SC EDUCATION OVERSIGHT COMMITTEE

Reporting facts. Measuring change. Promoting progress.

EDUCATION OVERSIGHT COMMITTEE

AGENDA

Monday, April 11, 2016
1:00 p.m.
<u>433 Blatt Building</u>

I.	Welcome and Introductions	Mr. Robinson
II.	Introduction of New EOC Member, April Allen	Mr. Robinson
III.	Approval of the February 8, 2016 Minutes	Mr. Robinson
IV.	Evaluations of TransformSC Initiatives: Montessori Education & New Tech Dr. Brooke Culclasure Research Director Center for Education Policy and Leadership Riley Institute at Furman University	
V.	Subcommittee Reports	
A.	Academic Standards and Assessment Discussion: Proposal to Delay Single Accountability System	Dr. Merck
B.	EIA and Improvement Mechanisms Information: Update on FY2016-17 Budget Information: Report on Military-Connected Students Information: Report on Parent Survey Results, 2015	Dr. Couch
C.	Public Awareness Information: Results of Focus Group Meetings	Ms. Hairfield

Adjournment

Neil C. Robinson, Jr.
CHAIR

Daniel B. Merck
VICE CHAIR

April Allen

Anne H. Bull

Bob Couch

Mike Fair

Raye Felder

Barbara B. Hairfield

Nikki Haley

R. Wesley Hayes, Jr.

Dwight A. Loftis

John W. Matthews, Jr.

Joseph H. Neal

Molly Spearman

Patti J. Tate

Ellen Weaver

Melanie D. Barton
EXECUTIVE DIRECTOR

SOUTH CAROLINA EDUCATION OVERSIGHT COMMITTEE

Minutes of the Meeting

February 8, 2016

Members in Attendance: Mr. Neil Robinson (Chair); Ms. Anne Bull; Dr. Bob Couch; Sen. Mike Fair; Rep. Raye Felder; Ms. Barbara Hairfield; Rep. Dwight Loftis; Sen. John Matthews; Rep. Joe Neal; State Superintendent of Education Molly Spearman; Ms. Patti Tate; and Ms.Ellen Weaver

EOC Staff in Attendance: Dr. Kevin Andrews; Mrs. Melanie Barton; Ms. Paulette Geiger; Dr. Rainey Knight; Ms. Bunnie Ward; and Ms. Dana Yow

Mr. Robinson called the meeting to order. He welcomed the members and audience in attendance. He then introduced the newest member of the EOC, Ellen Weaver, who is filling the unexpired term of Deb Marks, the business appointee of the Chairman of the Senate Education Committee, Sen. John Courson.

The minutes of the December 14, 2015 meeting of the EOC were approved as distributed.

Key Constituency:

The chairman recognized the superintendent of the Greenville County School District, Dr. Burke Royster. Dr. Royster opened by reiterating that the quality of the teacher in the classroom and the quality of the principal at the school are the single most important factors in improving the educational achievement of children. He then described how the Greenville County School District has strategically and systematically incorporated the Profile of the South Carolina Graduate throughout its schools in an initiative known as G+. The goal of G+ is that every student graduating from a Greenville County public high school will graduate with a diploma plus an industry certificate or with college credit. The initiative, G+ affects every child in the district. For example, Greenville has expanded its prekindergarten classes by increasing the average class from 20 to 23 students, resulting in a 15 percent increase in the number of children served. The district has also worked with private childcare providers to coordinate and provide quality professional development. In elementary schools the district has set a goal that by the end of 2nd grade all students will be reading on grade level. To achieve the goal, the district has set a balanced literacy approach. At the middle grades level, the district has implemented several initiatives: (1) implementing Pro Team to expose students to teaching as a career; (2) increasing the rigor of instruction in the middle grades; (3) collaborating with Junior Achievement to assist students living in poverty in the districts with social programs; and (4) implementing an electronic early warning system to identify students with attendance, behavior and course grades that would signal that the student is at risk of dropping out. Beginning in eighth grade, the district uses Naviance to track 25,000 students with their career interests, transcripts and individualized graduation plans to ensure that the students are on a track toward college and career readiness. Finally at the high school level, the district is implementing opportunities for accelerated learning and expanding manufacturing career opportunities to students.

Rep. Loftis asked about the success of the Naviance system in its first year of operation and the impact of these initiatives at Carolina High School. Sen. Fair recognized Pam Mills with the Greenville County school system. And, Sen. Matthews asked for clarification on the challenges that Greenville has to ensure students can earn national industry credentials.

Subcommittee Reports:

Joint Report from Academic Standards and Assessment and Public Awareness Subcommittees - The chair then recognized Barbara Hairfield to give the report from the joint meeting of the Academic Standards and Assessment and Public Awareness Subcommittees. Ms. Hairfield directed the members of the EOC to the staff's working document and to other information that describes the components of the new federal law, Every Student Succeeds Act that was approved by Congress and signed into law by President Obama in December of 2015. She asked Mrs. Barton to highlight the key issues in ESSA that impact the work of the EOC. Then Ms. Hairfield asked the members to review the parent support documents including the new *Family Friendly Standards in English Language Arts* and *Family Guide to Assessments: Student Success Tools*, a flyer for parents describing the various types of assessments. She asked members to contact Dana Yow of the EOC staff if they had any suggested changes.

EIA and Improvement Mechanisms Subcommittee – Dr. Couch provided a summary of the EOC's budget and proviso recommendations that were endorsed by Governor Haley in her Executive Budget recommendations. Sen. Matthews noted that one of the issues affecting his districts is the need for more funds for equipment that students need in order to earn national industry certifications. Rep. Loftis concurred stating that more collaboration with technical colleges is needed.

Special Reports:

Full-Day 4K Report – Bunnie Ward and Dr. Bill Brown of the University of South Carolina summarized the findings of the full-day 4K report. The report, which is Part I of the CDEP evaluation for 2015-16, reports on the quality, impact and access of the program in 2014-15 and through the first six months of 2015-16. When the EOC receives assessment data from the fall 2015 administrations of early literacy assessments, the EOC will release Part II of the evaluation. Of concern were the initial CIRCLE assessment results from school year 2014-15. Students enrolled in CDEP in nonpublic settings (private or faith-based centers) scored higher on CIRCLE in the fall of 2014 administration than did students enrolled in public school settings. However, in comparing the 5K assessment results, there was no difference in the early literacy skills of students who attended public or private childcare care centers. There is a need to better measure the quality of instruction in both public and private childcare centers. Finally, concerning growth, about 51 percent of children eligible for the free or reduced price lunch program and/or Medicaid are being served in a federally funded or state-funded full-day 4K program. Income-eligible students living in 79 percent of the school districts in the state are now eligible to participate.

Both Rep. Neal and Sen. Matthews concurred that the program must measure and improve the teacher-child interaction in 4K classrooms.

SC Community Block Grants for Education Pilot Program 2015-16 – Ms. Ward updated the EOC on the implementation of the block grants. Awards will be made by mid-February.

Accountability Report – The annual accountability report that the EOC is required by law to provide to the General Assembly was provided to the EOC for its input The report must be submitted by March 1. Ms. Yow asked that members provide any updates to her by February 28.

There being no further business, the meeting was adjourned.

SC EDUCATION OVERSIGHT COMMITTEE

Reporting facts. Measuring change. Promoting progress.

MEMORANDUM

TO: Members, Education Oversight Committee

FROM: Neil Robinson, Chair
Melanie Barton

DATE: March 28, 2016

IN RE: Proposal to Delay Single Accountability System

Attached are the following: (1) a memorandum from State Superintendent of Education Molly Spearman; and (2) a letter from David Mathis, President of the Superintendents' Division of the South Carolina Association of School Administrators. Both documents are requesting that the law, Section 59-18-325(C) be amended to delay for one year the merging of the federal and state accountability systems.

When Act 200 of 2014 was adopted, the General Assembly directed the Education Oversight Committee (EOC) to suspend the state accountability system for school years 2014-15 and 2015-16 while the new assessments were being implemented. In addition, the law required the EOC to make by the Fall of 2016 recommendations to the General Assembly on a new joint accountability system, which merges the state and federal systems.

Two issues have impacted this timeline:

1. The change in assessments for grades 3-8 in English language arts and mathematics. In 2014-15 ACT Aspire was administered. In 2015-16 SC Ready will be administered in the spring. Per the information provided herein, the EOC will not be able to approve SC Ready until after its second administration which will occur in school year 2016-17. Unlike PACT or PASS, SC Ready must be vertically aligned per state law. Determining whether the assessment is vertically aligned will not occur until after the second administration. The EOC has never approved criteria for state school and district rankings without first approving the assessment.

Neil C. Robinson, Jr.
CHAIR

Daniel B. Merck
VICE CHAIR

April Allen

Anne H. Bull

Bob Couch

Mike Fair

Raye Felder

Barbara B. Hairfield

Nikki Haley

R. Wesley Hayes, Jr.

Dwight A. Loftis

John W. Matthews, Jr.

Joseph H. Neal

Molly Spearman

Patti J. Tate

Ellen Weaver

Melanie D. Barton
EXECUTIVE DIRECTOR

2. Congress enacted the Every Student Succeeds Act (ESSA) in December of 2015. ESSA gives states until school year 2017-18 to design and implement a new accountability system. In the meantime the staffs at the South Carolina Department of Education and the EOC are working with stakeholders to design a joint system. However, key guidance from the federal United States Department of Education is pending and likely will not be released until later this year.

We met with the State Superintendent of Education Molly Spearman on March 16, 2016. Given the above issues, it would be our suggestion to the full EOC that, if the General Assembly chooses to suspend statewide ratings for a third consecutive year, that the EOC will recommend to the General Assembly the following. The EOC would determine the format of a transitional report card for the fall of 2016 and 2017 that identifies underperforming schools and districts. At a minimum, the public would have information on student assessment data in reading and mathematics in grades 3 through 8; high school graduation rates; and measures of student college and career readiness at the school, district, and statewide level. Attached is draft language to be considered.

Attachments

Amendment to: SECTION 59-18-325(C)

(Updated through March 28, 2016)

(6) Within thirty days after providing student performance data to the school districts as required by law, the department must provide to the Education Oversight Committee student performance results on assessments authorized in this subsection and end-of-course assessments in a format agreed upon by the department and the Oversight Committee. The Education Oversight Committee must use the results of these assessments in school years 2014-2015, ~~and~~ 2015-2016, and 2016-2017 to report on student academic performance in each school and district pursuant to Section 59-18-900. The committee may not determine state ratings for schools or districts, pursuant to Section 59-18-900, using the results of the assessments required by this subsection until after the conclusion of the 2015-2016 school year; provided, however, state ratings must be determined by the results of these assessments beginning in the ~~2016-2017~~ 2017-2018 school year. The Oversight Committee also must develop and recommend a single accountability system that meets federal and state accountability requirements by the Fall of ~~2016~~ 2017. While developing the single accountability system that will be implemented in 2017-2018, the Education Oversight Committee will determine the format of a transitional report card released to the public in the fall of 2016 and 2017 that will also identify underperforming schools and districts. These transitional reports will, at a minimum, include the following: (1) school, district and statewide student assessment results in reading and mathematics in grades 3 through 8; (2) high school and district graduation rates; and (3) measures of student college and career readiness at the school, district, and statewide level. These transitional reports will inform schools and districts, the public, and the Department of Education of school and district general academic performance and assist in identifying potentially underperforming schools and districts and in targeting technical assistance support and interventions in the interim before ratings are issued.

Note: In Red are the amendments proposed by SCDE. Language underlined is draft language to be presented to EOC on April 11, 2016.

STATE OF SOUTH CAROLINA
DEPARTMENT OF EDUCATION

MOLLY M. SPEARMAN
STATE SUPERINTENDENT OF EDUCATION

MEMORANDUM

TO: Mr. Neil Robinson, Jr.
Education Oversight Committee, Chair
Nexsen Pruett, LLC
205 King Street
Suite 400
Charleston, South Carolina 29401

Mrs. Melanie Barton
Education Oversight Committee, Executive Director
1205 Pendleton Street
Brown Building, Room 502
Columbia, SC 29211

FROM: Molly Spearman

DATE: March 15, 2016

RE: Accountability Model Extension Request

Dear Mr. Robinson and Mrs. Barton:

Beginning in August 2015, the State Department of Education and the Education Oversight Committee initiated a series of activities in conjunction with other states and South Carolina stakeholders to meet the requirements of the Education Accountability Act [1] for a single accountability system that meets federal and state requirements by Fall of 2016.

The reauthorization of the *Elementary and Secondary Education Act* (December 2015) by the *Every Student Succeeds Act* has given South Carolina an unprecedented opportunity to align federal and state priorities into one accountability model; however, it has also initiated several new data element requirements on the report card, new assessments (ie English language proficiency measures), and new school quality and/or student success measures that have not been previously validated in accountability model.

Several aspects of the new federal law are easily interpreted in the language of *ESSA*. Nonetheless, there are numerous sections of the law that are vague and subject to interpretation by the US Department of Education. Regulatory guidance on *ESSA* is forthcoming from the USED, but those guidance documents are not expected until August 2016 or later.

South Carolina, through the State Department and superintendent, must ensure that a comprehensive accountability plan is sent to the USED that is in total compliance with federal law. In order to ensure that the state and federal plans align, we respectfully seek the Education Oversight Committee's support in our request to delay the development of South Carolina's "single accountability system" from the state timeline of 2016-17 to the federal timeline of 2017-18.

For the reasons enumerated below, we believe that this request is not only prudent but also absolutely necessary to create an accountability system in South Carolina that is fair, effective, and aligned with the intent of the Education Accountability Act to have one system that meets federal and state regulations. The specific benefits of moving to the federal 2017-18 timeline as they relate to accountability are as follows:

1. SC Ready assessments will be implemented for the first time in grades 3-8 in ELA and mathematics in the spring of 2016. *ESSA* requires states to set long-term goals and annual measureable targets in the accountability model on all statewide assessments. It is difficult to set appropriate long-term goals and annual targets on a new assessment with no baseline data. The standard setting process that must occur after new assessments are given will occur in the summer 2016. From the standard-setting process, performance levels will be determined and baseline data on the ELA and mathematics grades 3-8 will be established. The EOC will not receive these data until mid to late summer 2016, which does not give either SCDE or the EOC time to review and approve the standard-setting process and performance levels and set reasonable long-term goals and objectives for an accountability plan that would effectively begin with the new school year in August 2016.

2. A second issue related to the new SC Ready assessments is that the Education Accountability Act (§59-18-325) requires the new statewide ELA and mathematics assessments to have a vertical scale to measure growth. At the recommendation of the SCDE's Technical Assessment Committee (TAC), the development of the vertical scale will take two implementations of the SC Ready assessment (2015-16 and 2016-17) so that enough items can be field tested with SC students to design a valid vertical scale. According to state law, the EOC must approve the statewide assessments used in accountability. SC Ready cannot be submitted to the EOC for approval until summer 2017. Moving to the federal 2017-18 timeline for accountability supports the intent of the Education Accountability Act for the EOC to give final approval of all tests prior to their being used for accountability.

[1] §59-18-325(6) "The Oversight Committee also must develop and recommend a single accountability system that meets federal and state accountability requirements by the Fall of 2016."

3. In the SCDE's and EOC's joint work on the new accountability model, rich discussions have occurred around comprehensive definitions for college and career readiness. These definitions will play into the accountability model for a significant percentage and will hinge upon a.) re-procurement of the "college readiness test" in the spring/summer 2016; b.) opportunities for re-assessment on the "two year college entrance test" that will be procured by the state technical schools; c.) pilots of the college remediation courses offered in the senior year; and d.) funding and training for high schools and career centers to provide more opportunities for students to earn industry certifications that lead to a living wage. Moving to the federal 2017-18 timeline for accountability will allow South Carolina to procure and implement the assessments, pilot remediation courses, and increase student accessibility to industry certifications to evaluate their impact on setting long term goals in the accountability model.

4. *ESSA* requires several new data elements on the report card disaggregated by all subgroups for every school and district that are not currently collected by the SCDE. Districts and schools have previously provided these data elements directly to the Office of Civil Rights in a report that is required every two years, but they have never reported on them annually to the SCDE. Therefore, the SCDE needs to design the methodologies in PowerSchool to collect these data elements from schools, train school personnel to input these data, and set up a way for schools to validate the data before it is reported annually on a state report card. The training will be extensive, particularly in small, rural districts that do not have district accountability coordinators or district test coordinators to support its schools. Moving to the 2017-18 timeline will allow the State Department to ensure ALL of the state's schools are ready to implement the new report card elements.

5. The USED must approve each state's *ESSA* accountability model. States are required to develop an accountability model that complies with all elements of the law, put the model out for public comment for a minimum period of time, submit the model to the USED for review (feedback from the USED must be provided back to the State within 120 days of the submission), and make changes as required to the model from the feedback. The SCDE cannot meet these federal requirements with the accountability plan by the Fall 2016 timeline. Moving to the federal 2017-18 timeline will allow the SCDE to meet these federal requirements prior to the EOC making a recommendation to the General Assembly for the "single accountability system" and will ensure that schools have enough time to get training and see simulations on the new accountability model prior to implementing it at the beginning of the 2017-18 school year.

In conclusion, development of a robust accountability model is vitally important to all stakeholders in ~~the~~ South Carolina. The SCDE respectfully requests the time that other states are taking to ensure our State's model aggressively leverages all of the elements in *ESSA* and the *Profile of the South Carolina Graduate* to move our state forward. Upon your request, the SCDE is available to address the EOC in person to provide additional information or answer any questions you have.

