest that the other types of information to consider when assessing students' skills and knowledge should include alternative assessments like samples of student schoolwork, portfolios, grades, end of course examinations, or classroom observations that will provide information about student performance.

These concerns are especially important for students with learning differences who may require different formats to demonstrate their knowledge. Many students are not validly assessed by standardized tests because of the artificial format which is removed from real-world applications of knowledge, the use of distractors among responses, complex language that impedes getting a clear performance on the construct of interest, and class- or culture-based differences in the life experiences students have had that may be used as background knowledge for questions (e.g. sledding for New Mexico students, stocks and bonds for low-income students). Many tests also do not meet universal design principles that have been specifically outlined for assessments. These principles (listed on Table 5) make tests more user-friendly, as well as valid and accessible for all students.

**Table 5: Elements of Universally Designed Assessments**

<b>Element</b>	<b>Explanation</b>
Inclusive Assessment Population	Tests designed for state, district, or school accountability must include every student except those in the alternate assessment, and this is reflected in assessment design and field-testing procedures.
Precisely Defined Concepts	The specific constructs tested must be clearly defined so that all construct irrelevant cognitive, sensory, emotional, and physical barriers can be removed.
Accessible, Non-Biased Items	Accessibility is built into items from the beginning, and bias review procedures ensure that quality is retained in all items.
Amenable to Accommodations	The test design facilitates the use of needed accommodations (e.g., all items can be Brailled).
Simple, Clear, and Intuitive Instructions and Procedures	All instructions and procedures are simple, clear, and presented in understandable language.

Maximum Readability and Comprehensibility	A variety of readability and plain language guidelines are followed (e.g., sentence length and number of difficult words are kept to a minimum) to produce readable and comprehensible text.
Maximum Legibility	Characteristics that ensure easy decipherability are applied to text, tables, figures, illustrations, and response formats.

Based on Thompson, Johnstone, and Thurlow (2002)

A multiple measures approach also increases validity by accessing skills and knowledge that tests cannot alone measure. As Achieve has noted:

States. . . will need to move beyond large-scale assessments because, as critical as they are, they cannot measure everything that matters in a young person’s education. The ability to make effective oral arguments and conduct significant research projects are considered essential skills by both employers and postsecondary educators, but these skills are very difficult to assess on a paper-and-pencil test (Achieve, Inc. 2004, p. 3).

With a multiple measures approach that goes beyond the consideration of on-demand tests, evaluation of learning does not have to happen in one day – or even several days – on one test. A system that includes measures beyond tests can incorporate more challenging and authentic tasks like extended writing assignments, scientific inquiries, and in depth research projects that require planning, data collection, and analysis. Even open-ended items on tests do not enable students to use the kinds of skills that will actually prepare students for success later in life. A multiple measures approach that includes these ambitious kinds of assessment can encourage districts and schools to implement assignments that will prepare students not only to meet high school standards but also to meet expectations in college and beyond.

A multiple measures system can also provide more information than a set of on-demand tests about what a student can do and where more attention is needed, in a way that allows for more diagnostic teaching. For example, collections of evidence or portfolio systems can help teachers assess student learning at multiple points in time on several dimensions of the standards, and provide a cumulative assessment of students’ skills and knowledge. Such a system can also show how a student performs in different genres or on different kinds of tasks. Thus, a multiple measures approach with a dynamic set of assessments can provide more opportunities to assess different aspects of students’ learning of state content standards as well as to provide information for improvement.

A multiple measures approach can strengthen pedagogy by using assessments that encourage teachers to teach students problem solving, analysis, and evaluation of information. Research shows that teachers working in high stakes accountability systems teach the skills required for students to do

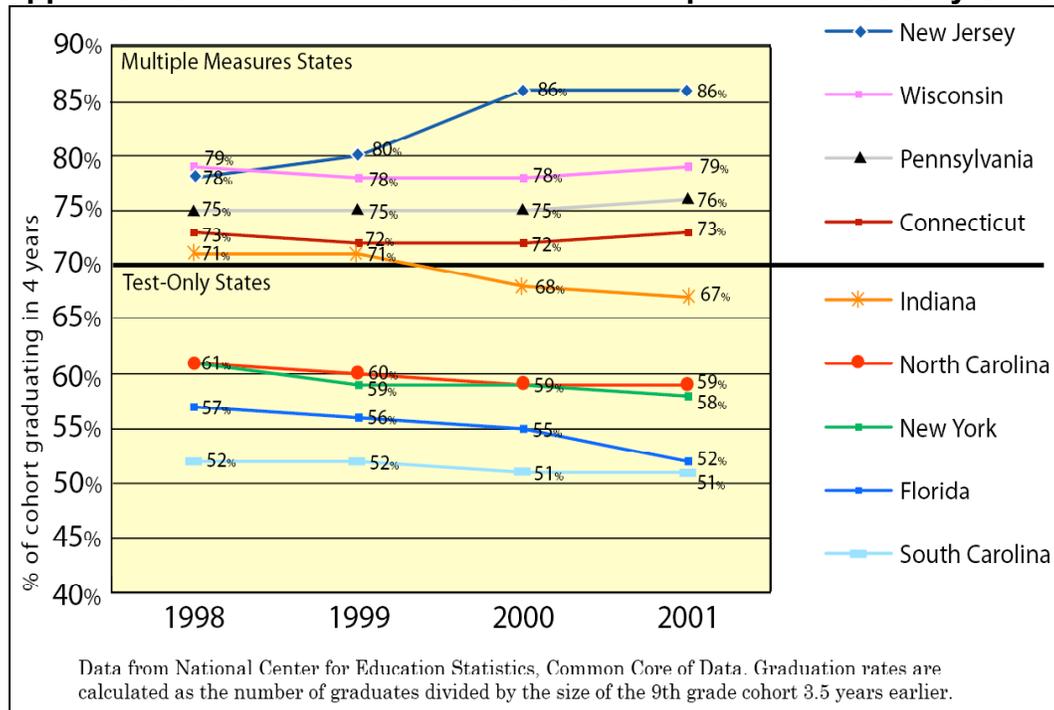
well on tests (Clarke et al., 2003; Pedulla et al., 2003; Darling-Hammond, et al., 2005). When tests narrowly define what counts as learning, teachers typically respond by narrowing their instruction to the skills and knowledge only necessary to pass the test (Darling-Hammond and Rustique-Forrester, 2005). For example, in a survey of 1,000 teachers by *Education Week*, 85% of the teachers reported that their school gives less attention to kinds of learning that are not represented on the state test (Education Week, 2001). As a Texas teacher noted:

I have seen more students who can pass the TAAS but cannot apply those skills to anything if it's not in the TAAS format. I have students who can do the test but can't look up words in a dictionary and understand the different meanings.... As for higher quality teaching, I'm not sure I would call it that. Because of the pressure for passing scores, more and more time is spent practicing the test and putting everything in TAAS format (Haney, 2000, part 6).

On the other hand, teachers' instruction is influenced positively where measures are used that access a wider, more dynamic set of skills. Studies have found that states like Kentucky, Vermont, and Connecticut have improved instruction as well as student performance by using portfolios and authentic performance tasks that require extensive writing, research projects and mathematical problem-solving. This combination of assessments provides room to include skills like complex thinking and communication that are often not adequately measured on multiple-choice, short answer tests. In addition, the local assessment components of these state systems (portfolios and performance tasks) encouraged teachers to learn and use formative assessment strategies, which have been shown to offer a particularly powerful means of improving student learning, especially for previously low-performing students (Darling-Hammond & Rustique-Forrester, 2005).

Finally, a multiple measures approach can help increase standards and improve achievement while maintaining high school graduation rates. A Stanford University study found that, in contrast to the declining graduation rates in many test-only states, states that used a multiple measures approach to graduation during the 1990s had stable or increasing graduation rates (Darling-Hammond, Rustique-Forrester, & Pecheone, 2005). A multiple measures approach may keep students engaged in school by giving teachers more useful information about learning that allows them to adapt their teaching to support individual students and by offering a wider range of opportunities for students to demonstrate their learning.

**Figure 2: Graduation Rates in States Requiring Multiple Measures Approaches to Graduation Prior to 2001 Compared to Test-Only States**



*From Darling-Hammond, et al. (2005) Multiple Measures Approaches to High School Graduation*

## Evaluation of Options

As we have noted, there are many different approaches to alternatives used by states across the country. In this section, we review various approaches, how they have been or might be used, and the strengths and limitations of these approaches.

### **Alternate or Substitute tests**

Washington's current legislation institutes an alternative using substitute tests in a limited fashion. This option asks that students obtain a score on the mathematics components of the PSAT, ACT, or SAT that is deemed equivalent to passing the WASL in mathematics. This does provide one alternative pathway for students not passing the mathematics portion of the WASL, it is limited by the tests selected and by its use only in mathematics. In line with other states' practices, this option could be considerably broadened.