Below is the section of the Education Accountability Act that is impacted by the change in timeline request with possible amendments for your consideration:

§59.18.325 (6): The Education Oversight Committee must use the results of these assessments in school years 2014-2015, ~~and~~ 2015-2016, and 2016-17 to report on student academic performance in each school and district pursuant to Section §59-18-900. The committee may not determine state ratings for schools or districts, pursuant to §59-18-900, using the results of the assessments required by this subsection until after the conclusion of the ~~2015-2016~~ 2016-17 school year; provided, however, state ratings must be determined by the results of these assessments beginning in the ~~2016-2017~~ 2017-18 school year. The Oversight Committee also must develop and recommend a single accountability system that meets federal and state accountability requirements by the Fall of ~~2016~~ 2017.

SCASA
Leading South Carolina Schools

SCASA DIVISIONS

Superintendents

Instructional Leaders

Personnel Directors

Education Specialists

Principals
Secondary
Middle
Elementary

Adult Education Directors

Career and Technology Directors

Higher Education

March 2, 2016

The Honorable Molly M. Spearman
State Superintendent of Education
SC Department of Education
Rutledge Building
1429 Senate Street
Columbia, SC 29201

Dear Mrs. Spearman:

The superintendent's division of the South Carolina Association of School Administrators (SCASA) took a unanimous vote on February 4, 2016 in support of extending the timeline of the state accountability system to coincide with the newly passed Every Student Succeeds Act (ESSA). Our new state accountability system will have a long-lasting impact on education in South Carolina and additional time is needed to ensure the state's system is designed to provide continuous growth and improvement. We have the opportunity to design a 21st century accountability system that will serve our students well and provide the continuous improvement that we all seek while ensuring school effectiveness.

Our considerations for this extension include:

The new federal accountability was passed in December 2015 and regulations for full implementation have not been finalized. It is projected that USED regulations will not be available until summer. It would be a mistake to make decisions about the state model before federal regulations have been finalized and communicated to states. Regulatory guidance on ESSA could impact approval of South Carolina's accountability plan. Additional time is needed to effectively merge our state system with ESSA.

A major component of ESSA is to measure growth in achievement of each student. The South Carolina assessments are in flux. Assessments have been changed. The 3-8 grade state assessment students will take in spring 2016 will be the third of three different assessments and new assessment selection has been delayed with procurement of the college entrance assessment in 2016-17. Additional time is needed to secure assessments, develop understanding of the new assessments by teachers and most importantly secure the data needed to accurately set growth indexes and establish effective data systems required for accountability.

Quality implementation is essential to the effectiveness of our new accountability system. Time is needed to develop implementation material, data systems, training for district staff and other implementation tasks once approval is gained by both USED and the state legislature.

South Carolina Association of School Administrators
121 Westpark Boulevard ■ Columbia, SC 29210
(803) 798-8380 ■ FAX (803) 731-8429
www.scasa.org

South Carolina would be joining states across the nation designing a new state system and aligning the state system within the ESSA timeline which is appropriate considering the task of developing a state model.

Superintendents strongly feel that an extension of this timeline is needed to adequately craft an accountability model that reports school and district progress, builds capacity for continuous improvement, provides intervention procedures for underperforming schools and merges with federal accountability. We are confident that South Carolina's system of state accountability can be a national model.

Please contact me if I can provide further explanation or clarification.

Sincerely,

David M. Mathis, Ed.D.
President
Superintendents' Division
South Carolina Association of School Administrators

SC EDUCATION OVERSIGHT COMMITTEE

Reporting facts. Measuring change. Promoting progress.

MEMORANDUM

TO: Members, Education Oversight Committee

FROM: Bob Couch, Chair
Melanie Barton
Bunnie Ward

DATE: March 23, 2016

IN RE: H.5001, Fiscal Year 2016-17 General Appropriation Bill

On March 23, 2016 the House of Representatives gave third reading to H.5001, 2016-17 General Appropriation Bill. The following is a summary of budget and proviso recommendations for Fiscal Year 2016-17. Additional information on provisos and the Education Improvement Act (EIA) budget are attached.

Attachments

Summary of H.5001, as Adopted by the House
Regarding Public Education
Increases of General Fund, Lottery and EIA Revenues

Recurring General Funds

Education Finance Act – Increased base student cost from $2,220 to $2,350 and added a weight for dual enrollment of 0.2. The EFA appropriation was increased by $217.6 million.

Virtual SC – Increase of $1.2 million for 18 additional teachers for online courses.

Operations at South Carolina Department of Education – Increase of $2.8 million for IT, technology infrastructure, staff, educator certification, alternative certification, etc.

EEDA - $10.0 million increase for guidance counselors, etc.

Transportation

Bus Driver Salary $19.2 million
Hazardous Transportation $3.5 million

Other

Education Outreach/State Museum (SCDE) $275,000
Coding Curriculum (SCDE) $300,000

Non-Recurring General Funds:

Technology Technical Assistance (SCDE) $13,977,209 - Funds are to increase the capacity of districts in the Abbeville law suit. Funds must be used by Department to assist school districts in procuring appropriate technology to include devices and infrastructure in accordance with the recommendations made by the technology review team to begin to build capacity to offer online testing and increased access. (*Another $2.8 million in Capital Reserve Fund monies, H.5002)*

Districts with a poverty index of 80% or greater - $9.1 million for teacher recruitment and retention.

Statewide Facilities Assessment (SCDE) $1.5 million

Lottery Revenues - Allocated through SCDE

- K-12 Technology Initiative - $29.3 million, which is the same level of funding as in the prior two fiscal years. The allocation formula has not changed as well.
- Instructional Materials - $18.0 million
- School Bus Lease/Purchase - $6.5 million for purchase or lease of new buses (*Another $7.2 million in Capital Reserve Fund monies, H.5002*)
- College & Career Readiness of $3.0 million – "Funds are to be used first to increase the capacity of districts that are or were the original trial and plaintiff school districts in the Abbeville law suit. Funds shall be used by the department to provide assistance to districts using appropriately experienced educators with demonstrated effectiveness in instructional leadership. Support shall include professional development, standards and learning support, instructional support, data analysis and leadership development resources to ensure that educators are equipped with the tools to provide students with high quality, personalized learning that supports the Profile of the South Carolina Graduate."
- Efficiency Studies - $3.1 million
- Dynamic Report Card System (EAA) - $1,695,000 (Plus an additional $1,952,000 in non-recurring EIA revenues for PowerSchool)
- Reading Partners $400,000
- Mobile Device Access & Management $3.0 million– Funds will be used to procure high-speed mobile internet service for students that do not have such internet service at home and are participating in a course of study that requires such access.

EIA – Increase of $54.1 million in recurring and $12.1 million in non-recurring EIA revenues, including $3.0 million for industry certification exams, were allocated to:

- Teacher Pay - $23.2 million for 2% teacher salary increase (1% paid for out of EIA and 1% out of EFA) and for increasing the statewide minimum teacher salary schedule from 22 to 23 years.
- SC Public Charter School District $12.1 million
- Rural Teacher Initiative $8.2 million (Governor's Initiative)
- AdvancedEd $2,501,301
- Teacher Supply $750,000
- Read to Succeed Office (SCDE) $257,400
- Office of First Steps:
    - BabyNet $2.0 million
    - Family Connection $300,000
    - New Positions (4.0 FTEs) $365,068
    - First Steps County Partnerships $1.7 million

Working Draft
4/4/2016 10:39 AM

H. 5001 General Appropriation
FY2016-17 Education Proviso Summary

Proviso	Proviso Type	Title	Description (As Adopted by House)	Passed by House 03.24.16
1A.68	EIA	4K Early Literacy Competencies Assessments	AMENDED - To delete the directive that SCDE select up to three formative assessments and instead directs SCDE to manage the administration of the assessments as done in the prior fiscal year. Directs that accommodation that do not invalidate assessment results must be provided in the manner set for by the student's IEP or 504. Deletes the requirement that districts and private providers be allocated $15 per child to cover the cost of the assessment and for them to report electronically to SCDE the results of each individual assessment. SCDE is to use the same assessments until the KEA work on numeracy in NC is completed.	
1A.22	EIA	4K Targeting	EIA funds allocated for 4K shall be utilized for service provision to age eligible children qualifying for free/reduced lunch or Medicaid. If more students seek to enroll than available space, districts shall prioritize students on the basis of family income expressed as a percentage of federal poverty guidelines with the lowest family incomes given the highest enrollment priority. Children with developmental delays who do not already qualify for special needs services should be considered for enrollment.	
1A.21	EIA	Accountability Program Implementation	EOC may carry forward unexpended EAA funds authorized specifically for the administration of the EOC. For current fiscal year, EOC may carry forward prior year Community Block Grants funding not awarded by the grants committee. Funds must be used for an independent common evaluation of each awarded grant to ensure high quality program that maximize a return on the state's investment.	
1A.27	EIA	Adult Education	AMENDED - To allow up to $300,000 to be used to establish an initiative for qualifying adult ed students to qualify for a free high school equivalency test and direct SCDE to establish guidelines for this initiative.	
1A.2	EIA	African-American History	AMENDED - To direct that funds, which are currently a salary line item, be reallocated for development of instructional materials and programs and implementation of professional learning opportunities that promote African American history and culture. Amend the direct that SCDE expend 70% through a competitive grants process.	
1A.8	EIA	Arts in Education	Funds appropriated for Arts Curricula shall be used to support innovative practices. Funds to be distributed through a competitive grants program with up to 33% of grant fund available as "Aid to Other Agencies".	
1A.26	EIA	Assessments G&T, AP, IB	AMENDED - To increase from $4.6 million to $5.4 million allocation of funds for AP and IB testing due to increased participation.	
1A.41.	EIA	Assessments Projection	DELETED - To require SCDE to institute a plan for reviewing SAT, ACT, WorkKeys and other high school exams to determine student strengths and weaknesses and to provide guidance to school districts	

H. 5001 General Appropriation
FY2016-17 Education Proviso Summary

Proviso	Proviso Type	Title	Description (As Adopted by House)	Passed by House 03.24.16
1A.60	EIA	BabyNet Autism Therapy	AMENDED - To increase increasing BabyNet provider rate to $13.58/hour or current Medicaid rate, whichever is higher.	
1.3	General Fund	Base Student Cost	AMENDED - To Increase base student cost from $2,220 to $2,350 and add .15 pupil weighting for dual credit enrollment. Direct students in poverty be identified using USDA community eligibility criteria and include students eligible for Medicaid, children classified as transient and/or homeless. (Increase in EFA of $217.6 million)	
1A.74	EIA (New)	CATE Equipment	ADDED - Funds shall be distributed to districts and career centers based on prior year actual student enrollment in CATE courses with no district or center receiving less than $50,000. Funds may be expended to purchase equipment, to up fit facilities and to purchase consumables. District must include info on current CATE equipment, alignment of equipment to job needs in District plan. Plans must include charter schools in district offering at least one CATE completer program. Districts and centers may carry forward funds to be used for same intended purposes.	
1.79 and 1A.71	General Fund	CDEPP Unexpended Funds	AMENDED - To permit First Steps to retain the first $2 million in unexpended prior year CDEP funds. If by August 15, school districts eligible to participate in full-day 4K opt not to participate, SCDE may utilize available funds to increase participation on a per pupil basis for districts eligible to participate and who have a documented waiting list. Delete requirements that First Steps allocated $4.25 million for CDEP to SCDE and $2 million for Community Block Grants to EOC.	
1A.20	EIA	Certified Staff Technology Proficiency	Amend proviso to delete preferred method for demonstrating technology proficiency and instead direct that district adopted technology proficiency standards and plan should be at a minimum aligned to ISTE teacher standards.	
1A.28	EIA	Clemson Agriculture Education Teachers	AMENDED - To authorize funds to be used for a Regional Coordinator if there are sufficient funds. (An additional $100,000 allocated to Clemson Agricultural Education for this initiative)	
1A.72	EIA (New)	College and Career Readiness	ADDED - Funds appropriated to SCDE for college and career readiness assistance must first be used to improve educational outcomes of students residing in Abbeville districts. SCDE shall use funds to provide districts assistance with using educators with effectiveness in instructional leadership. Equip educators with tools to provide students with high quality, personalized learning that supports Profile of SC Graduate.	

Working Draft
4/4/2016 10:39 AM

H. 5001 General Appropriation
FY2016-17 Education Proviso Summary

Proviso	Proviso Type	Title	Description (As Adopted by House)	Passed by House 03.24.16
1.89	General Fund (New)	College Readiness Assessment	ADDED-Directs SCDE to procure a college readiness assessment that must be administered to all students in Grade 11. A testing accommodation made for a student must be selected from vendor-approved accommodations list and includes in the student's IEP or an accommodations plan. SCDE will collect district data about percentage of students taking the assessment.	
1A.75	EIA (New)	Digital Learning	ADDED - $1.9 million of surplus EIA funds allocated to EOC for Partnerships for Innovation will be utilized for one-year agreements with public and private entities to pilot computer science initiatives in schools. Initiatives must focus on improving the digital literacy skills of students and teachers, expanding opportunities for students to learn coding, or providing computer science curriculum. $1.5 million must be authorized for schools or districts with 80% or greater poverty indices or are an Abbeville district. The EOC will pilot the program.	
1.86	General Fund	Educational Credits Exceptional Needs Children	AMENDED - To require eligible schools to have graduation certificate requirements for special needs children and "provides a specially designed program or learning resource center to provide needed accommodations based on the needs of exceptional needs students or provides onsite educational services or supports to meet the needs of exceptional needs students, or is a school specifically existing to meet the needs of only exceptional needs students with documented disabilities." Allows participating schools to notify EOC of continuing participation by December 30. Any school that did not participate in the program in the prior fiscal year may apply for participation in the program at any time in the fiscal year. EOC must publish the approved list on its website by September 1 and shall update the list upon approval of additional eligible schools.	
1A.77	EIA (New)	Family Connection (EIA)	ADDED - Funds appropriated ($300,000) to be transferred in quarterly installments from SCDE to Family Connection. Funds shall be used to provide support to families of children with disabilities and shall include home visits, transition assistance, education assistance, parent support and parent training. Family Connection to provide planning documents to SCDE no later than July 15 and quarterly reporting thereafter.	
1.74	General Fund	First Steps 4K Technology	AMENDED - To delete reference to "early literacy and language development" and instead reference "required school readiness" assessments. Broadens the definition of assessment to consistent with upcoming planned changes. Still provides $75,000 in 4K carry forward to fund electronic devices.	
1.62 and 1A.30	EIA & General Fund	Full Day 4K	AMENDED - To Increase per student reimbursement from $4,218 to $4,323, to increase transportation reimbursement for private providers from $550 to $564.	

H. 5001 General Appropriation
FY2016-17 Education Proviso Summary

Proviso	Proviso Type	Title	Description (As Adopted by House)	Passed by House 03.24.16
1.53 and 1.83	General Fund	Health Education	AMENDED - To require each district to publish on its website the title and publisher of all health ed materials it used in the classroom. Adds the requirement that a district's base student cost be reduced by 1% if a complaint is determined to be founded and corrective action is not taken and instead establish a 30 day process for corrective action to be taken. If SCDE determines corrective action not taken 1% of districts SHFA funds shall be withheld until district is in compliance. (Proviso 1.83 DELETED and merged with 1.53)	
1.90	General Fund (New)	Highly Qualified Teachers	ADDED- For current fiscal year, the requirement that teachers be highly qualified in the core academic subjects is suspended.	
1A.73	EIA (New)	Industry Certifications/ Credentials	ADDED - Funds appropriated ($3.0 million) for national industry exams to be allocated to districts based on number of national industry exams administered in prior school year. SCDE to work with other agencies to ensure students are award of industry required credentials for current job availability.	
1.39	General Fund	Instructional Materials Fees	AMENDED - To replace word "textbook" with "instructional materials."	
1.66 and 1A.66 in EIA	General Fund	Literacy Coaches	AMENDED - To change "reading coaches" to "literacy coaches" and "not met" to "below proficient." Clarify the duties associated with literacy coaches. Direct that unspent or unallocated literacy coach funds be used to fund summer reading camps and require SCDE report the amount of funds used for this purpose.	AMENDED - School districts must request a waiver from SC Superintendent of Education to utilize literacy coaches as interventionists as included in the district reading plan and when supported by assessment data showing the district reading plan is improving literacy.
1.82	General Fund	Literacy Initiatives	DELETED - Proviso had required SCDE to evaluate and coordinate available literacy initiatives.	
1A.78	EIA (New)	Low Achieving Schools (EIA)	ADDED - Of the funds allocated to EOC for Partnerships for Innovation, $500,000 to be allocated to parent support initiatives and afterschool programs in historically underachieving communities.	
1.87	General Fund (New)	Moving Cost Study	ADDED - To direct SCDE to submit a report on the costs associated with moving SCDE from Rutledge Building to legislature by September 1, 2016.	

Proviso	Proviso Type	Title	Description (As Adopted by House)	Passed by House 03.24.16
1.88	General Fund (New)	Technology Technical Assistance	ADDED - To direct funds appropriated by SCDE for technology technical assistance to be used to increase capacity of Abbeville districts. SCDE shall use funds to assist districts in procuring appropriate technology to include devices and infrastructure in order to build capacity to offer online testing and increased access.	
1A.29	EIA	National Board Certification Incentive	Teachers who complete the application process on or after July 1, 2010 shall be paid a $5000 salary supplement in the year of achieving certification. The supplement shall be added to the annual pay of the teacher not to exceed ten years of the national certificate.	
1A.42	EIA	National Board Certification Incentive	Teachers who complete the application process prior to July 1, 2010 shall be paid a $7500 salary supplement in the year of achieving certification. Supplement shall be added to the annual pay of the teacher for the length of the national certificate. Teachers may receive a loan equal to the amount of the application fee. Provides loan forgiveness when a teacher achieves certification: one half of loan principal and interest forgiven for teachers who applied for certification prior to July 1, 2010. Teachers who applied prior to July 1, 2010 and attained certification within three years of receiving loan will have full amount of loan forgiven. SC recertification cycle will be consistent with recertification cycle for national board for teachers who applied prior to July 1, 2010.	
1A.39	EIA	PowerSchool Dropout Data	SCDE will collect data on number of student who previously dropped out of school and who reenrolled to pursue a high school diploma. EOC will work with SCDE to determine how to calculate a dropout recovery rate for annual school and district report cards.	
1A.25	EIA	Professional Development	25% of funds must target Abbeville districts. SCDE authorized to carry forward and expend any funds for the same purpose.	
1A.23	EIA	Reading	SCDE shall ensure resources are utilized to improve student achievement in reading/literacy. 50% of appropriation shall be directed toward reading proficiency. 40% shall be directed toward classroom instruction and intervention for struggling readers and writers in grades 4-8. 10% should be directed toward acceleration to provide additional opportunities for deepening and refinement of literacy skills. 50% shall be allocated to school districts. 50% shall be allocated to SCDE to provide districts research-based strategies and PD. SCDE shall provide for an evaluation to review first year implementation and establish measurements for monitoring impact on student achievement.	

Proviso	Proviso Type	Title	Description (As Adopted by House)	Passed by House 03.24.16
1.42	General Fund	Residential Treatment Facilities	AMENDED - To require students placed in licensed residential treatment facilities that are identified on the State Qualified Providers list and meet requirements of state law.	AMENDED - If student in a RTF remains enrolled in a school district, district is eligible to receive a base student cost weighted funding of 2.10.
1A.64	EIA	Rural Teacher Recruiting Initiative	AMENDED - To direct CERRA to "publish" rather than "develop" eligibility requirements. Direct CERRA to include incentives for subsidized tuition and instructional materials; for enhanced student loan forgiveness or repayment; for establishment and maintenance of a teaching mentorship program and for technical support and recruiting incentives. Direct CERRA to also develop a process for recovering an amount equal to the incentive given to a participant who does not meet the obligations associated with the incentive.	
1.28	**General Fund**	School Districts and Special Schools Flexibility	AMENDED - In order for a district to use flexibility provisions, at least 75% of the district's per pupil expenditures must be utilized within the InSite categories of instruction and instructional support. Removed requirement that expenditures must be utilized in InSite category of non instruction pupil services. Added that no portion of the 75% could be used for operations.	AMENDED - Removed the requirement that no portion of 75% could be used for operations. Instead stated no portion could be used for facilities. Also allowed for 75% to be used for some specific non instruction pupil services (transportation, food service, and safety).
1A.76	EIA (New)	SC IT Academy	ADDED - To direct SCDE to procure an IT Academy for public schools statewide and direct the academy offer certification opportunities for educators to receive Teacher Certification Exams and for middle school students to receive certification in an office suite of products and for high school students to receive programming credentials.	
1A.54	EIA	SC Public Charter School District Funding	AMENDED - To extend the timelines for ruling on charter school application FY 2016-17 by 60 days if the Public Charter School district determines an applicant should be allowed to amend its application to meet the requirements based on their proposal to address an existing achievement gap in an underserved area of the state including, but not limited to, Abbeville districts. Direct the EOC to report at the end of the application cycle on the outcome of extending the hearing time.	
1A.70.	EIA	Statewide Assessment Procurement	DELETED - Proviso required SCDE to procure 11th grade college readiness assessment and grades 3-8 ELA and math assessment	

Proviso	Proviso Type	Title	Description (As Adopted by House)	Passed by House 03.24.16
1.50	General Fund	Student Health and Fitness	AMENDED - To allow funds for school nurses to be distributed to school districts directly rather than through a grant process.	
1A.24.	EIA	Students at Risk of School Failure	AMENDED - To change definition of poverty index to "students identified using USDA Community Eligibility guidelines and students in Medicaid, students identified as Migrant and students identified as homeless."	
1.63	General Fund	Summer Reading Camps	AMENDED - To designate $700,000 to SC Afterschool Alliance for summer reading camps in schools that have a poverty index of 40% or greater. Requires SLED background check for all volunteers, mentors, tutors. No more than 10% may be retained by Alliance for administration and planning document must be submitted to SCDE by September 15.	
1A.52	EIA	Surplus (EIA)	AMENDED - To require EIA surplus funds from prior fiscal year be allocated accordingly: $2.8 million to EOC for Partnerships for Innovation; Allendale District $150,000; Vocational Equipment Modernization $1,501,307; Industry Certification $3 million; Adult Ed $1.5 million; Power School/Data Collection $1.952 million; IT Academy $750,000, Instructional Development and Digital Content Curation $493,443	
1A.6	EIA	Teacher Recruitment	Funds appropriated for Teacher Recruitment Program, CHE will distribute 92% to CERRA for a state teacher recruitment program, with at least 78% to be used for Teaching Fellows Program, of which 22% must be used for other components including the Teacher Cadet Program and $166,302 for specific programs to recruit minority teachers. 8% will be allocated to SC State. Includes additional program requirements: CERRA will report annually by October 1 to EOC. SC State's program will extend beyond geographic area it currently serves. CHE shall evaluate the effectiveness of teacher recruitment projects and report annually by October 1 to legislature and EOC. CERRA will appoint and maintain the SC Teacher Loan Advisory Committee.	

H. 5001 General Appropriation
FY2016-17 Education Proviso Summary

Proviso	Proviso Type	Title	Description (As Adopted by House)	Passed by House 03.24.16
1.91 and 1A.79	EIA (New)	Teacher Salaries (EIA)	ADDED - SCDE directed to increase the statewide salary schedule by 2%. Local districts must provide all certified paid teachers a 2% increase. If districts already pay above the statewide salary schedule, districts shall use the additional funds made available from Teacher Salary Supplement appropriation to fund a 1% increase. District must increase salary compensation for all eligible certified teachers by an amount equal to a step on the salary schedule for any teacher entering the 23rd year if district's salary schedule does not go beyond 22 years. District may apply to SCDE for a waiver if it believes it will be unable to provide the additional step without incurring a deficit. Teachers are defined by SCDE Professional Certified Staff System. (An additional $23.2 million plus carry forward funds of $10.0 million used to pay for the increase, along with increase in EFA)	AMENDED - Clarifies that of the total 2% increase, 1% of the increase is funded by EFA and 1% of the increase is funded by EIA.
1A.38	EIA	Teacher Salaries/ Southeast Average	AMENDED - To update projections for Southeastern average teacher salary from $49,975 to $51,495.	
1A.4	EIA	Teacher Salaries/State Agencies	AMENDED - To direct that affected agencies receive increases for teacher salaries as recommended by the EOC rather than an allocation from the specific line item. Deletes the requirement that the Executive Budget Office distribute the line item funds and instead direct that teacher salary increases recommend by the EOC and funding in this act be incorporated into each agencies EIA appropriations.	
1.75	General Fund	Teacher Salary Schedule Study	AMENDED - To direct SCDE to include salary needs information from Abbeville districts. Allows work to continue by deleting the deadline for submission of recommendations.	
1A.9	EIA	Teacher Supplies	AMENDED - To add on certified public school teachers and career specialists to those individuals eligible to receive the reimbursement and to also allow those eligible individuals employed by a publically funded full day 4K classroom.	

H. 5001 General Appropriation
FY2016-17 Education Proviso Summary

Proviso	Proviso Type	Title	Description (As Adopted by House)	Passed by House 03.24.16
1A.12	EIA	Technical Assistance (EIA)	AMENDED - To provide intensive support to schools and districts with the lowest percentages of students meeting state standards on state assessments on the most recent state assessments or with the lowest high school graduation rates or with an absolute rating of below average or at risk on school report card. SCDE will create a system of tiers on TA and low performing schools and districts shall be placed within the tiered framework no later than December 15. Low performing schools shall receive a diagnostic review through SCDE and new identified schools and districts must be reviewed by an External Review Team in the year of designation and every third year after. Schools and districts shall submit an updated school renewal or district strategic plan; plans shall address specific strategies designed to increase student achievements and must include measures to evaluation success. TA experts shall assist schools and districts in designing and implementing the strategies and measurements identified. SCDE must monitor plan progress on implementation and report findings to local legislative delegation and Governor. SCDE Superintendent may declare a state or emergency in a school if the accreditation status is probation or denied or if the school fails to show improvement on the state accountability system. SCDE Superintendent may take over management of the school district. Eliminates direct funding to underperforming schools.	
1A.63.	EIA	Technology and Device Pilot Project	DELETED - SCDE to use surplus funds to pilot devices or digital content in up to six school districts	
1.73	General Fund	Transition Funds to Districts	DELETED - Proviso had allocated funds to districts that were negatively impacted by changes in EFA weightings in prior years and in realignment of EIA and lottery appropriations.	
1.78	General Fund	Transition Funds to Districts	DELETED - Proviso had allocated funds to districts that were negatively impacted by changes in EFA weightings in prior years and in realignment of EIA and lottery appropriations.	
1A.5	EIA	Work Based Learning	$75,000 of funds appropriated for work based learning program will be used by SCDE to provide regional PD in career guidance for teachers, guidance counselors and training mentors. $500,000 to be allocated for Regional Career Specialists.	

NOTE: Excluded from the list above are any proviso changes that update fiscal year or date references.
SOURCE: House Ways and Means Committee, Committee Postings and Reports, through March 14, 2016. General Appropriation Bill H. 5001 through March 22, 2016.

	2014-15	2015-16	Revised 2015-16 Base (Act 92)	EOC Recommendations	Governor's Executive Budget	Explanation and Notes	House Budget March 24
A. STANDARDS, TEACHING, LEARNING, ACCOUNTABILITY							
1. Student Learning							
Personal Service Classified Positions	58,629	58,629	58,629				$58,629
Other Operating Expenses	136,739	136,739	136,739				$136,739
High Achieving Students							
Aid to Districts	37,386,600	37,386,600	37,386,600	**$36,435,524**	**($37,386,600)**	Governor: Funded Base student cost at $2,300, an $80 increase with $165.3 million increase in EFA EOC: Balance of EIA revenues to school districts for initiatives to improve college & career readiness. Prov. 1.3	$37,386,600
School Health & Fitness Act -- Nurses	6,000,000	6,000,000	6,000,000				$6,000,000
Tech Prep	3,021,348	3,021,348	3,021,348				$3,021,348
Modernize Vocational Equipment *	6,682,406	7,260,261	7,260,261	**$1,501,307**	**$8,040,029**	EOC, Governor and House: Annualization and proviso increasing minimum district allocation from $20,000 to $50,000. Prov. 1A.52 provides $1,501,307 for modernization.	$13,798,983
New: Mobile Device Access & Management					**$5,220,000**	Governor: Provide mobile access to high-speed internet for underserved children. House: Funded $3.0 million in lottery funds for mobile device access and management	
New: Student Connectivity Infrastructure					**$5,781,862**	Governor: Provide 'behind the wall' supports for education technology infrastructure in depressed areas and support from state network engineers House: Funded $13.9 million in non-recurring General Funds and $2.8 million in Capital Reserve Fund monies (H.5002)	
Arts Curricula	1,487,571	1,487,571	1,487,571				$1,487,571
Adult Education	13,573,736	15,073,736	15,073,736		**$1,500,000**	Governor: To support adult education programs statewide. Also see Prov. 1A.52 for additional $1.5 million allocated to Adult Ed.	$15,073,736
Students at Risk of School Failure	79,551,723	79,551,723	79,551,723				$79,551,723
High Schools That Work	2,146,499	2,146,499	2,146,499	**$1,309,051**	**$0**	EOC: In FY2015-16 there are 432 sites in 227 high school/career centers and 205 middle schools. To assist districts in defraying cost of program, $8,000 per site is recommended	$2,146,499
Summer Reading Camps		1,500,000	1,500,000				$1,500,000
Reading Coaches		4,961,278	4,961,278				$9,922,556
EEDA	6,013,832	6,013,832	6,013,832				$8,413,832
Subtotal	**156,059,083**	**164,598,216**	**164,598,216**				**$178,498,216**
2. Student Testing							
Personal Service Classified Positions	488,518	488,518	488,518				$488,518
New: Additional Positions					**$300,000**	Governor: As recommended by SCDE	$60,000
Other Operating Expenses	332,948	332,948	332,948		**$29,000**	Governor: As recommended by SCDE. House: increased operations and funded one FTE.	$678,748
Assessment / Testing	27,261,400	27,261,400	27,261,400		**$10,200,000**	Governor: Funding for new state assessments	$27,261,400
Subtotal	**28,082,866**	**28,082,866**	**28,082,866**				**$28,488,666**
3. Curriculum & Standards							
Personal Service Classified Positions	126,232	126,232	126,232				$126,232
Other Personal Service	4,736	4,736	4,736				$4,736
Other Operating Expenses	41,987	41,987	41,987				$41,987
Reading	6,542,052	6,542,052	6,542,052				$6,542,052

	2014-15	2015-16	Revised 2015-16 Base (Act 92)	EOC Recommendations	Governor's Executive Budget	Explanation and Notes	House Budget March 24
Instructional Materials	20,922,839	20,922,839	20,922,839		$6,930,401	Governor: Increase above non-recurring appropriation. House: Funded $18.0 million in lottery revenues.	$20,922,839
Subtotal	**27,637,846**	**27,637,846**	**27,637,846**				**$27,637,846**
4. Assistance, Intervention, & Reward							
Personal Service Classified Positions	1,236,436	1,236,436	1,236,436				$1,236,436
Other Operating Expenses	1,174,752	1,174,752	1,174,752		$200,000	Governor: As recommended by SCDE	$1,374,752
EAA Technical Assistance	8,800,000	8,800,000	8,800,000			AdvancEd implementation by SCDE.	$11,301,301
PowerSchool/Data Collection	7,500,000	7,500,000	7,500,000		$1,952,000	Governor: For PowerSchool student information systems	$7,500,000
Subtotal	**18,711,188**	**18,711,188**	**18,711,188**				**$21,412,489**
B. Early Childhood							
Personal Service Classified Positions	376,246	376,246	376,246				$376,246
New: Additional Positions					$325,000	Governor: As recommended by SCDE. 7 FTEs.	$455,000
Other Operating Expenses	556,592	556,592	556,592				$556,592
Alloc EIA - 4 YR Early Child	15,513,846	15,513,846	15,513,846				$15,513,846
SCDE-CDEPP	34,324,437	34,324,437	34,324,437				$34,324,437
Subtotal	**50,771,121**	**50,771,121**	**50,771,121**				**$51,226,121**
C. TEACHER QUALITY							
1. Certification							
Personal Service Classified Positions	1,068,102	1,068,102	1,068,102				$1,068,102
Other Personal Service	1,579	1,579	1,579				$1,579
Other Operating Expenses	638,999	638,999	638,999				$638,999
Subtotal	**1,708,680**	**1,708,680**	**1,708,680**				**$1,708,680**
2. Retention & Reward							
Special Items							
Teacher of the Year Award	155,000	155,000	155,000				$155,000
Teacher Quality Commission	372,724	372,724	372,724				$372,724
Teacher Salary Supplement	127,640,691	127,640,691	127,640,691			House: 2% teacher salary increase; 1% to be funded by EIA and 1% to be funded by EFA. Also adds a 23rd step to minimum teacher salary schedule. $23.2M from EIA. See also Prov. 1.91 and 1A.79.	$150,823,453
Allocation-EIA Employer Contributions	15,766,752	18,266,752	18,266,752				$18,266,752
National Board Certification	55,500,000	54,000,000	54,000,000				$54,000,000
Rural Teacher Recruiting Initiative		1,500,000	1,500,000		$13,500,000	Governor: To expand teacher recruitment and retention policies in rural districts	$9,748,392
Teacher Supplies	13,596,000	13,596,000	13,596,000	$750,000	$750,000	EOC and Governor: Estimated 52,000+ teachers at $275	$14,346,000
Subtotal	**213,031,167**	**215,531,167**	**215,531,167**				**$247,712,321**
3-4. Professional Development and ADEPT							
Special Items							
Professional Development	5,515,911	9,515,911	9,515,911				$9,515,911
New: Additional Positions					$130,000	Governor: As recommended by SCDE	$65,000
New: Other Operating					$50,000	Governor: As recommended by SCDE	
ADEPT	873,909	873,909	873,909				$873,909
Subtotal	**6,389,820**	**10,389,820**	**10,389,820**				**$10,454,820**
D. LEADERSHIP							