A number of states allow students to use scores from alternative or substitute tests as alternatives to their exit exams. In some cases, states allow students to use an alternative assessment created by the state (NJ, MD) or local district (CT, ME) that uses a performance task or portfolio format rather than a

traditional testing format so students who traditionally do not test well on standardized tests will be able to show their knowledge in another way. Often these assessments can be considered after a student has failed the exit exam, but in some cases, they are considered as equivalent measures and can be used as evidence in lieu of the exit exam.

In other cases, states honor scores from substitute tests like the SAT or ACT in order to encourage students to take these tests for pre-collegiate requirements or to enroll in advanced classes that have rigorous end-of-course exams like International Baccalaureate (IB) and Advanced Placement (AP). To minimize testing burdens on students, these substitute tests are frequently accepted in lieu of the exit exam. As seen in Idaho and other states, more and more states are also honoring students' scores on exit exams from other states when students move from another state (Idaho State Board of Education, 2005).

**Table 6: Key Definitions**

<p><b>Alternate Assessment:</b> An assessment developed by a state or local district for students to take as an alternative to the state's exit exam. Typically the assessment is aligned with state standards and measures students' skills and knowledge in a different way than the exit exam, often through a set of performance tasks, work samples, or a portfolio.</p>
<p><b>Substitute Test:</b> A test that measures students' skills and knowledge in areas related to the exit exam that is developed by an independent organization. Substitute tests typically include measures like the ACT, SAT, AP, IB, or even the military entrance examination.</p>

Of the 25 states with planned or current exit exams, nine states allow the direct use of alternate or substitute assessments (Center on Education Policy, 2006). Maryland and New Jersey allow students to use alternative assessments that are created by the state, while Alabama, Florida, Idaho, Maryland, New York, North Carolina, Virginia and now Washington allow students to use substitute tests. New York also allows the use of alternative performance assessments for some schools that have developed a strong system of assessments and secured a special waiver from the state. Other states, including Connecticut, Pennsylvania, and Maine, require or allow locally determined alternative assessments that complement or substitute for the state exam.

Among the states that allow substitute tests, many of the assessments are college admissions tests like the SAT and the ACT. As one example, students in Florida who do not pass that state's exit exam after three attempts may use scores from the ACT or SAT to satisfy the graduation requirement. Table 7 shows the corresponding scores selected by Florida to allow for this substitute assessment (Florida Department of Education, 2005).

**Table 7: Corresponding Scores for Florida's Exit Exam and Substitute Tests**

Reading		Mathematics	
FCAT	1926	FCAT	1889
SAT	410	SAT	370
ACT	15	ACT	15

Other tests used as substitute tests include SAT II, IB and AP end-of-course exams, and the University of Cambridge Advanced International Certificate of Education (AICE). New York, Virginia, and Maryland allow students to substitute scores from end-of-course IB and AP exams for their graduation requirements. Students who take specific AP or IB exams will be able to meet Maryland's end-of-course exam requirement as long as they receive a score of a 3 or higher on the AP exam or a 4 or higher on the IB exam (Maryland State Department of Education, 2006). States like Idaho allow students who do not pass their exit exam to appeal for another measure approved by their local school board. Alternative measures are not specified by the state, but they must be aligned at a minimum to 10th grade state content standards for the subject matter in question and be valid and reliable. In addition, the criteria of the measure (or combination of measures) must be based on academic proficiency and performance (Idaho State Board of Education, 2004).

The AICE is the most sophisticated of the assessments used as substitute exams, as it evaluates students' learning of a curriculum that promotes in-depth skills and knowledge. Only a few states, including New York and Virginia, allow students to use the AICE as a substitute test at this point, but other states are considering it, and some believe it could inform future assessment policy more widely. The AICE is an internationally recognized curriculum and examination program modeled on the British pre-collegiate curriculum and "A-level" exams. Started in 1994, it was piloted in Florida in 1997 and 2000 where it is still used and funded today. The assessments are closely aligned to a set of courses based on three strands of curriculum: 1) mathematics and science, 2) languages, and 3) arts and humanities. Students must earn six credits in these three strands of curriculum by taking a combination of courses and exams based on specific curriculum designed by the University of Cambridge. One of the credits may be achieved by completing a certified research project. The introduction of this option in 2002 was prompted by a call for more emphasis on the Cambridge curriculum's focus on independent research and investigation, the use of initiative and creativity, and the application of skills and knowledge (University of Cambridge International Examinations).

Another aspect of the AICE that is important to recognize is that, like most examination systems in Europe, Asia, Canada, Australia and the Caribbean, it incorporates both on-demand tests and syllabus-based student work samples – like research papers – directly into the examination system and scoring process. Thus, rather than adding on a variety of options, substitutes, or alternatives, the

assessment inherently provides a multiple measures system for all students, one which increases both the challenge and rigor of the work that students are asked to demonstrate and enhances the validity of decisions inferred from the scores, since these are based on a combination of assessments. We will return to this property of many other countries' examination systems when we discuss the collection of evidence alternative later.

One other test which could easily be used as an alternative to an exit exam but has not yet been accessed by states for this purpose, in the Northwest Evaluation Association's Measures of Academic Progress (MAP) assessment, a computerized adaptive test that measures students' instructional level and growth over time. Over 2200 school districts and educational partners use NWEA's MAP assessments. Since MAP assessments are aligned with state standards, they are often used as an indicator of a student's preparation for state assessments. States like Idaho use the MAP assessment for informative purposes. States like Delaware participating in NCLB's pilot growth model will start using the MAP assessments to measure growth (Northwest Evaluation Association Website; Idaho State Department of Education; State of Delaware, 2006). A computerized adaptive test like MAP would allow the assessment to be more diagnostic and informative in support of instruction.

There are a number of potential benefits when states use alternate or substitute tests. The first potential benefit is to students who have options to demonstrate their learning. This may both make the assessment decision more valid, by providing alternative means to demonstrate knowledge, and, in some cases, enhance student learning opportunities. For example, students may be encouraged to take advanced classes or take tests important to post-secondary decisions. In other cases, as seen with New Jersey's Special Review Assessment (SRA), which offers a set of Performance Assessment Tasks embedded in remedial coursework for students who have not passed the exit exam, the alternate tests measure students skills and knowledge in a more dynamic way that can better inform ongoing instruction and provide a more thorough assessment of students' learning (New Jersey Office of Assessment and Evaluation). Also, alternate assessments are helpful for students, especially special education students and students with limited English who may be able to show their learning on performance assessments that measure performance in less artificial ways, i.e. without distracters and other "tricks" of standardized tests that often unnecessarily confuse students.

Finally, allowing substitute tests can reduce overall test burden, where these are tests students would be taking in any event and where, as in New York, students do not have to fail the state test in order to use substitute test scores. Students may already be taking some of these tests, thus making scores easily accessible. This option is lower in cost compared to other alternatives.

A down side of many substitute tests is that as on-demand tests, they are still a fairly remote proxy for actual performance and do not necessarily have high predictive validity. This is especially true of tests like the SAT, which is mostly multiple-choice and not curriculum-based. Studies on the SAT find, for example, that it predicts a small proportion – less than 20% -- of the variation in freshman year grades, because it does not measure many kinds of knowledge, skills, and abilities that matter in college, including the ability to find and organize information, write extensively, plan and manage multi-faceted tasks, be organized, exercise self-discipline and so on. More extensive tasks, like research papers, are better measures of these abilities. Tests like the IB and AICE have stronger curricular validity and are more authentic– in many cases requiring in-depth analysis and explanation and allowing for more ambitious work samples. These are closer to actual performance and may have more salutary effects on instruction. Alternative assessments may also be constructed to be more performance-oriented and thus closer to the kinds of performances that are actually called for in real-life situations.

Other considerations include the added costs if Washington chooses to pay test entrance fees for students from low socio-economic backgrounds, and the fact that substitute tests were not directly constructed to measure state standards, although there is likely to be considerable overlap. (See Table 8 for a list of potential benefits and limitations for using alternate and substitute tests.)

**Table 8: Potential Benefits and Limitations of using Alternate or Substitute Tests**

POTENTIAL BENEFITS	POTENTIAL LIMITATIONS
<ul style="list-style-type: none"> <li>● Substitute and alternative tests provide another measure of learning to enhance the validity of the high-stakes decision</li> <li>● Substitute tests are low cost to the state</li> <li>● Students already take some of these tests</li> <li>● Some options can encourage students to pursue stronger coursework and college</li> <li>● Alternatives that are performance tasks can be aligned to state standards and measure students knowledge and skills in a way that may be more valid for special needs students</li> <li>● Alternatives can provide assessments that measure performance in less artificial ways, i.e. without distracters and other tricks</li> <li>● Some alternatives have potential for curricular validity</li> </ul>	<ul style="list-style-type: none"> <li>● Tests as measures generally have fairly low levels of predictive validity</li> <li>● Substitute tests may not measure exactly the same standards as the WASL, although there is likely to be considerable overlap</li> <li>● Students may need financial support to take the tests</li> <li>● Alternative tests would need to be developed if they were to be considered as an option</li> </ul>

**Recommendations.** The idea of allowing substitute tests as an alternative to the WASL is a sound idea, used in many other states. Although an alternative assessment constructed by the state could also be recommended, the collection of evidence option (discussed below), if well constructed, could serve equally well in this role for most students. The enhanced validity of inferences about student achievement provided by the current provision for a substitute test could be further enhanced in three ways.

- Allow the use of substitute tests for subjects beyond mathematics. The rationale for such tests is equally compelling in reading, writing, and science and, although fewer students overall are currently failing the reading and writing tests, the extent to which a given test may be more or less valid for various students or subpopulations can differ from one subject area or test to the next.
- Allow the use of a wider range of substitute tests, including those that are performance-based and emphasize ambitious tasks which may have salutary effects on the curriculum and students' course-taking (e.g. the AP, IB, or AICE) and those that can be used diagnostically to measure growth (e.g. the MAP), as these can help to strengthen instruction. By providing a broader range of substitute tests, as some other states do, Washington could provide more options for students while increasing the skills and knowledge accessed by exit level assessments.
- Allow the use of a substitute test routinely, as in some other states, or after only one failure of the WASL (if the state's desire is to be sure all students take the WASL for purposes of data-gathering and instructional guidance). This will reduce test burden and allow the state to conduct research on student performance on multiple measures in relation to WASL performance, which should be helpful in the analyses needed to continually fine-tune the tests and the curriculum.
- In addition to providing for the use of substitute tests, the state should evaluate the design of the current assessments to see how well they meet universal design principles and how well they assess the knowledge and skills of English language learners. It may be that alternate forms of the test should be considered to reduce unnecessary linguistic complexity. Research has found that such modifications can provide a more valid assessment of English learners' knowledge and skills and improve performance (Abedi, Lord, & Plummer, 1997). The Center for Research on Evaluation, Student Standards, and Testing (CRESST) recommends that all state tests undergo rigorous review for language difficulty and that test questions be modified to reduce the level of unnecessary linguistic complexity and cultural bias (Abedi & Dietel, 2004).

### ***Grade Point Average***

Some states use students' Grade Point Average (GPA) as an alternative assessment to their exit exam. Most of these states use GPAs in a broader way than is currently proposed in Washington State. Washington's current alternative

that incorporates GPAs compares the students' grades with the grades of other students who took the same courses and who met the standard on the WASL.

In some respects, there is even stronger evidence for using a student's coursework and grades as a supplement to an exit exam than there is for using substitute tests, since a student's grades represent a body of work indicating student learning over time, whereas an exam gives only a snap shot of a small sample of student learning evaluated in a fashion that is removed from the ways in which knowledge is usually demonstrated. In fact, research shows that GPAs and class rank are typically a stronger predictor of college success than test scores (see e.g. Fernald, 2002; Walker, et al., 2002). As a result of this research, many current university admissions policies give less weight to SAT scores and place more weight on class rank and GPAs.

At least six states -- including Arizona, Indiana, Mississippi, New York, Massachusetts, Ohio, and now Washington -- factor in course grades as a substitute for passing an exit exam (Center on Education Policy, 2006).

Arizona's use of GPA as an augmentation to the exit exam factors in additional information about student performance and provides incentives for students to work hard in their courses. Among the alternatives offered in Arizona is an option that allows students to augment their highest exit exam score with additional points derived from classroom performance in specific courses that satisfy the standard. An "A" provides an additional 20 points per credit. A "B" is 15 times and "C" is 12 times the amount of credit for the class. These points are transformed into a proportional augmentation to the student's test score. Thus, an "A" in a one-credit course could increase a student's exit exam score by 20%. To get a sense of the calibration for the test, the score scale for the reading portion of the test indicates that a score of 500-626 is "Falls Far Below," 627-673 is "Approaches," 674-772 is "Meets," and 773-900 is "Exceeds" the standard (Arizona Department of Education, Spring 2006).

**Figure 3: Arizona's formula for calculating a pupil's score augmentation**

$$\frac{\text{Avg. Additional Points per Credit}}{100} \times (\text{Pupil's Original Score}) = \text{Augmentation points}$$

*(Arizona Department of Education)*

New York also allows course grades to substitute for exit exam scores under certain circumstances. Students who entered ninth grade after 2005 in New York who do not receive a passing grade on their Regents Examinations may use their course grade to substitute for a passing score on the Regents

exam if they have passed certain courses intended to prepare them for the Regents exam and have earned a 65 course average. In addition, their highest score on the Regents exam needs to be within three points of the 65 passing score (New York State Education Department, “Regents and High School Diploma Graduation Requirements”). This attention to both specific courses and grades provides information about a more comprehensive body of evidence representing students’ learning than the single test.

Massachusetts has a useful, but limited approach to factoring in the GPA as an alternative to its exit exam. In Massachusetts, once students meet a set of criteria and are eligible for an appeal, one of two alternatives includes a cohort GPA comparison, very similar to Washington’s proposed model. The Massachusetts criteria for filing an appeal have until recently included: having taken the exit exam at least three times, achieving at least a score of 216 on one try, participating in exit exam tutoring or mentoring programs, and maintaining a 95% attendance rate. A settlement in a recent lawsuit has eliminated the 216 score criterion as a prerequisite to the appeal. Now any student can take advantage of the GPA alternative.

**Table 9: Key Definitions**

<p><b>Grade Point Average (GPA):</b> States use a point system based on an average of student grades to summarize students’ performance in their courses. The GPA may be based on coursework and assessment for specific courses.</p>
<p><b>GPA Cohort Comparison:</b> States compare GPAs in specific courses for a group of students who passed the exit exam to that of a student who did not pass the exit exam.</p>

The specifics of the Massachusetts GPA cohort comparison include completing a cohort comparison worksheet that compares the student’s GPAs in courses taken in grades 10 and 11 in the area of the appeal to all of the students (but at least 6 others) in the school who took the same sequence of courses and passed the state exit exam test with scores between 220-228. When there are not six other students who fit these criteria, a collection of evidence or portfolio can be used. (This pathway will be discussed under collection of evidence.) This policy provides only a very limited approach to using GPAs in graduation decisions because it is available under very restricted conditions. The number of appeals granted in Massachusetts was only about 3,000 (of 5,000 filed) between 2002 and 2005 (Massachusetts Department of Education, 2005).

In a study conducted for the Office of the Superintendent of Public Instruction in Washington, David Conley (2005) found a moderately strong relationship between students’ GPAs and WASL scores: The results included a .52 correlation between the WASL mathematics scores and students’ GPAs in courses rated as having high content match to the WASL, a correlation of .39 between students’ GPAs in matched courses and reading WASL scores, and a .33 correlation between students’ GPA in matched courses and the writing

WASL. This suggests that the GPA may be a useful complement to WASL scores. An obvious concern about the use of grades may be the reliability of grades across teachers and schools, yet the evidence that grades more accurately predict success in college than do test score suggests that grades are nonetheless useful measures of student learning, and tests have their own limitations. While there may be some data collection challenges associated with this option, Conley expects improvements in the state data system to allow for GPAs to be a productive option.

(See Table 10 for a list of potential benefits and difficulties of using GPA cohort comparisons.)

**Table 10: Potential Difficulties, Benefits of using GPAs**

POTENTIAL BENEFITS	POTENTIAL DIFFICULTIES
<ul style="list-style-type: none"> <li>● Represents a more complete body of work</li> <li>● Provides additional evidence</li> <li>● A stronger predictor of college success than tests</li> <li>● Low cost</li> <li>● Encourages alignment of courses and state standards</li> </ul>	<ul style="list-style-type: none"> <li>● Variability in grades across teachers and schools</li> <li>● Data collection issues must be worked out</li> </ul>

Even given the limitations, GPAs have some real benefits: they represent a body of student work and are measured at a relatively low cost, since most of the structure for this assessment is already in place. Research suggests that GPAs are a stronger predictor of college success than test scores like the SAT. If properly administered, the use of GPAs as an alternative could encourage schools to align their courses to state standards, using grades from those courses to inform graduation decisions.

**Recommendations.** While the current proposal for using GPAs as an alternative to passing the exit exam is a reasonable one, it is likely to be quite limited in its effects. Washington could also consider an alternative that acknowledges student coursework as a more integral part of the graduation decision in the way that Arizona’s approach integrates grades in specific courses into the augmented exit exam score or New York’s approach counts course grades in courses organized around the state standards assessed on the test. Courses considered as evidence of performance should be directly linked to the WASL standards. Using GPAs in this way would encourage teachers to more closely align their instruction with the WASL standards, and would honor the work students complete in courses, giving them an incentive to work hard on their schoolwork.

***Collection of Evidence***

Currently, Washington presents a third alternative that evaluates a collection of work samples (also known as a collection of evidence) with specific

collections designed for students in programs leading to a national or state industry certificate. This is the most extensive alternative of the three alternatives authorized by the Washington legislature.