	2014-15	2015-16	Revised 2015-16 Base (Act 92)	EOC Recommendations	Governor's Executive Budget	Explanation and Notes	House Budget March 24
1. Schools							
2. State							
Personal Service Classified Positions	82,049	82,049	82,049				$82,049
Other Personal Service	83,121	83,121	83,121				$83,121
Other Operating Expenses	279,032	279,032	279,032				$279,032
Technology	10,171,826	12,271,826	12,271,826				$12,271,826
Employer Contributions	1,064,221	1,064,221	1,064,221		**$241,600**	Governor: For state employees at SCDE funded through EIA	$1,249,821
Subtotal	**11,680,249**	**13,780,249**	**13,780,249**				**$13,965,849**
E. PARTNERSHIPS							
1. Business and Community							
2. Other Agencies & Entities							
State Agency Teacher Pay (F30)	73,861	73,861	73,861	**$217,474**		EOC: Salaries of Teachers in Special Schools to be at level of district of residence	$73,861
Education Oversight Committee (A85)	1,643,242	1,793,242	1,793,242				$1,793,242
Center for Educational Partnerships (H27)	715,933	715,933	715,933				$715,933
SC Council on Economic Education	300,000	300,000	300,000				$300,000
Science PLUS	503,406	563,406	563,406				$563,406
Gov. School Arts & Humanities (H63)	959,994	959,994	959,994	**$138,025**	**$138,025**	EOC, Governor and House: Salaries of Teachers in Special Schools to be at level of district of residence	$1,192,439
Wil Lou Gray Opportunity School (H71)	605,294	605,294	605,294				$605,294
School for Deaf & Blind (H75)	7,439,286	7,439,286	7,439,286				$7,439,286
Disabilities & Special Needs (J16)	613,653	613,653	613,653	**($65,000)**		EOC and House: Requested by DDSN	$548,653
John De La Howe School (L12)	417,734	417,734	417,734				$417,734
Clemson Ag Ed Teachers	889,758	889,758	889,758			House: Prov. 1A.28: Increase of $100,000 allocated to Clemson PSA.	$989,758
Centers of Excellence-CHE (H03)	1,137,526	1,137,526	1,137,526				$1,137,526
Teacher Recruitment Program-CHE (H03)	4,243,527	4,243,527	4,243,527				$4,243,527
SC Program for the Recruitment and Retention of Minority Teachers, SC State University (Base: $339,482)							
Center for Ed, Recruitment, Ret, and Adv	531,680	531,680	531,680				$531,680
Teacher Loan Program-State Treasurer (E16)	5,089,881	5,089,881	5,089,881				$5,089,881
Gov. School Science & Math (H63)	533,130	533,130	533,130	**$63,241**		EOC and House: Salaries of Teachers in Special Schools to be at level of district of residence	$719,425
STEM Centers SC	1,750,000	1,750,000	1,750,000	**$350,000**		EOC: STEM Teacher Fellows Initiative recruiting STEM teachers at the secondary levels in rural communities	$1,750,000
Teach For America SC	3,000,000	3,000,000	3,000,000				$3,000,000
ETV - K-12 Public Education	2,829,281	2,829,281	2,829,281				$3,394,281
ETV - Infrastructure	2,000,000	2,000,000	2,000,000				$2,000,000
SC Youth Challenge Academy	1,000,000	1,000,000	1,000,000				$1,000,000
Literacy & Distance Learning	415,000	415,000	415,000				$415,000
Regional Education Centers (P32)	1,302,000	1,302,000	1,302,000	**$300,000**	**$300,000**	EOC and House: STEM Premier in at least 100 high schools; Currently, 15 districts and 50 high schools participating	$1,802,000
Reach Out and Read (A85) **		1,000,000	1,000,000				$1,000,000
Arts Curricula (H910)		1,000,000	1,000,000				$1,000,000
Family Connection							$300,000
Subtotal	**37,994,186**	**40,204,186**	**40,204,186**				**$42,022,926**
F. TRANSPORTATION/BUSES							
Other Operating	12,575,684	12,575,684	12,575,684				$12,575,684

	2014-15	2015-16	Revised 2015-16 Base (Act 92)	EOC Recommendations	Governor's Executive Budget	Explanation and Notes	House Budget March 24
New: Allocation to Districts - Driver Salaries					**$19,200,000**	Governor: To improve driver salaries targeted at rural and underserved districts. House: Funded in General Funds.	
New: Hazardous Transportation Route					**$3,500,000**	Governor: To reimburse districts for providing bus services along hazardous routes. House: Funded in General Funds.	
Subtotal	**12,575,684**	**12,575,684**	**12,575,684**				**$12,575,684**
G. Charter School District							
Charter School District	**56,253,692**	68,131,619	68,131,619	**$12,987,128**	**$11,547,128**	EOC, Governor and House: Growth in students in SC Public Charter School District schools	**$81,118,747**
Charter Schools Chartered by Institutions of Higher Education		1,440,000	1,440,000				
Subtotal		**69,571,619**	**69,571,619**				
H. First Steps to School Readiness							
Personal Services	2,182,993	2,182,993	2,182,993				$150,000
Classified Positions							$1,911,453
Unclassified Positions							$121,540
New: New Positions					**$268,432**		$268,432
Other Operating	1,872,789	1,872,789	1,872,789		**$783,436**		$1,906,225
County Partnerships	11,262,214	12,693,265	12,693,265		**$1,741,963**		$14,435,228
CDEPP	9,767,864	9,767,864	9,767,864				$9,767,864
BabyNet Autism Therapy	437,476	1,699,848	1,699,848		**$3,186,560**		$5,186,408
Employer Contributions							$1,015,485
Subtotal	**25,523,336**	**28,216,759**	**28,216,759**				**$34,762,635**
EIA TOTAL	**$646,418,918**	**$681,779,401**	**$681,779,401**	**$53,986,750**	**$58,428,836**		**$751,585,000**
Non-Recurring Appropriations				**$12,146,750**	**$12,146,750**	EOC and Governor: Recommend additional EIA revenues collected in FY16 to be allocated to instructional materials	
Proviso 1A.52							
EOC Partnerships for Innovation		$900,000					$2,800,000
Allendale County School District		$150,000					$150,000
Modernize Vocational Equipment		$1,501,307					$1,501,307
Industry Certification		$7,300,000				EOC, Governor and House: SCDE Requested $5.0 million. EOC recommends phasing-in payment of industry exams. See also Prov. 1A.73.	$3,000,000
Adult Education		$625,000					$1,500,000
PowerSchool/Data Collection		$1,952,000					$1,952,000
IT Academy		$750,000				See also Prov. 1A.76	$750,000
Digital Content Curation		$493,443					$493,443
Subtotal:		**$13,671,750**					**$12,146,750**

EDUCATION OVERSIGHT COMMITTEE

Subcommittee: EIA and Improvement Mechanisms

Date: April 11, 2016

INFORMATION
Educational Performance of Military-Connected Children

PURPOSE/AUTHORITY
Act 289, the Military Family Quality of Life Enhancement Act, was enacted in 2014. The law requires the Education Oversight Committee (EOC) to develop an annual report on the educational performance of military connected children:

> The Education Oversight Committee, working with the State Board of Education, is directed to establish a comprehensive annual report concerning the performance of military connected children who attend primary, elementary, middle, and high schools in this State. The comprehensive annual report must be in a reader-friendly format, using graphics wherever possible, published on the state, district, and school websites, and, upon request, printed by the school districts. The annual comprehensive report must address at least attendance, academic performance in reading, math, and science, and graduation rates of military connected children.

CRITICAL FACTS
EOC staff worked with staff from the SC Department of Education, the Department of Defense State Liaison Office, and the Virginia Board of Education on the report.

TIMELINE/REVIEW PROCESS
Report issued annually in March or April.

ECONOMIC IMPACT FOR EOC

Cost: No fiscal impact beyond current appropriations

Fund/Source:

ACTION REQUEST

☐ **For approval** ☒ **For information**

ACTION TAKEN

☐ **Approved** ☐ **Amended**

☐ **Not Approved** ☐ **Action deferred (explain)**

2016

EDUCATIONAL PERFORMANCE OF MILITARY-CONNECTED CHILDREN

SC EDUCATION
OVERSIGHT COMMITTEE

PO Box 11867 | 227 Blatt Building | Columbia SC 29211 | WWW.SCEOC.ORG

TABLE OF CONTENTS

Introduction 1

Acknowledgements 3

Summary of Findings and Recommendations 5

Section I: Recent Developments 7

Section II: Military-Connected Student Population 15

Section III: Military-Connected Student Performance 19

Appendices

A: Resources for Military-Connected Students and Families 25

B: Impact Aid Section of Every Student Succeeds Act 28

C: State of Virginia Law: Uniformed Services-Connected Students 38

D: Military-Connected Students by District, January 2016 39

Introduction

March 21, 2016

In 2014, the General Assembly passed Act 289, the Military Family Quality of Life Enhancement Act. The Act's purpose is to "enhance many quality of life issues for members of the armed forces" (Act 289 Preamble). Part V requests the SC Education Oversight Committee (EOC) to develop an annual report on the educational performance of military connected children:

> The Education Oversight Committee, working with the State Board of Education, is directed to establish a comprehensive annual report concerning the performance of military connected children who attend primary, elementary, middle, and high schools in this State. The comprehensive annual report must be in a reader-friendly format, using graphics wherever possible, published on the state, district, and school websites, and, upon request, printed by the school districts. The annual comprehensive report must address at least attendance, academic performance in reading, math, and science, and graduation rates of military connected children.

Based upon Act 289's direction, this report does not address military-connected students educated in Department of Defense schools, private schools and home school settings. The EOC evaluation team worked closely with the military and education community as it developed this report. Professionals, who directly support military families, including school liaison officers and organizations that support the National Guard and Reserves, provided input and reviewed report drafts. Both the South Carolina Department of Education (SCDE) and Defense Manpower Data Center provided data. The 2016 report provides:

- An overview of the federal Impact Aid Program;
- Recent policy developments that impact the identification and reporting on military-connected students;
- Details regarding the demographics of military-connected students;
- An update on the academic performance and school attendance of military-connected students.

Acknowledgements

The EOC is grateful for the assistance of local, state and national organizations and staff in the development of this report. Report contributors include:

Dino Teppara, SC Department of Education

Cynthia Hearn, SC Department of Education

Wanda Davis, SC Department of Education

Kevin Bruch, Department of Defense State Liaison Office

Melissa Luchau, Virginia Board of Education

Summary of Findings and Recommendations

Finding 1: Using federal Fiscal Year 2014-15 US Department of Education data, twelve school districts in South Carolina applied for Impact Aid funding under Sections 8002 and 8003.[1] Estimated FY 2015 payments were approximately $1.1 million. If the Impact Aid Program was fully funded at the federal level, payments to South Carolina districts would be $13.6 million. Total payments represented about 7,835 federally-connected students, with a majority of those students enrolled in Berkeley, Charleston, Richland 2 and Sumter school districts.

Finding 2: The federal Every Student Succeeds Act (ESSA) includes modifications to the reporting of military-connected students and Impact Aid. Impact Aid changes will improve the timeliness of payments, adjust the Basic Support formula, eliminate the Federal Properties "lockout" provision and provide additional budget certainty for districts.

The reporting requirement will provide more consistent, easily identifiable data regarding military-connected students with a parent on active duty. As student identification improves, additional supports may be implemented to assist students who live with perpetual challenges due to frequent moves, parental and sibling deployments and transitions that include reintegration and dealing with profoundly changed parents.

Recommendation 1: The State of Virginia passed a law last year requiring the Virginia Department of Education to provide non-identifiable aggregate data on newly-enrolled military-connected students that will be made available to local, state and federal entities. Virginia's reporting policy should be explored further as an approach that will support further the unique needs of military-connected students while also meeting ESSA reporting requirements.

Finding 3: As a state, South Carolina continues to underreport the number of military-connected students. It is important to note that based upon Act 289's direction, this report does not address military-connected students educated in Department of Defense schools, private schools and home school settings. Data provided by the Department of Defense Education Activity (DoDEA), indicates there were 12,335 military-connected students with at least one active duty parent in November 2015.[2] January 2016 data provided by the SC Department of Education (SCDE) indicates there are 7,763 military-connected students, representing a 37 percent underreporting of the DoDEA estimate. However, SCDE reported 7,308 military-connected students in 2014 and 7,763 military-connected students with an active duty parent in January 2016. This represents a six percent increase in reported military-connected students. Since the actual number of military-connected students reported by DoDEA decreased from 13,597 to 12,335, it is possible the six percent increase in military-connected students is due to improved reporting by the districts.

---

[1] FY 2015 SCDE data for Impact Aid was not available. Most recent SCDE data for Impact Aid was FY 2013.

[2] The **Department of Defense Education Activity** (**DoDEA**) is a civilian agency of the United States Department of Defense that manages all schools for military children and teenagers in the United States and also overseas at American military bases worldwide.

Recommendation 2: While there is a modest six percent increase in reporting of military-connected students, SCDE should continue its efforts to improve reporting of military connected students since 37 percent of military connected students in the state are not being reported in PowerSchool.

Finding 4: Approximately 90 percent of military-connected students attend school in the following districts: Richland 2, Berkeley, Dorchester 2, Beaufort, Lexington 1, Sumter, Kershaw, Oconee, Charleston and the SC Public Charter School District.

Finding 5: Military-connected students in South Carolina continue to outperform their peers on state-administered tests.

- On average, an additional 9.6 percent of military-connected students tested as “ready” or “exceeding” on ACT Aspire in reading, and an additional 7.8 percent of military-connected students tested as “ready” or “exceeding” on ACT Aspire in mathematics. On the science test of SCPASS the performance gap was even larger; on average an additional 12 percent of military-connected students tested as “met” or better. In eighth grade, an additional 11.9 percent of military-connected students tested as “met” or better on SCPASS science and an additional 11.3 percent of military-connected students tested as “ready” or “exceeding” on ACT Aspire reading.
- On the EOCEP tests, military-connected students continue to outperform all students statewide. However, the gap between military-connected students and all students statewide may be narrowing. The average difference between military-connected students and all students statewide is 3.8 points.

I. Recent Developments

Impact Aid Background

The Impact Aid Program was signed into law in 1950; approximately 1,300 school districts enrolling more than 11 million students receive Impact Aid funding. Impact Aid was incorporated into the Elementary and Secondary Education Act in 1994. It reimburses school districts for the loss of local tax revenue due to the presence of the Federal Government. Federal activities reduce local taxes because Federal property is removed from the tax rolls and/or the school district is educating students with no or reduced tax revenue associated with federally-connected students. Examples of federal impaction include: military installations, Indian Trust, Treaty and Alaska Native Claims Settlement Act Lands, civil service activities such as veterans hospitals, Federal agencies and national parks, and low-rent housing properties owned by the Federal Government.

Each school district must submit an Impact Aid application annually to the U.S. Department of Education (USDE). USDE allocates funding in multiple installments until all available funds are distributed. The Impact Aid Program has not been fully funded since 1969. Local school districts can qualify for Impact Aid through various sections of the Program. States are restricted from reducing state aid for a federally-connected school district because of its receipt of Impact Aid Funding. However, if a state has a school finance formula designed to equalize expenditure for all school districts in that state, and if the state meets several other criteria, the state can reduce the amount of state funding allocated to a specific school district based on its Impact Aid payment. This equalization policy currently does not apply to South Carolina's Education Finance Act.

Section 8002 reimburses school districts that have lost significant local revenue due to federal ownership of land within school district boundaries. Section 8003 payments for federally-connected students compensate school districts for educating students, whose parents or legal guardians reside and/or work on federal property, including children of members of the uniformed services, and children who reside on Native American Lands. Section 8003 is the largest component of the Impact Aid Program. Federally-connected children are defined as children whose parent(s):

- live on Indian trust, treaty land or ANSCA land,
- are members of the uniformed services and reside on a military installation,
- are members of the uniformed services but who reside off the military installation,
- reside in Federal low-rent housing (not Section 8 housing),
- both live and work on Federal property

- work or live on federal property.

Section 8003(D) provides payments to school districts for the additional costs associated with educating military and Indian Lands students with disabilities, identified as those with an active Individual Education Plan. Funding has to be spent on an IDEA-eligible activity. Section 8004 funding is available to districts with children residing on Indian Lands. Affected districts must establish Indian Policies and Procedures to ensure the school district meets certain requirements. Payments for districts to meet capital or construction needs of Federal students are allowable under Section 8007.

Table 1
School District Eligibility for Impact Aid Funding[3]

Section	Description	School District Eligibility
Section 8002 (Federal Property)	Eligible federal lands include: national parks and grasslands, national laboratories, Army Corps of Engineers projects, military testing grounds, expansions of military installations and environmental waste sites.	Property was acquired by the Federal Government after 1938 and the assessed valuation of the property, when it was acquired by the Federal Government, is at least 10% of the school district's total assessed value.
Section 8003 (Basic Support Payments for Federally-Connected Children)	A district can choose one of the following two options to count students: (1) use a USDE-approved form for counting each enrolled child or (2) use SourceCheck, a USDE-approved means of counting the members of a school district's federally-connected children. It is provided to a parent's employer, a housing officials and a tribal official.	A school district must educate at least 400 Federal students in average daily attendance or have at least a 3% average daily attendance of Federal students.
Section 8003(D) (Children with Disabilities)		Military and Indian Lands students with disabilities with an active IEP.
Section 8004 (Children Residing on Indian Lands)	Districts with children residing on Indian Lands must establish Indian Policies and Procedures to ensure district meets certain requirements.	
Section 8007 (Construction)	Payments for districts to meet capital or construction needs of Federal students.	Two different allocations: 60% is dedicated for competitive grants and 40% is allocated in formula payments to eligible school districts.

Source: National Association of Federally Impacted Schools, "The Basics of Impact Aid."

[3]National Association for Federally Impacted Schools ,"The Basics of Impact Aid." May be accessed at: http://media.wix.com/ugd/423d5a_751601531b7c42948bf292f68a8c8a77.pdf.

Actual payments to a school district are complicated due to the program not being full funded since 1969. It is estimated there is a 45 percent unmet needs. Payments are reduced and distributed on a needs-based formula. The Learning Opportunity Threshold (LOT) is the percentage that shows how dependent a school district is on Impact Aid funding. As the LOT percentage increases, a school district's payment also increases.

School districts with a total student enrollment of fewer than 1,000 students in average daily attendance and have a per-pupil expenditure that is less than the state or national average per-pupil expenditure qualify for an automatic 40 percent of the Learning Opportunity Threshold. In addition, districts may qualify for additional funding if they are considered to be heavily impacted by the presence of the Federal government.

Using federal Fiscal Year 2014-15 USDE data, twelve school districts in South Carolina applied for Impact Aid funding under Sections 8002 and 8003.[4] Estimated FY 2015 payments were approximately $1.1 million. Data from the National Association of Federally Impacted Schools are included below. SCDE data are not included because the most recent data available is from Fiscal Year 2012-13.

If the Impact Aid Program was fully funded at the federal level, payments to South Carolina districts would be $13.6 million. Total payments represented about 7,835 federally-connected students, with a majority of those students enrolled in Berkeley, Charleston, Richland 2 and Sumter school districts.

[4] FY 2015 SCDE data for Impact Aid was not available. Most recent SCDE data for Impact Aid was FY 2013.