**Table 11: Key Definitions**

<p><b>Collection of Evidence:</b> A collection of work samples such as papers, projects, experiments, reports, and other assignments that are representative of student work that meets learning standards. Collections of evidence can be as simple as a coursework file or organized as a set of work samples, performance tasks, or a portfolio meeting specifications as described below.</p>
<p><b>Portfolio:</b> A portfolio offers a specific approach to a collection of evidence. It usually contains a set of work samples that address certain learning standards through the collected work, and is often organized to include student commentary or reflections on the work and how it meets the standards. A portfolio may include progressive drafts or work samples showing revision and improvement in the quality of the work.</p>

Assessments based on a collection of evidence are in one sense the most valid assessment of students' learning and performance, because the evidence comes directly from the learning process and is presented in an authentic fashion. Also, a collection of evidence approach increases validity by expanding the amount of data available and the range of abilities assessed, accessing skills and knowledge that can not always be accessed by paper and pencil tests. A collection of evidence can be designed to allow assessment of a dynamic set of skills reflected in the standards, including planning; finding, organizing, and managing resources; synthesizing and analyzing information; problem-solving, and multi-tasking, that are not assessed by traditional standardized tests.

While standardized tests are seen as more reliable, a collection of evidence from samples of student work is generally a more valid measure of learning because this evidence is closer to the ways in which knowledge is generally used and demonstrated. Research has found that these samples can also be scored reliably, both individually and when they are assembled into a collection or portfolio, when the specifications for producing the work are clear, well-designed scoring rubrics have been established, and scorers are trained.

While standardized on-demand tests are less costly to implement, they offer a more remote proxy for student learning, because they are disconnected from the teaching and learning process, and they offer a much more artificial format for displaying knowledge and skills. A collection of evidence can be used to evaluate more of the abilities students have developed in their learning. The assessments included in such a collection also offer teachers information about student thinking and learning that is typically not available from multiple-choice or short answer tests, and may not be made available in a timely way from any test that is scored outside the school with information sent back in the form of scores rather than direct work samples teachers can see and evaluate.

There are many different approaches to a collection of evidence assessment, ranging from portfolios to individual work samples or tasks that are written or constructed to specifications and scored using a common set of standards or rubrics. In the most limited use of these approaches, several states provide a portfolio alternative or specific collection of student work for some students who apply for appeals from the exit exam or who fall into protected classifications, such as students with disabilities or English language learners.

In Massachusetts, for example, portfolios may be used as an alternative for students fail the exit exam and who do not meet the requirements for the GPA cohort comparison during the appeals process because of an insufficient number of cohort members for comparison. The portfolios must reflect the Massachusetts learning standards and include a table of contents that lists each piece of evidence submitted with the learning standards addressed by the evidence. The standards demonstrated in the ELA portfolio must include language (vocabulary, grammar and usage, mechanics) as seen in the writing or other work samples, reading and literature (three pieces of writing reflecting on a piece of grade 10 literature that discusses the meaning of a work of literary non-fiction, fiction, and poetry or drama as well as two writing samples that are compositions or essay that discusses a theme in literature and another sample of reflective, persuasive or creative writing). Multiple drafts of each work sample must be included.

The standards demonstrated in the mathematics portfolio must include work samples from five learning strands and its various learning standards. The strands are 1) number sense and operations, 2) patterns, relations, and algebra, 3) geometry, 4) measurement, and 5) data analysis, statistics, and probability. The mathematics portfolio should include four, independently produced, teacher-scored work examples of problems solved correctly for each learning standard. Evidence of students' thinking and problem-solving are important, and the work sample should preferably be original work. A percentage must be indicated on each piece of evidence regarding how much of the work was completely independently by the student and any accommodations are noted (Massachusetts Department of Education, May 2006).

Some states use classroom evidence as an alternative measure for evaluation, waiver or appeal, including a collection of coursework students have completed in class that demonstrates certain graduation standards. States using the classroom evidence approach include Arizona, Georgia, Indiana, Mississippi, Nebraska, and Ohio. In some cases, as in Indiana, either classroom work or a substitute test can be used as evidence of meeting standards for student requesting a waiver from the exit exam (Indiana Department of Education, 2006).

In a portfolio approach, the evidence generally includes multiple, multi-dimensional products from students' learning that are worked on during courses and added to the collection through a structured process that involves a set of criteria. Perhaps the state with the greatest experience with state-level portfolio

assessment is Vermont, which began the practice in 1988. Vermont's Comprehensive Assessment System (CAS) includes statewide portfolios and locally developed assessments in writing and mathematics, as well as statewide reference exams in English/language arts, mathematics, and writing that feature multiple choice questions, open-ended responses, and hands-on performance tasks. Each district must develop a local assessment system consistent with the Vermont Comprehensive Assessment System (CAS) which assesses students in both classroom-based and school-wide assessments in the Vermont Fields of Knowledge (art, language, and literature; history and social sciences; science, mathematics, and technology) and Vital Results (communications, reasoning and problem solving; personal development; and civic responsibility). To graduate from high school, students must demonstrate that they are attaining or meeting the Vermont standards as measured by results on state or local performance-based assessments or that they have completed a set of specified course credits in the core academic subjects, the arts, and physical education (Koretz, et al, 1992; State of Vermont Department of Education).

Studies found that these kinds of performance assessments, used by Vermont, Kentucky, and other states in the 1990's, caused teachers to assign more writing and mathematical problem solving similar to what was expected in statewide portfolios and locally developed assessment (Koretz, Stecher, & Deibert, 1992). While early studies of Vermont's initial portfolio system found that it was difficult to score reliably, later studies have found that, with more standardization of the portfolio components, and more work on moderated scoring with teachers, the portfolios can be scored with consistency and common standards. Indeed, teachers' participation in this process has been found to be an important professional development tool, supporting stronger instruction, ongoing curriculum improvements, and common understanding of the standards (Darling-Hammond & Ancess, 1994; Falk & Ort, 1997; Goldberg & Rosewell, 2000; Murnane & Levy, 1996).

Oregon uses a work sample model to assess students' learning as part of its high school graduation requirement. In order to graduate, Oregon students need to pass required courses and complete specific work samples in English, mathematics, science and social science, evaluated using a state scoring guide. They also need to build a collection of evidence to demonstrate extended application; demonstrate career-related knowledge and skills in areas like problem solving, communication, and teamwork; and develop an education plan. If students meet benchmark levels they can earn a Certificate of Initial Mastery in specific subject areas or overall. Students can also earn a Certificate of Advanced Mastery by meeting higher standards.

The Oregon system is built around the presumption that students should demonstrate their mastery of essential skills on authentic tasks. Students at the Certificate of Initial Mastery level must complete two work samples from different strands in mathematics problem solving and three samples in writing (one

persuasive, one expository and one imaginative or narrative). In speaking, CIM students must complete three samples (one persuasive, one informative, and one unrehearsed). For scientific inquiry and for social science analysis, students must complete one work sample in each, scored in a set of required dimensions. The state provides examples or models of content area tasks that may be utilized by districts as they develop their own local assessments. Work samples are locally scored, preferably by trained teachers, using official state scoring guides. Scores are considered reliable and consistent because of the use of benchmark exemplars of each of the score points for each trait from the official scoring guide (Oregon Department of Education Website, 2006d).

A student in grades 9-12 may also request a juried assessment as an option to the statewide CIM Benchmark assessment procedures, when he or she has mastered the standards for one or more content areas of the CIM, but is unable to demonstrate mastery through related statewide assessments. A juried assessment uses a collection of material to decide if a student has met the standards. Similarly, there are several options for earning a Certificate of Advanced Mastery. To earn this certificate in each subject area, a student may:

1. Demonstrate mastery through the state reading test and provide three speaking work samples and either three writing work samples or the state writing test.
2. Demonstrate mastery through one of the mathematics tests or two math, problem solving work samples, and either the science test or scientific inquiry work samples.
3. Develop an education plan and profile, and participate in career-related learning experiences as outlined in the education plan.
4. Demonstrate extended application through a collection of evidence and demonstrate career-related knowledge to meet standards adopted by the State Board.

Students who do not receive a Certificate of Initial Mastery receive from the school district a Certificate of Achievement that represents the student's progress toward achieving CIM performance standards in each applicable content area. The state K-12 assessment system is articulated with the higher education admissions system. The Proficiency-based Admissions Standards System (PASS) links the Certificates of Initial and Advanced Mastery to college admissions and the requirements of college-level work, thus enabling students to move continuously through secondary school to higher education based on their performance. (Oregon Department of Education Website, 2006a; Oregon Department of Education Website, 2006b; Oregon Department of Education

This work sample approach is not unlike the approach used in Great Britain, Canada, Australia, and many countries in Europe. In these countries, a substantial share of the exam score (generally between 20% and 50%) is comprised of syllabus-based work samples scored by teachers. The other

portion of the score comprises the sit-down test. The A-level or advanced level courses in Great Britain are considered pre-collegiate courses that students choose based on their career choices. The two-year long courses assess students using classroom assignments (known as coursework in Great Britain) and a set of formative assessments or one summative assessment. The summative assessments are written exams that students sit for, similar to the state tests in the United States. However, the difference with these exams is they are aligned completely with the course syllabi and are written year to year according to the changes in the syllabi. Coursework in Great Britain is considered a type of assessment of student performance consisting of projects and tasks specifically outlined in the course syllabi. Coursework that is assessed may include extended essays, investigations, practical experiments or performance work. For example, in the history A-level course, students must complete two written assignments totaling 2,500–3,000 words with the expectation that these are taught in the course. The coursework is graded locally by teachers (Qualifications and Curriculum Authority, 2005a; Qualifications and Curriculum Authority, 2005b; Website, 2006c).