Table 2
Impact Aid Section 8003 Payment Estimates, FY 2015[5]

School District	Learning Opportunity Threshold	Estimated FY 2015 Payment	Maximum FY 2015 Payment	Total Federally-Connected Students Average Daily Attendance
Section 8002 – Federal Property Payments				
Anderson 4		$216,608	$3,165,436	
Section 8003 – Basic Support Payments				
Aiken	2%	$5,019	$308,402	368.19
Beaufort	0%			0
Berkeley	10%	$306,137	$3,252,767	2525.51
Charleston	4%	$59,367	$1,744,148	1419.06
Dorchester 2	0%			0
Florence 3	0%			0
Marion	0%			0
Richland 1	1%	$942	$136,912	149.72
Richland 2	8%	$174,405	$2,344,149	1731.69
Sumter	12%	$292,120	$2,648,458	1554.67
Barnwell 29	40%	$18,493	$49,713	86.64
TOTAL		**$1,073,091**	**$13,649,985**	**7835.48**

Source: National Association of Federally Impacted Schools, "2015 Blue Book (8003)."

Every Student Succeeds Act

There have been recent changes to Impact Aid and the identification of military-connected students due to the passage of the federal Every Child Succeeds Act (ESSA) in December of 2015. Appendix B includes Impact Aid section of ESSA. ESSA requires the disaggregation of student-level data, including the identification, collection and reporting of military-connected students. ESSA also addresses Impact Aid. Funding authorization for Impact Aid is stagnant for the first three years of the four-year authorization. However, some changes to Impact Aid were made:

- technical and formula changes to federal properties that have already reduced program subjectivity and increased timeliness of payments were made permanent;
- the Federal Properties "lockout" provision that prevented eligible federally-impacted school districts from accessing Impact Aid funding was eliminated;
- adjusted the Basic Support formula to ensure equal proration when appropriations are sufficient to fund the Learning Opportunity Threshold;

[5] National Association for Federally Impacted Schools, "2015 Blue Book (8003)." May be accessed at: http://media.wix.com/ugd/423d5a_5bc3ae0d915648a08deffe8209c850ae.pdf.

- included a hold harmless provision to provide budget certainty to school districts facing a funding cliff or significant changes to their federally-connected student enrollment.[6]

ESSA also requires the state identification, collection and reporting of military-connected students in Title I, Part A, Section 1011:

> "(ii) For all students and disaggregated by each subgroup of students described in subsection (b)(2)(B)(xi), homeless status, status as a child in foster care, and status as a student with a parent who is a member of the Armed Forces (as defined in section 101(a)(4) of title 10, United States Code) on active duty (as defined in section 101(d)(5) of such title), information on student achievement on the academic assessments described in subsection (b)(2) at each level of achievement, as determined by the State under subsection (b)(1).[7]

This federal requirement will provide more consistent, easily identifiable data regarding military-connected students with a parent on active duty. As student identification improves, additional supports may be put into place to assist students who live with perpetual challenges presented by frequent moves, parental and sibling deployments, and transitions that include reintegration and dealing with profoundly changed parents. The well-being of these children depends heavily on a network of supportive adults who are trained to identify early signs of emotional or physical challenge.

Virginia – Model Reporting Policy

States have started to improve their identification and reporting on military-connected students. The State of Virginia passed a law last year requiring the Virginia Department of Education to provide non-identifiable aggregate data on newly-enrolled uniformed services-connected students that will be made available to local, state, and federal entities. The primary purpose of the law is to meet the needs of military-connected students. It will also be helpful for districts to determine their eligibility for non-general fund and Impact Aid funding. Appendix C includes the Virginia law.

A "uniformed services-connected student" is defined as a student enrolled in a public school whose parent is serving in either (i) the active component of the U.S. Army, Navy, Air Force, Marine Corps, Coast Guard, or National Guard, the Commissioned Corps of the National Oceanic and Atmospheric Administration, or the Commissioned Corps of the U.S. Public Health Services or (ii) the reserve component of the U.S. Army, Navy, Air Force, Marine Corps, Coast Guard, or National Guard.

---

[6] National Conference of State Legislatures, "Summary of the Every Student Succeeds Act, Legislation Reauthorizing the Elementary and Secondary Education Act." May be accessed at: http://www.ncsl.org/documents/capitolforum/2015/onlineresources/summary_12_10.pdf.

[7] Every Student Succeeds Act. May be accessed at: https://www.congress.gov/bill/114th-congress/senate-bill/1177/text#HBCB1043F254B467C880CA4632EB8661D.

The Virginia Department of Education student records collection system provides the mechanism for identifying uniformed services-connected students. A field was added to the student records collection system for identification of uniformed services-connected students. Beginning fall 2015, the Virginia Department of Education provides non-identifiable aggregate data on uniformed services-connected students available to local, state, and federal entities for the purposes of becoming eligible for non-general fund sources and receiving services to meet the needs of uniformed services-connected students. Uniformed services-connected students are not an "accountability subgroup," and the new law expressly prohibits the use of the data on such students for the purposes of student achievement, the Standards of Accreditation, student-growth indicators, the school performance report card, or any other school rating system. The military student identifier does not take the place of Impact Aid collection in Virginia. Since Impact Aid requires more extensive information than the identifier, Virginia determined it would be "overly burdensome" to require all school districts to ask the more detailed Impact Aid questions.[8]

While the federal Every Student Succeeds Act requires the identification and collection of military-connected student data, South Carolina has an established mechanism for collecting this information. The South Carolina Department of Education (SCDE) manages PowerSchool, the student identification system that is provided to school districts. It is the primary source for student data and is often used for state and federal reporting requirements. In PowerSchool, a "Parent Military Status" field includes a drop-down list with eight possible student status options:

- (blank) – Neither Parent nor Guardian is serving in any military service.
- 01 – A Parent or Guardian is serving in the National Guard but is not deployed.
- 02 – A Parent or Guardian is serving in the Reserves but is not deployed.
- 03 – A Parent or Guardian is serving in the National Guard and is currently deployed.
- 04 – A Parent or Guardian is serving in the Reserves and is currently deployed.
- 05 – A Parent or Guardian is serving in the military on active duty but is not deployed.
- 06 – A Parent or Guardian is serving in the military on active duty and is currently deployed.
- 07 – The student's Parent or Guardian died while on active duty within the last year.
- 08 – The student's Parent or Guardian was wounded while on active duty within the last year.[9]

[8] Luchau, Melissa; Virginia Department of Education. "Implementing Virginia's Military Student Identifier." May be accessed at: http://www.doe.virginia.gov/support/student_family/military/va_council/meetings/2015/092915_military_identifier_implementation.pdf.

[9] SC Department of Education, "PowerSchool Data Collection Manual, January Update 2015-16." May be accessed at: http://www.ed.sc.gov/scdoe/assets/File/data/information-systems/power-school/SC_PS_Data%20Collection-Specific_Fields_Combo_Jan%202016%20Update.pdf.

This field remains unchanged from the 2014-15 school year. In last year's report, the EOC recommended this field be revised to reflect criteria for qualification for federal impact aid and provide more information regarding students that may be helpful for district and school staff. In a previous section, Virginia's approach to a state-assigned code to identify military-connected students was discussed. Virginia's student data system includes three codes for military-connected students:

- 1 = student is not military connected
- 2 = Active duty; student is a dependent of a member of the Active Duty Forces (Army, Navy, Air Force, Marine Corps, Coast Guard, or National Guard, the Commissioned Corps of the National Oceanic and Atmospheric Administration, or the Commissioned Corps of the U.S. Public Health Services)
- 3 = Reserve; student is a dependent of a member of the National Guard or Reserve Forces (Army, Navy, Air Force, Marine Corps, Coast Guard, or National Guard.[10]

If accurately and consistently entered, SCDE's "Parent Military Status" field appears to provide additional information that is not collected in Virginia's student data system. However, by state law, Virginia requires the data be entered three times during the school year – fall, spring and end-of-year. Currently, South Carolina does not have a state law that requires the collection and entry of military-connected student data.

[10]Luchau, Melissa; Virginia Department of Education. "Implementing Virginia's Military Student Identifier." May be accessed at: http://www.doe.virginia.gov/support/student_family/military/va_council/meetings/2015/092915_military_identifier_implementation.pdf.

II. Military-Connected Student Population

National, state and local district collection of military-connected student data is inconsistent. ESSA requires the disaggregation of student-level data, including military-connected students. When this requirement is fully implemented, data collection should become more consistent and accurate. As a state, South Carolina continues to underreport the number of military-connected students. Data provided by the Department of Defense Education Activity (DoDEA), indicates there were 12,335 military-connected students with at least one active-duty parent in November 2015.[11] January 2016 data provided by the SC Department of Education (SCDE) indicates there are 7,763 military-connected students, representing a 37 percent underreporting of the DoDEA estimate. However, SCDE reported 7,308 military-connected students in 2014 and 7,763 military-connected students with an active duty parent in January 2016. Appendix D provides more detail about number of military-connected students by school district. This represents a six percent increase in reported military-connected students. Since the actual number of military-connected students reported by DODEA decreased from 13,597 to 12,335, it is possible the six percent increase in military-connected students is due to improved reporting by the districts.

Table 3
Estimated Number of Military-Connected Students with Active-Duty Parent, 2012-2016

Active Duty	Sept. 2012 DODEA	Nov. 2015 DODEA	2016 SCDE	2014 SCDE
Air Force	3,826	4,766	7,763	7,308
Army	5,855	3,832		
Marines	1,980	2,275		
Navy	1,936	1,462		
TOTAL	**13,597**	**12,335**	**7,763**	**7,308**

Sources: SC Department of Education; DoDEA

As of January 2016, approximately 957 military-connected students have at least one parent who is deployed, including National Guard, Reserves and active duty military personnel. In addition, 26 military-connected students were reported to have a parent who was on active duty but died within the last year; another 39 military-connected students have a parent who was on active duty and wounded within the last year. Improved reporting of this data is needed so district and school staff can identify students who may need additional support services. As noted in the prior section, military-connected students live with perpetual challenges presented by frequent moves, parental and sibling deployments, and additional transitions that include reintegration and dealing with profoundly changed parents. The well-being of these children

[11] The Department of Defense Education Activity (DoDEA) is a civilian agency of the United States Department of Defense that manages all schools for military children and teenagers in the United States and also overseas at American military bases worldwide.

depends heavily on a network of supportive adults who are trained to identify early signs of emotional or physical challenge. Appendix A provides a list of resources for military-connected students and families.

Table 4
Military-Connected Students,
by Parental Military Branch and Deployment Status, January 2016

Military Connection	Number	Percent
National Guard - Not Deployed	835	8.72
Reserves - Not Deployed	716	7.48
National Guard - Deployed	164	1.71
Reserves - Deployed	94	0.98
Active Duty Military - Not Deployed	6,999	73.12
Active Duty Military - Deployed	699	7.30
Active Duty Military - Deceased in last year	26	0.27
Active Duty Military - Wounded in last year	39	0.41
Total	**9,572**	

Source: SC Department of Education

Of the 9,572 military-connected students (including active duty, Reserves and National Guard) reported by school districts to SCDE, approximately 90 percent of the students attend one of the ten school districts in Table 5 included below. South Carolina's largest military installations are located in Charleston, Beaufort, Richland and Sumter counties.

The Charleston Air Force Base and the Naval Weapons Station in Goose Creek comprise Joint Base Charleston (JBC). Joint Base Charleston is one of 12 DoD Joint Bases and is host to over 60 DoD and Federal agencies. The 628th Air Base Wing delivers installation support to a total force of over 90,000 Airmen, Sailors, Soldiers, Marines, Coast Guardsmen, civilians, dependents, and retirees across four installations including Charleston AFB and Naval Weapons Station Charleston. The base maintains $7.5 billion in base property and capital assets spanning three seaports, two civilian-military airfields, 38 miles of rail, and 22 miles of coastline total almost 24,000 acres. A few unique programs include operation of three locomotives critical to moving crucial munitions, MRAPS, and tanks, as well as participation in the Global Threat Reduction Initiative. Joint Base Charleston also offers unique installation

support missions including the Harbor Security Patrol, the Transportable Isolation System (TIS), and the Cooper River Dredging.[12]

Marine Corps Air Station Beaufort, home of the Marine Corps' Atlantic Coast fixed-wing, fighter-attack aircraft assets, is located in the heart of the South Carolina Lowcountry and is among the United States military's most important and most historically colorful installations. Consisting of some 6,900 acres 70 miles southwest of Charleston, South Carolina on Highway 21, the installation is home to seven Marine Corps F/A- 18 squadrons. Two additional Navy F/A-18 squadrons joined the Fightertown community in March 2000, strengthening the installation's economic contribution to the local area. Three versions of the F/A-18 Hornet are found aboard MCAS Beaufort, the F/A-18 Hornet A/Cs and the F/A-18 D.[13] The Marine Corps Recruit Depot is located nearby at Parris Island, which is also located in Beaufort County.

Both Fort Jackson and Shaw Air Force Base are located in the Midlands. Located in Richland County, Fort Jackson is the Army's main production center for Basic Combat Training. Approximately 50 percent of the Army's Basic Combat Training is completed at Fort Jackson, with more than 36,000 troops trained each year. Fort Jackson is home to the U.S. Army Soldier Support Institute, the Armed Forces Army Chaplaincy Center and School, the National Center for Credibility Assessment (formerly the Department of Defense Polygraph Institute), and the Drill Sergeant School, which trains all Active Duty and Reserve instructors.

Located about 40 miles east of Columbia, South Carolina, Shaw Air Force Base, S.C., is home to the 20th Fighter Wing, and headquarters, Ninth Air Force, U.S. Air Forces Central, Third Army and U.S. Army Central. The 20th FW is a unit of Ninth AF and Air Combat Command. With responsibility for the installation as host unit, the 20th FW supports nearly 30 associated units. The base is located on more than 3,569 acres within the city limits of Sumter, S.C. The base also has custodial responsibility for approximately 12,500 acres at Poinsett Electronic Combat Range Complex southwest of Sumter and for 23.5 leased acres at the Lake Wateree Recreational Area, 38 miles northwest of Sumter, near Camden, S.C.[14]

[12] Accessed at: http://www.charleston.af.mil/units/index.asp.
[13] Accessed at: http://www.beaufort.marines.mil/About.aspx.
[14] Accessed at: http://www.shaw.af.mil/AboutUs/FactSheets/Display/tabid/6148/Article/663885/shaw-air-force-base.aspx.

Table 5
School Districts with the Largest Military-Connected Student Populations, 2015 – 2016

District	Number	Percent of Students Statewide
Richland 2	1,985	20.7
Berkeley	1,627	17.0
Dorchester 2	1,600	16.7
Beaufort	1,055	11.0
Lexington 1	897	9.4
Sumter	651	6.8
Kershaw	310	3.2
Oconee	164	1.7
Charleston	162	1.7
SC Public Charter School District	140	1.5
Total	8,591	

Source: SC Department of Education

III. Military-Connected Student Performance

This section provides academic and attendance data for military-connected students for school year 2014-15 including:

- student achievement as measured by the ACT Aspire for third through eighth grades in English language arts and mathematics;
- student achievement as measured by SCPASS on science for students in grade four through eight;
- student achievement as measured by the End-Of-Course Evaluation Program (EOCEP);
- high school graduation rates; and
- student attendance.

Academic Data

The academic achievement of military-connected students was compared to the academic achievement of all students in South Carolina for students in third through eighth grades with ACT Aspire for reading and mathematics and SCPASS for science. For high school students, student performance on the South Carolina End-of-Course Evaluation Program (EOCEP) was considered.

Student Achievement in Grades Three through Eight

According to the company's website, ACT Aspire is a "vertically aligned system of summative and periodic assessments" linked to ACT's College and Career Readiness Benchmarks and predicts a students' performance on the ACT, the college readiness assessment.[15] The test measures student's readiness for college in five areas: English, math, reading, science, and writing.

Military-connected students performed significantly better than their peers in reading, mathematics and science. On average, an additional 9.6 percent of military-connected students tested as "ready" or "exceeding" on ACT Aspire in reading, and an additional 7.8 percent of military-connected students tested as "ready" or "exceeding" on ACT Aspire in mathematics. On the science test of SCPASS the performance gap was even larger; on average an additional 12 percent of military-connected students tested as "met" or better. In eighth grade, an additional 11.9 percent of military-connected students tested as "met" or better on SCPASS science and an additional 11.3 percent of military-connected students tested as "ready" or "exceeding" on ACT Aspire reading.

[15] http://www.act.org/content/act/en/products-and-services/act-aspire.html

Table 6
Performance of Military-Connected Students (MCS) Compared to Statewide Average on State-Administered Tests, 2014-15

	ACT Aspire Reading				ACT Aspire Mathematics			
Grade Level	Number MCS Tested	Percent MCS Ready or Exceeding	State Percent Ready or Exceeding	Difference	Number MCS Math	Percent Ready or Exceeding	State Percent Ready or Exceeding	Difference
3	825	41.6	31.8	9.8	827	69.2	58.0	11.2
4	786	43.4	32.8	10.6	787	59.3	49.4	9.9
5	669	42.6	33.6	9.0	668	58.2	48.2	10.0
6	737	45.9	36.9	9.0	736	57.1	53.2	3.9
7	677	44.6	36.9	7.7	677	40.9	36.4	7.5
8	657	58.0	46.7	11.3	655	39.2	32.0	7.2

Source: SC Department of Education

SCPASS Science				
Grade Level	Number MCS Science	Percent Met or Above	State Percent Met or Above	Difference
4	789	82.9	69.9	13.0
5	667	78.9	66.3	12.6
6	737	75.0	63.6	11.4
7	676	79.3	68.2	11.1
8	656	77.1	65.2	11.9

Source: SC Department of Education

Student Performance in End of Course Exams

In 2014, the mean score for military-connected students was typically six points higher than the mean score for all South Carolina students. On the EOCEP tests, military-connected students continue to outperform all students statewide. However, the gap between military-connected students and all students statewide may be narrowing. The average difference between military-connected students and all students statewide is 3.8 points.

Table 7
End-of-Course Assessment Performance of
Military Connected Students and All Students in South Carolina

Academic Year	Military Connected Students			All South Carolina Students	
	Number of Students	Mean	Letter Grade	Mean	Letter Grade
Algebra 1					
2013	398	83.5	C	78.2	C
2014	535	85.7	B	79.8	C
2015	668	85.7	B	82.6	C
English 1					
2013	350	81.3	C	75.3	D
2014	537	82.2	C	76.0	D
2015	636	83.6	C	79.4	C
Biology					
2013	310	84.2	C	78.1	C
2014	451	85.4	B	79.2	C
2015	580	86.5	B	82.3	B

Source: SC Department of Education

High School Graduation Rate

The federally-approved way of computing a graduation rate is to identify a cohort of students who were ninth grade students in a specific year and calculate the percentage of that cohort that graduates three years later. Students are removed from the cohort when they transfer to other degree-granting institutions or programs. Students who transfer into a district are added to the cohort.

For military-connected students this process was not possible because enrollment history of these students was not available. The EOC evaluation team could not know when students were initially in the ninth grade and could not document transfers into or out of a cohort of students who were initially in the ninth grade three years prior. Available data identifies students by grade level and graduation status. For students who were identified as being in

twelfth grade during the 2014-15 timeframe, the EOC evaluation team could identify: (1) those students who graduated, (2) those who received a certificate or did not graduate, and (3) those students who transferred to other degree-granting institutions and were removed from the graduation cohort. Based on this information, the graduation rates for military-connected students are included below.

Table 8
Four-Year High School Graduation Rates for Military-Connected Students (MCS) and Statewide Graduation Rates

Year	Total Number of MCS	Percent MCS Graduates	State Four Year Graduation Rate
2013	237	96.5	77.5
2014	309	97.4	80.1
2015	407	95.3	80.3

Attendance Data

Student attendance rates were computed using information provided by SCDE. Within any year, the number of students reported by school districts as military connected with an active duty parent is only 63 percent of the number reported by the Defense Manpower Data Center. The average percent of school days absent for all districts that reported military-connected students is 4.2 percent. There are nine districts with at least 30 reported military-connected students who average more than 4.2 percent absent days; these districts are shaded in the table below. At 5.6 percent, Spartanburg 7 reports the highest absence rate and Florence 1 has the lowest absence rate.

Table 9
Percent of Days Absent in School Districts with at least 30 Military-Connected Students

District	Number of MCS Students	Percent of Days Absent
Average for All Districts	9571	4.2
Spartanburg 7	30	5.6
Colleton	65	5.1
Greenville	58	5.0
Aiken	83	4.9
Pickens	127	4.8
Beaufort	1054	4.8
Charleston	162	4.5
Kershaw	310	4.4
Dorchester	1600	4.3
Berkeley	1627	4.2
Hampton 1	74	4.1
Lexington 1	897	4.0
Sumter	651	3.9
Oconee	164	3.9
Edgefield	62	3.8
Richland 2	1985	3.8
Richland 1	89	3.7
Horry	56	3.4
SC Public Charter School District	140	3.1

Appendix A
Resources for Military-Connected Students and Families

The Department of Defense Education Activity (DoDEA) is a civilian agency of the United States Department of Defense that manages all schools for military children and teenagers in the United States and also overseas at American military bases worldwide. This information is also helpful for local school districts to understand the needs of students and how to support them in a comprehensive manner. DoDEA's Website "Keeping Students at the Center" http://slmodules.dodea.edu/.

School Liaison Officers serve as a primary point of contact for students and their families transitioning to new communities and schools. They are also a resource for schools and school districts. To view a list of school liaison officers by branch, go to http://www.dodea.edu/Partnership/schoolLiaisonOfficers.cfm. Below is a list of additional information regarding school liaison officers by installation:

Fort Jackson: http://www.fortjacksonmwr.com/school_liaison and http://www.militaryinstallations.dod.mil/MOS/f?p=MI:CONTENT:0:::P4_INST_ID,P4_CONTENT_TITLE,P4_CONTENT_EKMT_ID,P4_CONTENT_DIRECTORY:4210,Education,30.90.120.30.30.0.0.0.0,12

Beaufort: http://www.mccs-sc.com/mil-fam/slp.shtml and http://www.militaryinstallations.dod.mil/MOS/f?p=MI:CONTENT:0:::P4_INST_ID,P4_CONTENT_TITLE,P4_CONTENT_EKMT_ID,P4_CONTENT_DIRECTORY:4240,Education,30.90.120.30.30.0.0.0.0,12.

Joint Base Charleston: http://www.charleston.af.mil/news/story.asp?id=123469714 and http://www.militaryinstallations.dod.mil/MOS/f?p=MI:CONTENT:0:::P4_INST_ID,P4_CONTENT_TITLE,P4_CONTENT_EKMT_ID,P4_CONTENT_DIRECTORY:7195,Education,30.90.120.30.30.0.0.0.0,12.

Shaw Air Force Base: http://www.militaryinstallations.dod.mil/MOS/f?p=MI:CONTENT:0:::P4_INST_ID,P4_CONTENT_TITLE,P4_CONTENT_EKMT_ID,P4_CONTENT_DIRECTORY:4270,Education,30.90.120.30.30.0.0.0.0,12

Military Impacted School Association is a national organization of school superintendents. MISA supports school districts with a high concentration of military children by providing detailed, comprehensive information regarding impact aid and resources for families and schools.

http://militaryimpactedschoolsassociation.org/

The **Military Interstate Children's Compact Commission (MIC3)** provides consistent policy in every school district and in every state that voluntarily joins MIC3. MIC3 addresses key

educational transition issues such as enrollment, placement, attendance, eligibility and graduation.

http://www.mic3.net

For more information about South Carolina's role in MIC3, go to http://ed.sc.gov/agency/lpa/mic.cfm.

South Carolina Military Kids is part of the National OMK initiative designated to provide support to the children and youth of families that are impacted by global contingency operations. This includes those served by Army installations, Air Force, Navy and Marine bases, and those families, children and youth who are geographically dispersed.

https://www.clemson.edu/extension/4h/kids_families/militarypartners/

The **Military Child Education Coalition (MCEC)** focuses on ensuring quality educational opportunities for all military children affected by mobility, family separation, and transition. A 501(c)(3) non-profit, world-wide organization, the MCEC performs research, develops resources, conducts professional institutes and conferences, and develops and publishes resources for all constituencies.

www.militarychild.org

Military OneSource is a confidential Department of Defense-funded program providing comprehensive information on every aspect of military life at no cost to active duty, National Guard, and reserve members, and their families.

Information includes, but is not limited to, deployment, reunion, relationships, grief, spouse employment and education, parenting and childhood services. It is a virtual extension to installation services.

The program also provides free resources to schools, including books and videos with relevant topics that help students cope with divorce and deployment.

www.militaryonesource.mil

South Carolina Programs

The **International Baccalaureate** Program helps students develop skills to create a better and peaceful world through intercultural understanding and respect. For more information, including a list of South Carolina schools participating in the IB Program, go to http://ed.sc.gov/agency/programs-services/127/.

Four-year-old kindergarten is available in the state and is offered in public schools and private child care centers. State-funded prekindergarten for four-year-olds serves children in the "most

at-risk” category, where family income falls 185% below poverty level or the family is Medicaid eligible. Families may also be eligible for other services such as Even Start, Head Start, state-funded family literacy programs, Social Security, food stamps, Medicaid or temporary assistance to needy families (TANF).

Children also qualify in case of a documented developmental delay, an Individual Education Plan (IEP) requiring pre-kindergarten, incarceration of a parent, placement in a foster home, or a child who is homeless. Documentation of family or child “most at-risk” conditions must be kept on file for review. Children who participate in free and reduced meal programs at the center/school they attend may also qualify, if income eligibility is verified on each child and records are kept on file for review.

Some districts use local funds to serve children who are not in the “at risk” category. Several districts serve all children who request services. A few districts charge a fee for non-qualifying children, but state regulations prohibit any fees for “at risk” children.

State law says that “students may enter kindergarten in the public schools of this State if they will attain the age of four on or before September first of the applicable school year.”

Appendix B
Impact Aid Section of Every Student Succeeds Act

TITLE VII—IMPACT AID
SEC. 7001. GENERAL PROVISIONS.

(a) IMPACT AID IMPROVEMENT ACT OF 2012.—Section 563(c) of National Defense Authorization Act for Fiscal Year 2013 (Public Law 112–239; 126 Stat. 1748; 20 U.S.C. 6301 note) (also known as the ''Impact Aid Improvement Act of 2012''), as amended by section 563 of division A of Public Law 113–291, is amended—(1) by striking paragraphs (1) and (4); and(2) by redesignating paragraphs (2) and (3), as paragraphs(1) and (2), respectively.

(b) REPEAL.—Section 309 of division H of the Consolidated Appropriations Act, 2014 (Public Law 113–76; 20 .S.C. 7702 note) is repealed.

(c) TITLE VII REDESIGNATIONS.—Title VIII (20 U.S.C. 7701 et seq.) is redesignated as title VII and further amended—(1) by redesignating sections 8001 through 8005 as sections 7001 through 7005, respectively; and
(2) by redesignating sections 8007 through 8014 as sections 7007 through 7014, respectively.

(d) CONFORMING AMENDMENTS.—Title VII (as redesignated by subsection (c) of this section) is further amended—
(1) by striking ''section 8002'' each place it appears and inserting ''section 7002''; (2) by striking ''section 8003'' each place it appears and inserting ''section 7003''; (3) by striking ''section 8003(a)(1)'' each place it appears
and inserting ''section 7003(a)(1)''; (4) by striking ''section 8003(a)(1)(C)'' each place it appears and inserting ''section 7003(a)(1)(C)''; (5) by striking ''section 8003(a)(2)'' each place it appears and inserting ''section 7003(a)(2)''; (6) by striking ''section 8003(b)'' each place it appears and inserting ''section 7003(b)'';
(7) by striking ''section 8003(b)(1)'' each place it appears and inserting ''section 7003(b)(1)''; (8) by striking ''section 8003(b)(2)'' each place it appears and inserting ''section 7003(b)(2)''; S. 1177—274 (9) by striking ''section 8014(a)'' each place it appears and inserting ''section 7014(a)''; (10) by striking ''section 8014(b)'' each place it appears and inserting ''section 7014(b)''; and (11) by striking ''section 8014(e)'' each place it appears
and inserting ''section 7014(d)''.

SEC. 7002. PURPOSE.
Section 7001, as redesignated by section 7001 of this Act, is amended in the matter preceding paragraph (1), by striking ''challenging State standards'' and inserting ''the same challenging State academic standards''.

SEC. 7003. PAYMENTS RELATING TO FEDERAL ACQUISITION OF REAL PROPERTY.
Section 7002, as redesignated and amended by section 7001 of this Act, is further amended—(1) in subsection (a)(1)(C), by striking the matter preceding clause (i) and inserting the following: ''(C) had an assessed value according to original records (including facsimiles or other reproductions of those records) documenting the assessed value of such property (determined as of the time or times when so acquired) prepared by the local officials referred to in subsection

(b)(3) or, when such original records are not available due to unintentional destruction (such as natural disaster, fire, flooding, pest infestation, or deterioration due to age), other records, including Federal agency records, local historical records, or other records that the Secretary determines to be appropriate and reliable, aggregating 10 percent or more of the assessed value of—''; (2) in subsection (b)— (A) in paragraph (1)(C) by striking ''section 8003(b)(1)(C)'' and inserting ''section 7003(b)(1)(C)''; (B) in paragraph (3), by striking subparagraph (B) and inserting the following: ''(B) SPECIAL RULE.—In the case of Federal property
eligible under this section that is within the boundaries of 2 or more local educational agencies that are eligible
under this section, any of such agencies may ask the Secretary to calculate (and the Secretary shall calculate) the
taxable value of the eligible Federal property that is within its boundaries by— ''(i) first calculating the per-acre value of the eligible Federal property separately for each eligible local educational agency that shared the Federal property, as provided in subparagraph (A)(ii); ''(ii) then averaging the resulting per-acre values of the eligible Federal property from each eligible local educational agency that shares the Federal property; and
''(iii) then applying the average per-acre value to determine the total taxable value of the eligible Federal
property under subparagraph (A)(iii) for the requesting local educational agency.''; (3) in subsection (e)(2), by adding at the end the following:
''For each fiscal year beginning on or after the date of enactment of the Every Student Succeeds Act, the Secretary shall treat S. 1177—275 local educational agencies chartered in 1871 having more than 70 percent of the county in Federal ownership as meeting the eligibility requirements of subparagraphs (A) and (C) of subsection (a)(1).'';
(4) by striking subsection (f) and inserting the following: ''(f) SPECIAL RULE.—For each fiscal year beginning on or after the date of enactment of the Every Student Succeeds Act, a local educational agency shall be deemed to meet the requirements of subsection (a)(1)(C) if the agency was eligible under paragraph (1) or (3) of section 8002(f) as such section was in effect on the day before the date of enactment of the Every Student Succeeds Act.'';
(5) by striking subsection (g) and inserting the following: ''(g) FORMER DISTRICTS.— ''(1) CONSOLIDATIONS.—For fiscal year 2006 and each succeeding fiscal year, if a local educational agency described in paragraph (2) is formed at any time after 1938 by the consolidation of 2 or more former school districts, the local educational agency may elect to have the Secretary determine its eligibility for assistance under this section for any fiscal year on the
basis of 1 or more of those former districts, as designated by the local educational agency. ''(2) ELIGIBLE LOCAL EDUCATIONAL AGENCIES.—A local educational agency referred to in paragraph (1) is— ''(A) any local educational agency that, for fiscal year 1994 or any preceding fiscal year, applied, and was determined to be eligible under, section 2(c) of the Act of September 30, 1950 (Public Law 874, 81st Congress) as that section was in effect for that fiscal year; or ''(B) a local educational agency— ''(i) that was formed by the consolidation of 2 or more districts, at least 1 of which was eligible for assistance under this section for the fiscal year preceding the year of the consolidation; and ''(ii) which includes the designation referred to in paragraph (1) in its application under section 7005 for a fiscal year beginning on or after the date of enactment of the Every Student Succeeds Act or any
timely amendment to such application. ''(3) AMOUNT.—A local educational agency eligible under paragraph (1) shall receive a foundation payment as provided for under subparagraphs (A) and (B) of subsection (h)(1), except

that the foundation payment shall be calculated based on the most recent payment received by the local educational agency based on its status prior to consolidation.''; (6) in subsection (h)(4), by striking ''For each local educational
agency that received a payment under this section for fiscal year 2010 through the fiscal year in which the Impact
Aid Improvement Act of 2012 is enacted'' and inserting ''For each local educational agency that received a payment under this section for fiscal year 2010 or any succeeding fiscal year''; (7) by repealing subsections (k) and (m);
(8) by redesignating subsection (l) as subsection (j); (9) in subsection (j) (as redesignated by paragraph (8)),
by striking ''(h)(4)(B)'' and inserting ''(h)(2)''; (10) by redesignating subsection (n) as subsection (k); and
S. 1177—276 (11) in subsection (k)(1) (as redesignated by paragraph (10)), by striking ''section 8013(5)(C)(iii)'' and inserting ''section 7013(5)(C)(iii)''.

SEC. 7004. PAYMENTS FOR ELIGIBLE FEDERALLY CONNECTED CHILDREN.
Section 7003, as redesignated and amended by section 7001 of this Act, is further amended— (1) in subsection (a)(5)(A), by striking ''to be children'' and all that follows through the period at the end and inserting ''or under lease of off-base property under subchapter IV of chapter 169 of title 10, United States Code, to be children
described under paragraph (1)(B), if the property described is— ''(i) within the fenced security perimeter of the
military facility; or ''(ii) attached to, and under any type of force protection agreement with, the military installation upon which such housing is situated.''; (2) in subsection (b)—(A) in paragraph (1)—(i) by striking subparagraph (E); and (ii) by redesignating subparagraphs (F) and (G) as subparagraphs (E) and (F), respectively; (B) in paragraph (2), by striking subparagraphs (B) through (H) and inserting the following: ''(B) ELIGIBILITY FOR HEAVILY IMPACTED LOCAL EDUCATIONAL AGENCIES.—''(i) IN GENERAL.—A heavily impacted local educational agency is eligible to receive a basic support payment under subparagraph (A) with respect to a number of children determined under subsection (a)(1) if the agency—''(I) is a local educational agency—''(aa) whose boundaries are the same as
a Federal military installation or an island property designated by the Secretary of the Interior to be property that is held in trust by the Federal Government; and ''(bb) that has no taxing authority; ''(II) is a local educational agency that—''(aa) has an enrollment of children described in subsection (a)(1) that constitutes a percentage of the total student enrollment of the agency that is not less than 45 percent; ''(bb) has a per-pupil expenditure that is
less than—''(AA) for an agency that has a total student enrollment of 500 or more students, 125 percent of the average per-pupil expenditure of the State in which the agency is located; or ''(BB) for any agency that has a total
student enrollment of less than 500 students, 150 percent of the average per-pupil expenditure of the State in which the agency is located or the average per-pupil S. 1177—277 expenditure of 3 or more comparable local educational agencies in the State in which the agency is located; and ''(cc) is an agency that has a tax rate for general fund purposes that is not less than 95 percent of the average tax rate for general fund purposes of comparable local educational agencies in the State; ''(III) is a local educational agency that—''(aa) has a tax rate for general fund purposes which is not less than 125 percent of the average tax rate for general fund purposes for comparable local educational agencies in the State; and ''(bb)(AA) has an enrollment of children described in subsection (a)(1) that constitutes a percentage of the total student enrollment of the agency that is not less than 30 percent; or ''(BB) has an enrollment of