Within each of these collections of evidence models, students perform tasks that respond to state standards and specifications. For the most part, the products from those tasks are then scored locally using state or national rubrics. The coursework in the collection is embedded in class activities and projects that encourage students to demonstrate the skills they learn.

Dr. David Conley conducted a feasibility report on alternatives to the WASL which comprehensively outlines the organization and management of a collection of evidence alternative for Washington. Conley describes how a collection of evidence approach would derive from actual assignments and materials being covered by the teacher. Collections of evidence would be judged on the same scoring criteria across the state with samples or benchmark materials provided. This alternative would provide students with formative assessment feedback from teachers during the class work and final assessment by off-site scorers. Among the benefits of a collection of evidence approach are greater student and teacher buy-in to the assessment system and the flexibility of the model, which provides a wide variety of evidence potentially to be used.

Conley notes some of the challenges to be addressed in implementing this approach, including the need to decide on criteria that would qualify an assessment to be included in a collection of evidence and the issue of rater reliability and describes how these would need to be addressed. The largest challenge in implementing a collection of evidence alternative will come from the costs of the management and training involved. To properly structure a collection of evidence model, appropriate models for the collections, specific criteria, and a set of expectations will need to be crafted by the state. There will also need to be proper monitoring and training completed by the state to ensure quality, validity, and reliability of the scoring. However, most of these challenges

can be mitigated if, as Conley (2005) suggests, there is a thoughtful and gradual implementation over time.

In our view, the potential benefits outweigh the difficulties. A collection of evidence approach in assessment provides a greater number of learning samples, which makes the judgment about achievement more valid and defensible, and it lies closest to the learning process, thus aligning student coursework and development with assessment. This alternative also encourages the evaluation of multi-dimensional tasks and applied skills used in college and work, such as inquiries involving researching, organizing, analyzing, and presenting information. By allowing students to include their coursework that is worked on over time rather than under time constraints, the students have more opportunities to hone their skills and present the best possible product to be assessed, working as they would in outside-of-school contexts. (See Table 12 for a list of potential benefits and difficulties of using a collection of evidence approach.)

**Table 12: Potential Benefits and Difficulties of Using Collection of Evidence**

POTENTIAL BENEFITS	POTENTIAL DIFFICULTIES
<ul style="list-style-type: none"> <li>● Closest to the learning process</li> <li>● Assesses multi-dimensional skills that are important to develop.</li> <li>● Use multiple pieces of evidence to assess student learning</li> </ul>	<ul style="list-style-type: none"> <li>● Costs for management and training</li> <li>● May be seen as less reliable, although specification of tasks or products and training of scorers can address this concern.</li> </ul>

**Recommendations.** The collection of evidence approach is worthwhile and will add significantly to the fairness, validity, and instructional benefits of the assessment system. To achieve these goals, the collection should be embedded in tasks that are part of students’ coursework, and these should be well-specified, connected to the standards and course syllabus, and commonly scored, so that they are supported by teaching and can be reliably evaluated. This alternative should ideally be pursued from the beginning of the students’ high school career and perhaps should be available to all students, not just those who fail, as a way of strengthening teaching and learning.

***Weighted Multiple Measure Approach***

One proposal that has surfaced in Washington is a proposal for a “weighted multiple measures” system of assessments. All of the proposals we have reviewed above would create a multiple measures system. The distinction in a *weighted* multiple measures system is that each assessment component would be associated with a particular value for its weight in rendering a decision. In the proposal that has been suggested, these values would be established at the state level.

**Table 13: Key Definitions**

<p><b>Multiple Measures Approaches to Assessment:</b> The use of numerous and varied assessments to evaluate learning and performance. A multiple measures approach incorporates evaluations of different kinds, not just repeated uses of the same measure (e.g. taking a test on multiple occasions). This is sometimes referred to as using a “combination of measures” to assess student learning.</p>
<p><b>Weighted Multiple Measures System of Assessment (WMM):</b> An assessment system that takes into account numerous and varied assessments to evaluate learning and performance, and attaches specific weights to each measure to designate its influence on a specific decision. A WMM system attempts to quantify the importance of each measure by establishing a formula for combining information to produce a standardized score or determination.</p>

As we have described, professional testing standards support the validity of approaches that use multiple measures when making high stakes decisions, since no test is sufficiently reliable and valid to be the sole basis for important decisions. Many states have developed approaches that incorporate more than one measure for making the high school graduation decision. However, no state in the U.S. has yet developed and maintained a multiple measures approach that attaches specific *weights* at the state level for various components.

A weighted multiple measures assessment system would have all the benefits of the multiple measures approaches we have described above, and would have the added benefit of clarifying how a set of different kinds of evidence would be combined in rendering a single decision. Thus, students would not have to fail the exit exam in order to access alternatives. They would be able to show their competence in several ways and have all of them taken into account in the final decision. This is much more like the examination systems in other countries that we described above.

Despite the supportive rationale behind a multiple measures approach, which is used by a number of states, there is no state that has yet fully developed and implemented a weighted multiple measures (WMM) approach that adopts a single set of state weightings for different measures, including non-test measures. Thus, a WMM approach has a number of unknowns that pose both educational and political challenges.

Most states that use multiple measures approaches to graduation do not have a weighting system attached. Many consider alternative assessments or measures when students have failed the exit exam. Others leave the decisions about how to combine information from different assessments up to local districts (e.g. Rhode Island, Connecticut, and Maine). There is no state that has yet fully developed and implemented a weighted multiple measures approach that adopts a single set of state weightings for different measures, including non-test measures. Delaware experimented with a weighted system based only on test scores (not alternative measures), but decided to discontinue the system because of political challenges. These challenges included differences of opinion about what measures ought to be emphasized in the weighting system, a concern that would certainly arise in any attempt to establish a WMM system.

The idea of a weighted multiple measures approach, while potentially desirable in the abstract, may be a challenge at this moment in time, especially given the paucity of state experience and the lack of research support for this model. Below, we discuss what various states have done to combine evidence from multiple measures to make graduation decisions.

*Rhode Island:* Rhode Island's graduation decision is to be based on a range of performance measures combined in a compensatory model. The commissioner in Rhode Island asked local districts to develop a high school diploma system that takes into account at least 20 Carnegie units, knowledge and skills representing the minimum statewide standards, a school wide diploma assessment that might include an exhibition, graduation portfolio, certificate of initial mastery, etc., as well as results from local and state assessments. Rhode Island is developing a new statewide assessment with New Hampshire and Vermont. The results of the state tests need to be used in some manner, but should not be the sole grounds to prohibit graduation from high school and may represent no more than 10% (and conceivably even less) of all of the weighted factors contributing to promotion or graduation (Rhode Island Department of Secondary and Elementary Education, 2003; Rhode Island Department of Education, 2005).

*Connecticut:* Graduation requirements in Connecticut vary from district to district as well. The state test must play a role in determining graduation, but cannot be the sole basis for denying a diploma. In making graduation decisions, districts are expected to use the state high school exam in combination with local assessments, which must include both performance assessments and coursework performance. For example, one district, West Hartford, requires students to satisfactorily earn 21.75 units of credit in various subject areas and meet four performance standards in reading/writing, mathematics, science, and technology. The performance tasks involve writing an essay about a piece of literature; demonstrating mathematics included in the state standards tests, including multi-step problems, as well as an explanation and justification for their answers; exhibiting investigational skills and competencies in science, and demonstrating technology competency within discipline-based learning tasks (West Hartford Public Schools Board of Education Policy; Connecticut State Board of Education). These requirements for the collection of evidence push instruction in the district to ensure these kinds of tasks are offered and taught to, as well as ensuring that students have mastered such skills.

*Maine:* Maine determines graduation at the local level using the Maine Learning Results in eight content areas (English Language Arts, Mathematics, Science, Social Studies, Health/Physical Education Career Preparation, Modern and Classical Languages and Visual and Performing Arts). The assessment of those content areas varies by district. Many districts use a combination of performance measures including classroom-based portfolios, observations, exhibitions, and district administered exams and tasks. For example, the city of

Bangor Maine uses teacher-created common assessments, course grades, and external assessments like the Maine Educational Assessment (MEA) (The Bangor School Department for the Citizens of Bangor; FairTest Examiner, Spring 2002). Some districts use alternative assessments like the SAT as one of the measures.