children described in subsection (a)(1) that constitutes a percentage of the total student enrollment
of the agency that is not less than 20 percent, and for the 3 fiscal years preceding the fiscal year for which the determination is made, the average enrollment of children who are not described in subsection (a)(1) and who are
eligible for a free or reduced price lunch under the Richard B. Russell National School Lunch Act constitutes a percentage of the total student enrollment of the agency that is not less than 65 percent; ''(IV) is a local educational agency that has a total student enrollment of not less than 25,000 students, of which—''(aa) not less than 50 percent are children described in subsection (a)(1); and ''(bb) not less than 5,000 of such children are children described in subparagraphs (A) and (B) of subsection (a)(1); or ''(V) is a local educational agency that—''(aa) has an enrollment of children described in subsection (a)(1) including, for purposes of determining eligibility, those children described in subparagraphs (F) and (G) of such subsection, that is not less than 35 percent of the total student enrollment of the
agency; ''(bb) has a per-pupil expenditure described in subclause (II)(bb) (except that a local educational agency with a total student enrollment of less than 350 students shall be deemed to have satisfied such per-pupil expenditure requirement) and has a tax rate for general fund purposes which is not less than 95 percent of the average tax rate for general fund purposes for comparable local educational agencies in the State; and ''(cc) was eligible to receive assistance under subparagraph (A) for fiscal year 2001. ''(ii) LOSS OF ELIGIBILITY.—''(I) IN GENERAL.—Subject to subclause (II), a heavily impacted local educational agency that met the requirements of clause (i) for a fiscal year shall be ineligible to receive a basic support payment under subparagraph (A) if the agency fails
to meet the requirements of clause (i) for a subsequent fiscal year, except that such agency shall continue to receive a basic support payment under this paragraph for the fiscal year for which the ineligibility determination is made.
''(II) LOSS OF ELIGIBILITY DUE TO FALLINGBELOW 95 PERCENT OF THE AVERAGE TAX RATE FOR GENERAL FUND PURPOSES.—In the case of a heavily impacted local educational agency described in subclause (II) or (V) of clause (i) that is eligible to receive a basic support payment under subparagraph (A), but that has had, for 2 consecutive fiscal years, a tax rate for general fund purposes that falls below 95 percent of the average tax rate for general fund purposes of comparable local educational agencies in the State, such agency shall be determined to be ineligible under clause
(i) and ineligible to receive a basic support payment under subparagraph (A) for each fiscal year succeeding such 2 consecutive fiscal years for which the agency has such a tax rate for general fund purposes, and until the fiscal year for which the agency resumes such eligibility in accordance with clause (iii).
''(III) TAKEN OVER BY STATE BOARD OF EDUCATION.—
In the case of a heavily impacted local educational agency that is eligible to receive a basic support payment under subparagraph (A), but that has been taken over by a State board of education in any 2 previous years, such agency
shall be deemed to maintain heavily impacted status for 2 fiscal years following the date of enactment
of the Every Student Succeeds Act.
''(iii) RESUMPTION OF ELIGIBILITY.—A heavily impacted local educational agency described in clause (i) that becomes ineligible under such clause for 1 or more fiscal years may resume eligibility for a basic support payment under this paragraph for a subsequent fiscal year only if the agency meets the requirements of clause (i) for that subsequent fiscal year, except that

such agency shall not receive a basic support payment under this paragraph until the fiscal year succeeding the fiscal year for which the eligibility determination is made.
‘‘(C) MAXIMUM AMOUNT FOR HEAVILY IMPACTED LOCAL EDUCATIONAL AGENCIES.—S. 1177—279
‘‘(i) IN GENERAL.—Except as provided in subparagraph (D), the maximum amount that a heavily impacted local educational agency is eligible to receive under this paragraph for any fiscal year is the sum of the total weighted student units, as computed under subsection (a)(2) and subject to clause (ii), multiplied by the greater of—
‘‘(I) four-fifths of the average per-pupil expenditure of the State in which the local educational agency is located for the third fiscal year preceding the fiscal year for which the determination is made; or ‘‘(II) four-fifths of the average per-pupil expenditure of all of the States for the third fiscal year preceding the fiscal year for which the determination is made. ‘‘(ii) CALCULATION OF WEIGHTED STUDENT UNITS.—‘‘(I) IN GENERAL.—‘‘(aa) PERCENTAGE ENROLLMENT.—For a local educational agency in which 35 percent or more of the total student enrollment of the schools of the agency are children described in subparagraph (D) or (E) (or a combination thereof) of subsection (a)(1), and that has an enrollment of children described in subparagraph (A), (B), or (C) of such subsection equal to at least 10 percent of the agency’s total enrollment, the Secretary shall calculate the weighted student units of those children described in subparagraph (D) or (E)of such subsection by multiplying the number
of such children by a factor of 0.55. ‘‘(bb) EXCEPTION.—Notwithstanding item (aa), a local educational agency that received a payment under this paragraph for fiscal year 2013 shall not be required to have an enrollment of children described in subparagraph (A), (B), or (C) of subsection (a)(1) equal to at least 10 percent of the agency’s total enrollment and shall be eligible for the student weight as provided for in item (aa).
‘‘(II) ENROLLMENT OF 100 OR FEWER CHILDREN.—
For a local educational agency that has an enrollment of 100 or fewer children described in subsection (a)(1), the Secretary shall calculate the total number of weighted student units for purposes of subsection (a)(2) by multiplying the number of such children by a factor of 1.75.
‘‘(III) ENROLLMENT OF MORE THAN 100 CHILDREN BUT LESS THAN 1000.—For a local educational agency that is not described under subparagraph (B)(i)(I) and has an enrollment of more than 100 but not more than 1,000 children described in subsection (a)(1), the Secretary shall calculate the total number of weighted student units for purposes of subsection (a)(2) by multiplying the number of such children by a factor of 1.25.
‘‘(D) MAXIMUM AMOUNT FOR LARGE HEAVILY IMPACTED LOCAL EDUCATIONAL AGENCIES.—
‘‘(i) IN GENERAL.—‘‘(I) FORMULA.—Subject to clause (ii), the maximum amount that a heavily impacted local educational agency described in subclause (II) is eligible to receive under this paragraph for any fiscal year shall be determined in accordance with the formula described in paragraph (1)(C).
‘‘(II) HEAVILY IMPACTED LOCAL EDUCATIONAL AGENCY.—A heavily impacted local educational agency described in this subclause is a local educational agency that has a total student enrollment of not less than 25,000 students, of which not less than 50 percent are children described in subsection (a)(1) and not less than 5,000 of such children are children described in subparagraphs (A) and (B) of subsection (a)(1). ‘‘(ii) FACTOR.—For purposes of calculating the maximum amount described in clause (i), the factor used in determining the weighted student units under subsection (a)(2) with respect to children described in subparagraphs (A) and (B) of subsection (a)(1) shall be 1.35.
‘‘(E) DATA.—For purposes of providing assistance under this paragraph, the Secretary shall use student, revenue,

expenditure, and tax data from the third fiscal year preceding the fiscal year for which the local educational agency
is applying for assistance under this paragraph.
''(F) DETERMINATION OF AVERAGE TAX RATES FOR GENERAL FUND PURPOSES.—
''(i) IN GENERAL.—Except as provided in clause (ii), for the purpose of determining the average tax rates for general fund purposes for local educational agencies in a State under this paragraph, the Secretary shall use either—
''(I) the average tax rate for general fund purposes for comparable local educational agencies, as determined by the Secretary in regulations; or ''(II) the average tax rate of all the local educational agencies in the State. ''(ii) FISCAL YEARS 2010–2015.—''(I) IN GENERAL.—For fiscal years 2010 through 2015, any local educational agency that
was found ineligible to receive a payment under subparagraph (A) because the Secretary determined that it failed to meet the average tax rate requirement for general fund purposes in subparagraph (B)(i)(II)(cc), shall be considered to have met that requirement, if its State determined, through an alternate calculation of average tax rates for
general fund purposes, that such local educational agency met that requirement. ''(II) SUBSEQUENT FISCAL YEARS AFTER 2015.— For any succeeding fiscal year after 2015, any local educational agency identified in subclause
S. 1177—281 (I) may continue to have its State use that alternate methodology to calculate whether the average
tax rate requirement for general fund purposes under subparagraph (B)(i)(II)(cc) is met.
''(III) AVAILABILITY OF FUNDS.—Notwithstanding any other provision of law limiting the period during which the Secretary may obligate funds appropriated for any fiscal year after 2012, the Secretary shall reserve a total of $14,000,000 from funds that remain unobligated under this section from fiscal years 2015 or 2016 in order to make payments under this clause for fiscal years 2011 through 2014.
''(G) ELIGIBILITY FOR HEAVILY IMPACTED LOCAL EDUCATIONAL AGENCIES AFFECTED BY PRIVATIZATION OF MILITARY
HOUSING.—
''(i) ELIGIBILITY.—For any fiscal year, a heavily impacted local educational agency that received a basic support payment under this paragraph for the prior fiscal year, but is ineligible for such payment for the current fiscal year under subparagraph (B) due to the conversion of military housing units to private housing described in clause (iii), or as the direct result of base realignment and closure or modularization as determined by the Secretary of Defense and force structure change or force relocation, shall be deemed to meet the eligibility requirements under subparagraph (B) for the period during which the housing units are undergoing such conversion or during such time as activities associated with base closure and realignment, modularization, force structure change, or force relocation are ongoing.
''(ii) AMOUNT OF PAYMENT.—The amount of a payment to a heavily impacted local educational agency for a fiscal year by reason of the application of clause (i), and calculated in accordance with subparagraph (C) or (D), as the case may be, shall be based on the number of children in average daily attendance in the schools of such agency for the fiscal year and under the same provisions of subparagraph (C) or (D) under which the agency was paid during the prior fiscal year.
''(iii) CONVERSION OF MILITARY HOUSING UNITS TO PRIVATE HOUSING DESCRIBED.—For purposes of clause (i), 'conversion of military housing units to private housing' means the conversion of military housing units to private housing units pursuant to subchapter IV of chapter 169 of title 10, United States Code, or pursuant to any other related provision of law.'';
(C) in paragraph (3)—(i) in subparagraph (B), by striking clause (iii) and inserting the following:

‘‘(iii) In the case of a local educational agency providing a free public education to students enrolled in kindergarten through grade 12, that enrolls students described in subparagraphs (A), (B), and (D) of subsection (a)(1) only in grades 9 through 12, and that received a final payment for fiscal year 2009 calculated under section 8003(b)(3) (as such section was in effect on the day before the date of enactment of the Every Student Succeeds Act) for students in grades 9 through 12, the Secretary shall, in calculating the agency’s payment, consider only that portion of such agency’s total enrollment of students in grades 9 through 12 when calculating the percentage under clause (i)(I) and only that portion of the total current expenditures attributed to the operation of grades 9 through 12 in such agency when calculating the percentage under clause (i)(II).’’; (ii) in subparagraph (C), by striking ‘‘subparagraph (D) or (E) of paragraph (2), as the case may be’’ and inserting ‘‘subparagraph (C) or (D) of paragraph (2), as the case may be’’; and (iii) by striking subparagraph (D) and inserting the following:

‘‘(D) RATABLE DISTRIBUTION.—For fiscal years described in subparagraph (A), for which the sums available exceed the amount required to pay each local educational agency 100 percent of its threshold payment, the Secretary shall distribute the excess sums to each eligible local educational agency that has not received its full amount computed under paragraphs (1) or (2) (as the case may be) by multiplying— ‘‘(i) a percentage, the denominator of which is the difference between the full amount computed under paragraph (1) or (2) (as the case may be) for all local educational agencies and the amount of the threshold payment (as calculated under subparagraphs (B) and (C)) of all local educational agencies, and the numerator of which is the aggregate of the excess sums, by ‘‘(ii) the difference between the full amount computed under paragraph (1) or (2) (as the case may be) for the agency and the amount of the threshold payment (as calculated under subparagraphs (B) or (C)) of the agency, except that no local educational agency shall receive more than 100 percent of the maximum payment calculated under subparagraph (C) or (D) of paragraph (2).

‘‘(E) INSUFFICIENT PAYMENTS.—For each fiscal year described in subparagraph (A) for which the sums appropriated are insufficient to pay each local educational agency all of the local educational agency’s threshold payment described in subparagraph (B), the Secretary shall ratably reduce the payment to each local educational agency under this paragraph.

‘‘(F) INCREASES.—‘‘(i) INCREASES BASED ON INSUFFICIENT FUNDS.—If additional funds become available under 7014(b) for making payments under paragraphs (1) and (2) and those funds are not sufficient to increase each local educational agency’s threshold payment above 100 percent of its threshold payment described in subparagraph (B), payments that were reduced under subparagraph (E) shall be increased by the Secretary on the same basis as such payments were reduced. ‘‘(ii) INCREASES BASED ON SUFFICIENT FUNDS.—If additional funds become available under section 7014(b) for making payments under paragraphs (1) and (2) and those funds are sufficient to increase each local educational agency’s threshold payment above 100 percent of its threshold payment described in subparagraph (B), the payment for each local educational agency shall be 100 percent of its threshold payment.

The Secretary shall then distribute the excess sums to each eligible local educational agency in accordance with subparagraph (D).

‘‘(G) PROVISION OF TAX RATE AND RESULTING PERCENTAGE.— As soon as practicable following the payment of funds under paragraph (2) to an eligible local educational agency, the Secretary shall provide the local educational agency with a description of—‘‘(i) the tax rate of the local educational agency; and ‘‘(ii) the percentage such tax rate represents of the average tax rate for general fund purposes of comparable local educational agencies in the State as determined under subclauses (II)(cc), III(aa), or (V)(bb) of paragraph (2)(B)(i) (as the case may be).’’; and (D) in paragraph (4)—(i) in subparagraph (A), by striking ‘‘through (D)’’and inserting ‘‘and (C)’’; and (ii) in subparagraph (B), by striking ‘‘subparagraph (D) or (E)’’ and inserting

‘‘subparagraph (C) or (D)’’; (3) in subsection (c), by striking paragraph (2) and inserting the following: ‘‘(2) EXCEPTION.—Calculation of payments for a local educational agency shall be based on data from the fiscal year for which the agency is making an application for payment if such agency—‘‘(A) is newly established by a State, for the first year of operation of such agency only; ‘‘(B) was eligible to receive a payment under this section for the previous fiscal year and has had an overall increase in enrollment (as determined by the Secretary in consultation with the Secretary of Defense, the Secretary of the Interior, or the heads of other Federal agencies)—‘‘(i)(I) of not less than 10 percent of children described in—‘‘(aa) subparagraph (A), (B), (C), or (D) of subsection (a)(1); or ‘‘(bb) subparagraphs (F) and (G) of subsection (a)(1), but only to the extent that such children are civilian dependents of employees of the Department of Defense or the Department of the Interior; or ‘‘(II) of not less than 100 of such children; and
‘‘(ii) that is the direct result of closure or realignment of military installations under the base closure process or the relocation of members of the Armed Forces and civilian employees of the Department of Defense as part of the force structure changes or movements of units or personnel between military installations or because of actions initiated by the Secretary of the Interior or the head of another Federal agency; or ‘‘(C) was eligible to receive a payment under this section for the previous fiscal year and has had an increase in enrollment (as determined by the Secretary)—‘‘(i) of not less than 10 percent of children described in subsection (a)(1) or not less than 100 of such children; and ‘‘(ii) that is the direct result of the closure of a local educational agency that received a payment under subsection (b)(1) or (b)(2) for the previous fiscal year.’’; (4) in subsection (d)(1), by striking ‘‘section 8014(c)’’ and inserting ‘‘section 7014(c)’’; (5) in subsection (e)—(A) by redesignating paragraph (3) as paragraph (4); (B) by striking paragraphs (1) and (2) and inserting the following:
‘‘(1) IN GENERAL.—In the case of any local educational agency eligible to receive a payment under subsection (b) whose calculated payment amount for a fiscal year is reduced by 20 percent, as compared to the amount received for the previous fiscal year, the Secretary shall pay the local educational agency, for the year of the reduction and the following 2 years, the amount determined under paragraph (2).
‘‘(2) AMOUNT OF REDUCTION.—Subject to paragraph (3), A local educational agency described in paragraph (1) shall receive—‘‘(A) for the first year for which the reduced payment is determined, an amount that is not less than 90 percent of the total amount that the local educational agency received under subsection (b) for the previous fiscal year; ‘‘(B) for the second year following such reduction, an amount that is not less than 85 percent of the total amount that the local educational agency received under subparagraph (A); and ‘‘(C) for the third year following such reduction, an amount that is not less than 80 percent of the total amount that the local educational agency received under subparagraph (B).
‘‘(3) SPECIAL RULE.—For any fiscal year for which a local educational agency would receive a payment under subsection (b) in excess of the amount determined under paragraph (2), the payment received by the local educational agency for such fiscal year shall be calculated under paragraph (1) or (2) of subsection (b).’’; and
(6) by striking subsection (g).

SEC. 7005. POLICIES AND PROCEDURES RELATING TO CHILDREN RESIDING ON INDIAN LANDS.
Section 7004(e)(9), as redesignated and amended by section 7001 of this Act, is further amended by striking ‘‘Affairs’’ both places the term appears and inserting ‘‘Education’’.

SEC. 7006. APPLICATION FOR PAYMENTS UNDER SECTIONS 7002 AND 7003.

Section 7005, as redesignated and amended by section 7001 of this Act, is further amended—(1) in the section heading, by striking ‘‘**8002 AND 8003**’’and inserting ‘‘**7002 AND 7003**’’; (2) by striking ‘‘or 8003’’ each place it appears and inserting ‘‘or 7003’’; (3) in subsection (b)—(A) in the matter preceding paragraph (1), by striking ‘‘, and shall contain such information,’’; and (B) by striking ‘‘section 8004’’ and inserting ‘‘section 7004’’; and (4) in subsection (d)(2), by striking ‘‘section 8003(e)’’ and inserting ‘‘section 7003(e)’’;

SEC. 7007. CONSTRUCTION.
Section 7007, as redesignated and amended by section 7001 of this Act, is further amended—(1) in subsection (a)—
(A) in paragraph (3)(A)(i)—(i) by redesignating the first subclause (II) as subclause (I); (ii) in subclause (II), by striking ‘‘section 8008(a)’’ and inserting ‘‘section 7008(a)’’; and (B) in paragraph (4), by striking ‘‘section 8013(3)’’ and inserting ‘‘section 7013(3)’’; and (2) in subsection (b)—(A) in paragraph (3)(C)(i)(I), by adding at the end thefollowing: ‘‘(cc) Not less than 10 percent of the property acreage in the agency is exempt from State and local taxation under Federal law.’’; and (B) in paragraph (6)—(i) in the matter preceding subparagraph (A), by striking ‘‘, in such manner, and accompanied by such information’’ and inserting ‘‘and in such manner’’; (ii) in subparagraph (A), by inserting before the period at the end the following: ‘‘, and containing such additional information as may be necessary to meet any award criteria for a grant under this subsection as provided by any other Act’’; and (iii) by striking subparagraph (F).