*Delaware:* Delaware instituted a three-tiered diploma system in 2000. At the time, students needed to fulfill a set of course requirements, any local requirements, and take Delaware's state high school test. Students received a basic, standard, or distinguished diploma based on a composite index score earned on the state high school tests in reading, writing, and mathematics. Test scores were factored together into a graduation index according to a 40% weight on the students reading score, a 40% weight on the mathematics score, and a 20% weight on the writing score. This weighted system only attached weights to test scores, not to other assessments that were part of the graduation decision. Thus, it was not a weighted multiple measures system. Its fate, however, suggests how problematic weighting systems can be.

The three-tiered diploma system quickly became controversial and caused debate among stakeholders. Delaware's Governor Minner supported the system, but some stakeholders pointed out that students with high GPAs or SAT scores could be denied distinguished diplomas under the three-tiered system. Some constituents thought it was unfair that students could take the test in tenth grade and be assured they'd receive a distinguished diploma from their high scores. Overall, most critics argued that the system gave too much emphasis to the test in making decisions about diploma status, and unfairly categorized students without accurately assessing their high school performance.

After more than a year of debate, the legislature and the governor passed a bill replacing the three-tiered system with a one-diploma system that allows students to use scores on the SAT and certain Advanced Placement end-of-course exams as well as their score on state tests to qualify for a distinguished diploma. By 2008, the current weighting system will be eliminated (Delaware Department of Education; State of Delaware, 2005; The Council of State Governments Eastern Regional Conference, 2005).

The potential benefits and difficulties of a weighted multiple measures system are outlined in Table 14. As noted above, there is experience with locally managed multiple measures systems that is encouraging in states like Rhode Island, Maine, and Connecticut. On the other hand, Delaware's failed attempt to use state-determined weights in a graduation policy suggests the sensitive nature of the approach. While a weighted multiple measures approach is grounded in valid theories about assessment and learning, the practicality of the approach has yet to be established. The most important concern is how to justify the weights in a way that has educational and political acceptability. There would need to be a number of studies conducted surrounding the use of a weighted

multiple measures system. The studies should explore how weights are selected, the impact weights have on the education system, the predictive validity of the weighted scores and the unintended consequences of various weights.

**Table 14: Potential Benefits and Difficulties of a Weighted Multiple Measures System**

POTENTIAL BENEFITS	POTENTIAL DIFFICULTIES
<ul style="list-style-type: none"> <li>● The use of multiple assessments meets professional testing standards</li> <li>● Depending on the measures, the system may include more authentic, formative assessment tasks that improve instruction</li> <li>● Depending on the measures and the weights, may increase graduation rates</li> </ul> <p>[Note these benefits would pertain to any multiple measures system, with or without weights.]</p>	<ul style="list-style-type: none"> <li>● There is no empirically-verified information or process available to select appropriate weights</li> <li>● The impact of the choices of various weights is unknown, thus unforeseen consequences may be caused by various models</li> <li>● Costs of researching and developing a WMM assessment system would be substantial</li> <li>● No state has fully developed or implemented a WMM approach at the state level; the only state that has attempted to attach specific weights to assessments had to terminate its system because of political non-sustainability</li> <li>● In the absence of strong empirical justification, weightings may be politically contentious and weaken the viability of enacting or maintaining a WMM system.</li> </ul>

To this date, there have been no studies of a weighted multiple measures approach, which leaves policy makers without the information needed to design and implement such a system well. Ideally the justification for weights would be based on the strength of different measures in guiding desirable school practices and predicting success in life and college. To gain insights about these issues, empirical studies would be needed to examine the predictive validity of different measures as well as the effects on schools of using different approaches. Studies should examine the impact weights have on students' pathways towards education. It would be important to ensure that the weights encourage high quality standards-based instruction and encourage rather than discourage students from persevering to graduation.

Constituents will for one reason or another want to place the weights on different performance measures higher or lower according to their perspectives. It will be important to know what impact there will be, either in terms of resource management, teacher instruction, or public opinion, caused by weighting one performance measure over another. Varying weights on different performance measures could potentially have unforeseen consequences. For example, too much weight on GPA could lead to grade inflation; too little on an on-demand test might lead to lack of confidence in the system, while too little on performance measures might lead to inadequate incentives for emphasizing ambitious kinds of learning in the curriculum, relative to other things tested. These unintended consequences will need to be thought through and examined in terms of their

impact. For example a weighted approach that places too little weight on a costly performance measure may be seen as not effectively recognizing that measure. To ease this decision-making process, studies outlining an appropriate process for selecting the weights and analyzing their impact would need to be presented to policymakers. Finally, there is yet no model for how to implement a WMM system in practice, once the weights are assigned. This, too, would need to be worked out before such a system could proceed.

**Recommendations.** Using multiple measures can provide a well-rounded assessment of students' learning that may also encourage higher graduation rates, improve pedagogy, and provide a better indication of student potential. As we have described throughout this paper, there are several ways to accomplish this, aside from using a state-determined weighting system. Although promising, a weighted multiple measures approach needs study and exploration prior to its consideration on the policy front and implementation in an assessment system

## Conclusions

Senate Bill 6475 offers three alternative assessment strategies to augment Washington's current exit exam. The findings from this report support all three of these alternatives and suggest some refinements that would expand the validity of the assessment system, encourage improvements in curriculum and instruction, support incentives for students to work hard in school, and inspire rather than discourage students from remaining in school to pursue their high school diploma.

We conclude that system would be strengthened by:

- Allowing students to use substitute tests in all of the fields tested by the WASL and, in addition to the SAT or ACT, considering as potential options assessments that would encourage students to undertake more rigorous coursework and challenging tasks, such as the Advanced Placement (AP), International Baccalaureate (IB), and Cambridge Advanced International Certificate of Education (AICE) assessments, and more diagnostic assessments like the Measures of Academic Progress (MAP) assessments. The use of these substitutes routinely, as in many states, rather than only for those who fail the exam, should be considered.
- Evaluating the design of the current assessments to see how well they meet universal design principles and how well they assess the knowledge and skills of English language learners, and, if necessary, creating a linguistically-modified version of the test that increases its validity for assessing English language learners.

- Considering student GPAs in standards-based coursework as an adjunct to the examination score, as some states do, rather than as an alternative measure only for those who have failed the exam.
- Ensuring that a collection of evidence approach included in the final assessment system uses tasks that are part of students' coursework and are well-specified, connected to the standards and course syllabus, and commonly scored, so that they are supported by teaching and can be reliably evaluated. This alternative should ideally be pursued from the beginning of the students' high school career and perhaps should be available to all students, not just those who fail, as a way of strengthening teaching and learning.

These measures should be made available to students as quickly as possible and ideally, as in some states, to all students so that the graduation decision is based on multiple forms of evidence. Those that are reserved as alternatives only for students who have failed the test should be available after one attempt. Especially in the case of a collection of evidence approach, implementation of the alternative should begin early in students' career both to ensure a high-quality process and to encourage students to work hard in school and to remain in school rather than dropping out.

## References

- Abedi, Jamal & Dietel, John (2004, Winter). Policy Brief #7. Los Angeles: University of California at Los Angeles. Center for the Study of Evaluation, Student Standards, and Testing.
- Abedi, J., Lord C., & Plummer, J. R. (1997). Final report of language background as a variable in NAEP mathematics performance. Los Angeles: Center for the Study of Evaluation. CSE Technical Report # 429.
- Achieve, Inc. (2004) *Do Graduation Tests Measure Up? A closer Look at State High School Exit Exams*. <http://www.achieve.org/files/TestGraduation-FinalReport.pdf>
- Advocates for Children (2002). Pushing out at-risk students: An analysis of high school discharge figures -- a joint report by AFC and the Public Advocate. Retrieved from: <http://www.advocatesforchildren.org/pubs/pushout-11-20-02.html>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). Standards for educational and psychological testing. Washington, D.C.: American Educational Research Association.
- Arizona Department of Education. "Rules" *AIMS Augmentation*. <http://www.ade.state.az.us/stateboard/AIMSAugmentationOverview.pdf>
- Arizona Department of Education. "AIMS Spring 2006, Scale Scores and Performance Levels, Adopted by the School Board May 12, 2006." <http://www.ade.state.az.us/ResearchPolicy/AIMSResults/2006ScaleScoreTable.pdf>
- Asimov, N. (March 29<sup>th</sup>, 2006). "California: Thousands still may not get their diplomas." San Francisco Chronicle. <http://www.sfgate.com/cgi-bin/article.cgi?f=/c/a/2006/03/29/BAGT0HVKDE1.DTL&hw=Exit+exam&sn=009&sc=459>
- The Bangor School Department for the Citizens of Bangor. *The Communique*. "State Law and High School Diplomas: Certifying Achievement of the Maine Learning Results." By Frank Hackett. <http://www.bangorschools.net/userimages/certifyachievement.pdf>
- Bernstein, D. (June 11-17, 2004). "Achievement gap: This is improvement" The Boston Phoenix. Retrieved from: [http://www.bostonphoenix.com/boston/news\\_features/this\\_just\\_in/documents/03902591.asp](http://www.bostonphoenix.com/boston/news_features/this_just_in/documents/03902591.asp)

Capriccioso, R. (March 28<sup>th</sup>, 2006). "Can ACT Improve the Pipeline?" *Inside Higher Ed*. <http://www.insidehighered.com/news/2006/03/28/act>

Center on Education Policy. (2003). "State High School Exit Exams: Put to the test." Washington, D.C.: Center on Education Policy. <http://www.cep-dc.org/highschool/exit/1/exitexam4.pdf>

Center on Education Policy. (2004). "State High School Exit Exams: A Maturing Reform." Washington, D.C.: Center on Education Policy. <http://www.cep-dc.org/highschool/ExitExamAug2004/ExitExam2004.pdf>

Center on Education Policy. (August, 2006). "State High School Exit Exams: A Challenging Year." Washington, D.C.: Center on Education Policy. <http://www.cep-dc.org/pubs/hseeAugust2006/CEPHSEEPresConference8-10.pdf>

Chan, Aldo. (October 2004). "The Class of 2008: Growing up with the WASL." *The Seattle Times*.

Clarke, M., Shore A., Rhoades, K., Abrams, L., Miao, J., Li, J. (January 2003) "Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from Interviews with Educators in Low-, Medium-, and High-Stakes States." National Board on Educational Testing and Public Policy. Lynch School of Education. Boston College. <http://www.bc.edu/research/nbetpp/statements/nbr1.pdf>

Conley, D. (January-August, 2005) Student of Alternative Methods to the Washington Assessment of Students Learning (WASL): Feasibility Study, January-August 2005, Final Report. [http://www.s4s.org/upload/WASL%20Final%20Report\\_093005.pdf](http://www.s4s.org/upload/WASL%20Final%20Report_093005.pdf)

Connecticut State Board of Education, Hartford. (February 6<sup>th</sup>, 2002). From Theodore S. Sergi, "Implementation of New Legislation Regarding Local Graduation Competency Requirements." [http://www.state.ct.us/sde/DTL/curriculum/hs02/sbememo\\_tsscomp.pdf](http://www.state.ct.us/sde/DTL/curriculum/hs02/sbememo_tsscomp.pdf)

The Council of State Governments Eastern Regional Conference. (March 25<sup>th</sup>, 2005). "Regional News: Delaware Debates Alternatives to a Three-tier program." <http://www.csgeast.org/page.asp?id=biweeklyeduc2>

Darling-Hammond, L., & Aness, J. (1994). *Authentic assessment and school development*. New York: National Center for Restructuring Education, Schools, and Teaching, Teachers College, Columbia University.

Darling-Hammond, L., Rustique-Forrester, E., and Pecheone R. L. *Multiple Measures Approaches to High School Graduation*. Stanford: School Redesign Network, 2005. [http://www.schoolredesign.net/srn/mm/pdf/multiple\\_measures.pdf](http://www.schoolredesign.net/srn/mm/pdf/multiple_measures.pdf)

Darling-Hammond, L. & Rustique-Forrester, E. (2005). "The consequences of student testing for teaching and teacher quality." In Ed Haertel and Joan Herman (eds.) *NSSE Year- book*. Chicago: National Society for the Study of Education.

Dee, Thomas S. and Jacob, Brian, "Do High School Exit Exams Influence Educational Attainment or Labor Market Performance?" (April 2006). National Bureau of Economic Research Working Paper No. W12199.

Delaware Department of Education. Delaware Student Testing Program. "Procedures for Earning A State of Delaware Distinguished High School Diploma for the class of 2006 and 2007." <http://www.doe.k12.de.us/AAB/>

Education Week. (January 11<sup>th</sup>, 2001). *Quality Counts 2001: A better balance*. Bethesda, MD: Editorial Projects in Education.

Eslinger, B. (August 23<sup>rd</sup>, 2006). "Exit exam opponents won't give up legal fight." *The San Francisco Examiner*. [http://www.examiner.com/a-236994~Exit\\_exam\\_opponents\\_won\\_t\\_give\\_up\\_legal\\_fight.html](http://www.examiner.com/a-236994~Exit_exam_opponents_won_t_give_up_legal_fight.html)

FairTest Examiner. (Spring, 2002). "Nebraska and Maine Assessment Models." [http://www.fairtest.org/examarts/Spring%2002/Maine\\_and\\_Nebraska.html](http://www.fairtest.org/examarts/Spring%2002/Maine_and_Nebraska.html)

FairTest Examiner. (Winter 2005). "Arizona Exit Exam Under Attack." [http://fairtest.org/examarts/Winter%202004-2005/Arizona\\_Exit.html](http://fairtest.org/examarts/Winter%202004-2005/Arizona_Exit.html)

FairTest Examiner. (Summer 2005). "Arizona Adds Exit Exam Options." [http://www.fairtest.org/examarts/Summer%2005/AZ\\_ExitExam.html](http://www.fairtest.org/examarts/Summer%2005/AZ_ExitExam.html)

FairTest Examiner. (May 2006). "Arizona Grad Test in Place...for Now." [http://www.fairtest.org/examarts/May%202006/AZ\\_Exit.html](http://www.fairtest.org/examarts/May%202006/AZ_Exit.html)

Falk, B., & Ort, S. (1997, April). *Sitting down to score: Teacher learning through assessment*. Presentation at the annual meeting of the American Educational Research Association, Chicago, IL.

Fernald, J. (2002) "Preliminary Findings on the Relationship between SAT Scores, High School GPA, Socioeconomic Status, and UCSC Freshman." Institutional Research and Policy Studies, University of California – Santa Cruz. <http://senate.ucsc.edu/cafa/SATGPA.htm>

Florida Department of Education, (December 2005) "Graduation Requirements."

*Florida Comprehensive Assessment Test.*  
<http://www.firn.edu/doe/sas/fcat/pdf/fcatpass.pdf>

Glenn, D. (July 3<sup>rd</sup>, 2006). "High-School Exit Exams Are Associated With Higher Dropout Rates, Researchers Find." *The Chronicle of Higher Education*.  
<http://chronicle.com/daily/2006/07/2006070301n.htm>

Goldberg, G.L., & Rosewell, B.S. (2000). "From perception to practice: The impact of teachers' scoring experience on the performance based instruction and classroom practice." *Educational Assessment*, 6(4), 257–290.

Haney, W. (2000). The myth of the Texas miracle in education. *Education Policy Analysis Archives*, 8(41). <http://epaa.asu.edu/epaa/v8n41/>

Haney, W. (2002). "Lake Wobegone guaranteed: Misuse of test scores in Massachusetts, Part 1." *Education Policy Analysis Archives*, 10(24). Retrieved from: <http://epaa.asu.edu/epaa/v10n24/>

Heubert, J. & Hauser, R. (eds.) (1999). High stakes: Testing for tracking, promotion, and graduation. A report of the National Research Council. Washington, D.C.: National Academy Press.

Idaho State Board of Education. "Phased-In Graduation Requirements." (March 20<sup>th</sup>, 2004). *Idaho Administrative Code*.  
<http://www.sde.state.id.us/instruct/docs/alternativeschools/Adminrules.pdf>

Idaho State Board of Education. (March 2005) "State High School Exit Exams Accepted in Idaho."  
<http://www.boardofed.idaho.gov/saa/documents/StateHSEExitExams-ID2.pdf>

Idaho State Department of Education. "Idaho Standards Achievement Test (ISAT) data." <http://www.sde.state.id.us/admin/isat/>

Indiana Department of Education, 2006. "What you need to know about Indiana's Graduation Qualifying Exam." <http://www.doe.state.in.us/GQE/welcome.html>

Jacob, B.A. (2001). "Getting tough? The impact of high school graduation exams." *Education and Evaluation and Policy Analysis*, 23 (2), 99-122.

Kato, G. (February 27<sup>th</sup>, 2006). "English Only Five Years Later." [policymatters.net](http://www.policymatters.net/2006/02/), the Goldman school of Public Policy, University of California, Berkeley. <http://www.policymatters.net/2006/02/>

Koretz, D. Stecher, B., Deibert, E. (August 1992). "The Vermont Portfolio Assessment Program: Interim Report on Implementation and Impact, 1991-1992 School Year." The RAND Corporation.

<http://www.cse.ucla.edu/Reports/TECH350.PDF#search=%22Vermont%20Performance%20Assessment%22>

Libreria, William (June 30<sup>th</sup>, 2005). Commissioner, New Jersey Department of Education. Memo to Members of the State Board of Education regarding the Special Review Assessment.

<http://www.state.nj.us/njded/assessment/hs/sra/sramemo.pdf>

Lilliard, D. & DeCicca, P. (2001). "Higher standards, more dropouts? Evidence within and across time." Economics of Education Review, 20(5): 459-73.

Marchant, G. J. & Paulson, S. E. (2005). "The relationship of high school graduation exams to graduation rates and SAT scores." Education Policy Analysis Archives, 13(6). <http://epaa.asu.edu/epaa/v13n6/>

Massachusetts Department of Education. (July 2005). "MCAS Performance Appeals: Commissioner's Message to Educators."

<http://www.doe.mass.edu/mcasappeals/educator.html>

Massachusetts Department of Education. (April, 2006). "Dropouts in Massachusetts public schools: District survey results."

[www.doe.mass.edu/infoservices/reports/dropout/06survey.doc](http://www.doe.mass.edu/infoservices/reports/dropout/06survey.doc)

Massachusetts Department of Education, (May 26<sup>th</sup>, 2006). "Settlement allows MCAS to stand as graduation requirement."

<http://www.doe.mass.edu/news/news.asp?id=2907>

Massachusetts Department of Education, (October, 2005). "Dropouts in Massachusetts Public Schools: 2003-2004."

<http://www.doe.mass.edu/infoservices/reports/dropout/0304/report.pdf>

Massachusetts Department of Education "Enrollment Data." Retrieved from

<http://www.doe.mass.edu/infoservices/reports/enroll/>.

Massachusetts Department of Education, "Plans of High School Graduates."

Retrieved from <http://www.doe.mass.edu/infoservices/reports/hsg/>.

Murnane, R., & Levy, F. (1996). *Teaching the new basic skills*. New York: The Free Press.

New Jersey's Office of Evaluation & Assessment, 2005. "Special review Assessment (SRA) Administration Manual 2005-2006 School Year." New Jersey Department of Education.

New Jersey Department of Education website, 2006. "Your Guide to the HSPA, High School Proficiency Assessment, March 2006."

[http://www.state.nj.us/njded/assessment/hs/hspa\\_guide\\_english.pdf](http://www.state.nj.us/njded/assessment/hs/hspa_guide_english.pdf)

New York State Education Department. "Regents and High School Diploma Graduation Requirements."

<http://www.emsc.nysed.gov/deputy/Documents/alternassess.htm>

Northwest Evaluation Association Website. <http://www.nwea.org/>

Office of Superintendent of Public Instruction. "Washington State Report Card."

<http://reportcard.ospi.k12.wa.us/>

Office of Superintendent of Public Instruction.. (October 2005). "Guidelines for Participation and Testing Accommodations for Special Populations in State Assessment Programs."

<http://www.k12.wa.us/Assessment/TestAdministration/pubdocs/GuidelinesSpecialPopulationsStateTests2005.pdf>

Oregon Department of Education website, 2006a. "Overview: Oregon's Assessment System."

<http://www.ode.state.or.us/teachlearn/subjects/general/asmtovervieworasmtsyste m0506.pdf>

Oregon Department of Education website, 2006b. "Certificate of Initial Mastery, Certificate of Advanced Mastery, and Diploma Requirements."

<http://www.ode.state.or.us/teachlearn/real/documents/CIMCAMDiploma.pdf>

Oregon Department of Education website, 2006c. "2006-2007 High School Graduation Requirements: Preparing each student for successful transitions to his or her next steps."

<http://www.ode.state.or.us/teachlearn/certificates/diploma/diploma-at-a-glance.pdf>

Oregon Department of Education website, 2006d. "Work Samples: Frequently Asked Questions."

[http://www.ode.state.or.us/teachlearn/testing/scoring/revisedworksamplfaq\\_03312005.pdf](http://www.ode.state.or.us/teachlearn/testing/scoring/revisedworksamplfaq_03312005.pdf)

Orfield, G., Losen, D., Wald, J., & Swanson, C.B. (2004). Losing our future: How minority youth are being left behind by the graduation rate crisis. Retrieved from:

<http://www.urban.org/url.cfm?ID=410936>

Pedulla, J.J., Abrams, L.M., Madaus, G.F., Russell, M. K., Ramos, M.A., Miao, J. (March 2003). "Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from a National Survey of Teachers." National Board on Educational Testing and Public Policy. Lynch School of Education, Boston College. <http://www.bc.edu/research/nbetpp/statements/nbr2.pdf>

Qualifications and Curriculum Authority. (2005a). "AS and A Levels: The Official Student Guide to the System."

[http://www.qca.org.uk/downloads/7005\\_student\\_alevel\\_guide\\_2005-6\(1\).pdf](http://www.qca.org.uk/downloads/7005_student_alevel_guide_2005-6(1).pdf)

Qualifications and Curriculum Authority. (2005b). "A review of GCE and GCSE coursework arrangements."

<http://www.qca.org.uk/downloads/qca-05-1845-coursework-report.pdf>

Rhode Island Department of Secondary and Elementary Education. (January 9<sup>th</sup>, 2003) *High School Reform*. "Regulations of the Board of Regents For Elementary and Secondary Education Regarding Public High Schools and Ensuring Literacy for Students Entering High School, Final Version 1.12"

<http://www.ridoe.net/HighSchoolReform/hsregulations.htm>

Rhode Island Department of Education. (October 2005) "Rhode Island's Diploma System: An Overview."

<http://www.ride.ri.gov/HighSchoolReform/DSLAT/pdf/dipsysover.pdf>

Roderick, M., Bryk, A.S., Jacob, B.A., Easton, J.Q. & Allensworth, E. (1999). Ending social promotion: Results from the last two years. Chicago: Consortium on Chicago School Research.

Rumberger, R. & Larson, K. (1998). "Student mobility and the increased risk of high school dropout." American Journal of Education, 107: 1-35.

Rustique-Forrester, E. (in press). "Accountability and the pressures to exclude: A cautionary tale from England." Education Policy Analysis Archives.

Shaw, Linda. (September 1<sup>st</sup>, 2005). "Beginner's Guide to the WASL." *The Seattle Times*.

Southeast Center for Teaching Quality. (2004). High-stakes accountability in California: A view from the teacher's desk. *Teaching Quality RESEARCH MATTERS*, 12, 1-2. <http://www.teachingquality.org/researchmatters/issue12-Aug2004.pdf>

State of Delaware. (January 25<sup>th</sup>, 2005). "2005 State of the State Address of Governor Ruth Ann Minner."

[http://www.state.de.us/governor/speeches/state%20of%20the%20state%202005\\_english%20version.shtml](http://www.state.de.us/governor/speeches/state%20of%20the%20state%202005_english%20version.shtml)

State of Delaware. (May 31<sup>st</sup>, 2006.) "Assessment Task Force Report."

[http://www.doe.k12.de.us/files/pdf/de-hjr4\\_assesstaskforce-ap20060531.pdf](http://www.doe.k12.de.us/files/pdf/de-hjr4_assesstaskforce-ap20060531.pdf)

Thompson, S. and Thurlow, M. (June 2002) "Universally Designed Assessments: Better Tests for Everyone!" *National Center on Educational*

*Outcomes Policy Directions*. Number 14, as cited by the National Joint Committee on Learning Disabilities (January 2004) "State and District-Wide Assessments and Students with Learning Disabilities: A Guide for States and School Districts." This document is available online at [www.ldonline.org/njclcd](http://www.ldonline.org/njclcd)

University of Cambridge International Examinations. "Advanced International Certificate of Education: Overview."  
<http://www.cie.org.uk/CIE/WebSite/qualificationsandawardshub/ciespecialisms/aice/aice.jsp>

Vermont Department of Education. State Board of Education: Manuals of Rules and Practices. "2120.8.7, Graduation Requirements; Instructional Practices; Curriculum Leadership, Content and Coordination."  
<http://www.state.vt.us/educ/new/pdfdoc/board/rules/2000.pdf>

Walker, B., Ashcroft, J., Carver, L.D., Davis, P., Torres, G., Lavergne, G.M., (January 25<sup>th</sup>, 2002). "A Review of the Use of Standardized Test Scores in the Undergraduate Admissions Process at the University of Texas at Austin." University of Texas, Austin.  
<http://www.utexas.edu/student/admissions/research/TxTestingTaskForce2001.pdf>

West Hartford Public Schools Board of Education Policy. "Graduation Requirements." Adopted June 17<sup>th</sup>, 2003.  
[http://www.state.ct.us/sde/DTL/curriculum/hs02/hswh\\_polreg.pdf](http://www.state.ct.us/sde/DTL/curriculum/hs02/hswh_polreg.pdf)

Wheelock, A. (2003). School awards programs and accountability in Massachusetts: Misusing MCAS scores to assess school quality. FairTest. Retrieved from  
<http://www.fairtest.org/arn/Alert%20June02/Alert%20Full%20Report.html>

Winegarner, B. (March 27<sup>th</sup>, 2006). "S.F. firm's injunction challenges exam." *The San Francisco Examiner*. [http://www.examiner.com/Search-a59141~S F firm s injunction challenges exam.html](http://www.examiner.com/Search-a59141~S_F_firm_s_injunction_challenges_exam.html)

Wise, L. L. Becker, D.D., Harris, C.D., Taylor, L.R., Johnstone, C.J., Miller, N., A., Thompson, S.J., Sun, S., Shen, X., Wang, X., Koger, L. E., Moody, R. (February 1<sup>st</sup>, 2006). *Independent Evaluation of the California High School Exit Examination (CAHSEE): Third Biennial Report*. Alexandria: Human Resources Research Organization, p. 156.  
<http://www.cde.ca.gov/ta/tg/hs/documents/thirdrpt.pdf>