SEC. 7008. FACILITIES.
Section 7008(a), as redesignated by section 7001 of this Act, is amended by striking ‘‘section 8014(f)’’ and inserting ‘‘section 7014(e)’’.

SEC. 7009. STATE CONSIDERATION OF PAYMENTS IN PROVIDING STATE AID.
Section 7009, as redesignated and amended by section 7001 of this Act, is further amended—(1) by striking ‘‘section 8011(a)’’ each place it appears and inserting ‘‘section 7011(a)’’; (2) in subsection (b)(1)—(A) by striking ‘‘or 8003(b)’’ and inserting ‘‘or 7003(b)’’; and (B) by striking ‘‘section 8003(a)(2)(B)’’ and inserting ‘‘section 7003(a)(2)(B)’’; and (3) in subsection (c)(1)(B), by striking ‘‘and contain the information’’ and inserting ‘‘that’’ after ‘‘form’’.

SEC. 7010. FEDERAL ADMINISTRATION.
Section 7010, as redesignated and amended by section 7001 of this Act, is further amended—(1) in subsection (c)— (A) in paragraph (1), in the paragraph heading, by striking ‘‘8003(a)(1)’’ and inserting ‘‘7003(a)(1)’’; (B) in paragraph (2)(D), by striking ‘‘section 8009(b)’’ and inserting ‘‘section 7009(b)’’; and (2) in subsection (d)(2), by striking ‘‘section 8014’’ and inserting ‘‘section 7014’’.

SEC. 7011. ADMINISTRATIVE HEARINGS AND JUDICIAL REVIEW.
Section 7011(a), as redesignated by section 7001 of this Act, is amended by striking ‘‘or under the Act’’ and all that follows through ‘‘1994)’’.

SEC. 7012. DEFINITIONS.
Section 7013, as redesignated by section 7001 of this Act, is amended—(1) in paragraph (1), by striking ‘‘and Marine Corps’’ and inserting ‘‘Marine Corps, and Coast Guard’’; (2) in paragraph (4), by striking ‘‘and title VI’’; (3) in paragraph (5)(A)—(A) in clause (ii), by striking subclause (III) and inserting the following: ‘‘(III) conveyed at any time under the Alaska Native Claims Settlement Act to a Native individual, Native group, or village or regional corporation (including

single family occupancy properties that may have been subsequently sold or leased to a third party), except that property that is conveyed under such Act—''(aa) that is not taxed is, for the purposes of this paragraph, considered tax-exempt due to Federal law; and ''(bb) is considered Federal property for the purpose of this paragraph if the property is located within a Regional Educational Attendance Area that has no taxing power;''; and (B) in clause (iii)—(i) in subclause (II), by striking ''Stewart B. McKinney Homeless Assistance Act'' and inserting S. 1177—287 ''McKinney-Vento Homeless Assistance Act (42 U.S.C. 11411)''; and (ii) by striking subclause (III) and inserting the following: ''(III) used for affordable housing assisted under the Native American Housing Assistance and Self-Determination Act of 1996 (25 U.S.C. 4101 et seq.); or''.

SEC. 7013. AUTHORIZATION OF APPROPRIATIONS.

Section 7014, as amended and redesignated by section 7001 of this Act, is further amended—(1) in subsection (a), by striking ''$32,000,000 for fiscal year 2000 and such sums as may be necessary for each of the seven succeeding fiscal years'' and inserting ''$66,813,000 for each of fiscal years 2017 through 2019, and $71,997,917 for fiscal year 2020''; (2) in subsection (b), by striking ''$809,400,000 for fiscal year 2000 and such sums as may be necessary for each of the seven succeeding fiscal years'' and inserting ''$1,151,233,000 for each of fiscal years 2017 through 2019, and $1,240,572,618 for fiscal year 2020''; (3) in subsection (c)—(A) by striking ''section 8003(d)'' and inserting ''section 7003(d)''; and (B) by striking ''$50,000,000 for fiscal year 2000 and such sums as may be necessary for each of the seven succeeding fiscal years'' and inserting ''$48,316,000 for each of fiscal years 2017 through 2019, and $52,065,487 for fiscal year 2020''; (4) by redesignating subsections (e) and (f) as subsections (d) and (e), respectively; (5) in subsection (d) (as redesignated by paragraph (4))—(A) by striking ''section 8007'' and inserting ''section 7007''; and (B) by striking ''$10,052,000 for fiscal year 2000 and such sums as may be necessary for fiscal year 2001, $150,000,000 for fiscal year 2002, and such sums as may be necessary for each of the five succeeding fiscal years'' and inserting ''$17,406,000 for each of fiscal years 2017 through 2019, and 18,756,765 for fiscal year 2020''; and (6) in subsection (e) (as redesignated by paragraph (4))—(A) by striking ''section 8008'' and inserting ''section 7008''; and (B) by striking ''$5,000,000 for fiscal year 2000 and such sums as may be necessary for each of the seven succeeding fiscal years'' and inserting ''$4,835,000 for each of fiscal years 2017 through 2019, and $5,210,213 for fiscal year 2020''.

Appendix C
State of Virginia Law: Uniformed Services-Connected Students, 2015

SB 1354 (Reeves) and HB 2373 (Ramadan)(2015)

Be it enacted by the General Assembly of Virginia:

1. That the Code of Virginia is amended by adding a section numbered **22.1-287.02** as follows:

§ **22.1-287.02**. Uniformed services-connected students.

A. For purposes of this section, a "uniformed services-connected student" means a student enrolled in a public school whose parent is serving in either (i) the active component of the U.S. Army, Navy, Air Force, Marine Corps, Coast Guard, or National Guard, the Commissioned Corps of the National Oceanic and Atmospheric Administration, or the Commissioned Corps of the U.S. Public Health Services or (ii) the reserve component of the U.S. Army, Navy, Air Force, Marine Corps, Coast Guard, or National Guard.

B. The Department of Education shall establish a process for the identification of newly enrolled uniformed services-connected students by local school divisions. Local school divisions shall identify newly enrolled uniformed services-connected students in accordance with such process.

C. Nonidentifiable, aggregate data collected from the identification of uniformed services-connected students shall be made available to local, state, and federal entities for the purposes of becoming eligible for nongeneral fund sources and receiving services to meet the needs of uniformed services-connected students residing in the Commonwealth.

D. Data collected from the identification of uniformed services-connected students shall not be a public record as defined in § **2.2-3701**. No person shall disclose such data except as permitted under the provisions of the federal Family Educational Rights and Privacy Act (20 U.S.C. § 1232g) and related regulations. No such data shall be used for the purposes of student achievement, the Standards of Accreditation, student-growth indicators, the school performance report card, or any other school rating system.

Appendix D
Military-Connected Students (MCS) by District, January 2016

District	Number of MCS	Percent	Cumulative Number	Cumulative Percent
Abbeville	1	0.01	1	0.01
Aiken	83	0.87	84	0.88
Anderson 1	5	0.05	89	0.93
Anderson 2	4	0.04	93	0.97
Anderson 3	1	0.01	94	0.98
Anderson 4	18	0.19	112	1.17
Bamberg 1	2	0.02	114	1.19
Barnwell 45	1	0.01	115	1.20
Beaufort 1	1055	11.02	1170	12.22
Berkeley 1	1627	17.00	2797	29.22
Charleston	162	1.69	2959	30.91
Cherokee	5	0.05	2964	30.97
Chester 1	7	0.07	2971	31.04
Chesterfield	5	0.05	2976	31.09
Clarendon 2	15	0.16	2991	31.25
Colleton 1	65	0.68	3056	31.93
Darlington	4	0.04	3060	31.97
Deaf & Blind School	3	0.03	3063	32.00
Dillon 4	1	0.01	3064	32.01
Dorchester 2	1600	16.72	4664	48.73
Dorchester 4	2	0.02	4666	48.75
Edgefield	62	0.65	4728	49.39
Fairfield	6	0.06	4734	49.46
Florence 1	51	0.53	4785	49.99
Georgetown	11	0.11	4796	50.10
Governor's School for Math and Science	6	0.06	4802	50.17
Greenville 1	58	0.61	4860	50.77
Greenwood 50	9	0.09	4869	50.87
Hampton	74	0.77	4943	51.64
Horry	56	0.59	4999	52.23
Kershaw	310	3.24	5309	55.46
Lancaster	12	0.13	5321	55.59
Laurens 55	4	0.04	5325	55.63
Laurens 56	7	0.07	5332	55.70
Lee	1	0.01	5333	55.71
Lexington 1	897	9.37	6230	65.09
Lexington 2	8	0.08	6238	65.17
Lexington 5	10	0.10	6248	65.27
Marion 10	5	0.05	6253	65.33
McCormick	2	0.02	6255	65.35
Newberry	27	0.28	6282	65.63
Oconee	164	1.71	6446	67.34
Orangeburg 3	3	0.03	6449	67.37
Orangeburg 4	1	0.01	6450	67.38
Orangeburg 5	29	0.30	6479	67.69
Pickens	127	1.33	6606	69.01

District	Number of MCS	Percent	Cumulative Number	Cumulative Percent
Richland 1	89	0.93	6695	69.94
Richland 2	1985	20.74	8680	90.68
SC Public Charter School District	140	1.46	8820	92.14
Saluda	2	0.02	8822	92.16
Spartanburg 1	2	0.02	8824	92.19
Spartanburg 2	2	0.02	8826	92.21
Spartanburg 5	4	0.04	8830	92.25
Spartanburg 6	2	0.02	8832	92.27
Spartanburg 7	30	0.31	8862	92.58
Sumter	651	6.80	9513	99.38
Union	8	0.08	9521	99.47
Williamsburg	8	0.08	9529	99.55
York 1	9	0.09	9538	99.64
York 2	11	0.11	9549	99.76
York 3	17	0.18	9566	99.94
York 4	6	0.06	9572	100.00

The SC Education Oversight Committee is an independent, non-partisan group made up of 18 educators, business persons, and elected leaders. Created in 1998, the committee is dedicated to reporting facts, measuring change, and promoting progress within South Carolina's education system.

ADDITIONAL INFORMATION
If you have questions, please contact the Education Oversight Committee (EOC) staff for additional information. The phone number is 803.734.6148. Also, please visit the EOC website at www.eoc.sc.gov for additional resources.

The Education Oversight Committee does not discriminate on the basis of race, color, national origin, religion, sex, or handicap in its practices relating to employment or establishment and administration of its programs and initiatives. Inquiries regarding employment, programs and initiatives of the Committee should be directed to the Executive Director 803.734.6148.

EDUCATION OVERSIGHT COMMITTEE

Subcommittee: EIA and Improvement Mechanisms

Date: April 11, 2016

INFORMATION
Results of the 2015 Parent Survey

PURPOSE/AUTHORITY
Section 59-28-190 of the Parental Involvement in Their Children's Education Act requires the Education Oversight Committee (EOC) to "survey parents to determine if state and local efforts are effective in increasing parental involvement." In addition Section 59-18-900 of the Education Accountability Act (EAA) requires that the annual school report cards include "evaluations of the school by parents, teachers, and students" as performance indicators to evaluate schools. The tool that has been adopted by the EOC and administered by the South Carolina Department of Education (SCDE) to meet these statutory requirements is the annual parent survey.

CRITICAL FACTS
The parent survey was commissioned by the EOC and designed by the Institute for Families in Society at the University of South Carolina in 2001. The survey is designed to determine parent perceptions of their child's school and to evaluate the effectiveness of state and local parental involvement programs. Since 2002 the South Carolina Department of Education has annually administered the survey, and the EOC has provided an annual review of the survey results. The attached report reflects the results of the 2015 administration of the parent survey and compares the results on key issues with the 2015 administration of the teacher survey. The report also documents the results of new items on the parent survey regarding bullying.

TIMELINE/REVIEW PROCESS
Study began in January 2016 and completed in March 2016

ECONOMIC IMPACT FOR EOC

Cost: No fiscal impact beyond current appropriations

Fund/Source:

ACTION REQUEST

☐ **For approval** ☒ **For information**

ACTION TAKEN

☐ **Approved** ☐ **Amended**

☐ **Not Approved** ☐ **Action deferred (explain)**

2016

Results of the 2015 Parent Survey

SC EDUCATION
OVERSIGHT COMMITTEE

PO Box 11867 | 227 Blatt Building | Columbia SC 29211 | WWW.SCEOC.ORG

CONTENTS

Page

Acknowledgements iii

Executive Summary 1

Part One – Administration of the 2015 Parent and Teacher Surveys 5

Part Two – Respondents of the 2015 Parent Survey 7

Part Three – Results for Items of the 2015 Parent Survey 13

Part Four – Results of the 2015 Teacher Survey 25

Appendices 37

A. Copy of the 2015 Parent Survey

B. Items of the 2015 Teacher Survey

Acknowledgements

The Education Oversight Committee (EOC) acknowledges the ongoing assistance of Cynthia Hearn and Ling Gao of the South Carolina Department of Education (SCDE) in providing data files, timely updates and important information on the annual administration of the parent survey. The EOC also appreciates the parents, teachers, and students who took the time to complete and return their annual surveys, because their perspectives are critical in evaluating public schools. Finally, the EOC is also grateful for principals and administrators who encouraged participation in the survey, and who oversaw the administration of the survey.

Executive Summary

Background: The parent survey was designed in 2001 to meet the requirements of the Education Accountability Act (EAA) and the Parental Involvement in Their Children's Education Act. Section 59-18-900 of the EAA requires that the annual school report card include "evaluations of the school by parents, teachers, and students" as performance indicators to evaluate schools. In addition Section 59-28-190 of the Parental Involvement in Their Children's Education Act requires the Education Oversight Committee (EOC) to "survey parents to determine if state and local efforts are effective in increasing parental involvement." The tool that has been adopted by the EOC and administered by the South Carolina Department of Education (SCDE) to meet these statutory requirements is the annual parent survey.

Since 2002 the SCDE has administered the parent survey to a sample of parents whose children attended public schools in South Carolina. From its inception, the parent survey contains items regarding parent perceptions of the learning environment in the school, home and school relations, and the social and physical environment of the school. Additional questions document characteristics of the parents and the children of the parents responding to the survey. The 2015 parent survey contained many of the same items as the 2014 parent survey. Three items were added for the 2015 survey to obtain information about student bullying.

The parents of students in the highest grade at all elementary, middle and high schools are surveyed. In high schools and career centers, parents of all 11th graders are surveyed. In schools with a grade configuration that spans multiple levels, parents of children in multiple grades are surveyed. For example, in a school with a grade span of grades 6 through 10, parents of children in grades 8 and 10 are surveyed. For parents in schools with a grade span of K-12, parents of children in grades 5, 8 and 11 are surveyed. Parents in schools containing grades 2 or lower (K-1, K-2, and 1-2 configurations) are not surveyed. Annually, the EOC has analyzed the results of the parent survey and issued reports. The reports are online at www.eoc.sc.gov.

Survey Responses: In 2015 the number of parent surveys completed and returned totaled 62,192, an increase of 2,899 surveys (4.9 percent) from the prior year. Estimates are that between 32.3 and 38.3 percent of all eligible parents surveyed responded to the 2015 parent survey. In 2015 the percentage of parents who completed the survey who identified themselves as Hispanic was 6.4 percent, compared to 5.7 percent in 2014. The percentage of survey respondents who are Hispanic has increased each of the past five years.

An analysis of the respondents to the 2015 parent survey concluded that the survey responses typically overrepresented the perceptions of parents who had children in elementary schools and underrepresented the perceptions of parents who had children in high school. Furthermore, the respondents typically obtained higher educational achievements and had greater median household

incomes than the general population of South Carolina. As in prior years, the "typical" parent responding to the survey was a white female having attended or graduated from college and having a household income of greater than $35,000. Furthermore, when compared to the enrollment of students in public schools, parents of African American students were underrepresented in the responses.

Parent Survey Results: The results of the 2015 parent survey demonstrate that parent satisfaction levels with the three characteristics measured - the learning environment and social and physical environment of their child's school—were consistent with the prior year's results. Significant changes are estimated as an annual increase or decrease of three or more percent. Satisfaction is defined as the percentage of parents who agreed or strongly agreed that they were satisfied with the learning environment, home and school relations, and social and physical environment of their child's school. After parent satisfaction with home and school relations declined from 2013 to 2014 and the number of missing responses for this item increased by a comparable amount, parent satisfaction in 2015 did not differ from 2014. The percentage of parents satisfied in 2014 was 71.7, and the percentage of parents satisfied in 2015 was 73.1 percent.

Percentage of Parents Satisfied with Each Characteristic: 2012-2015

Characteristic	**2015**	2014	2013	2012	**Difference between 2015 and 2014**
Learning Environment	**87.6**	86.7	87.0	87.2	**0.9**
Home and School Relations	**73.1**	71.7	83.3	82.9	**1.4**
Social and Physical Environment	**85.3**	84.4	84.3	84.1	**0.9**

Parents who responded to the 2015 annual survey reported levels of parental involvement comparable to previous years and identified work schedules as their greatest obstacle to involvement.

Parent Reported Obstacles to Parental Involvement in 2015

Work Schedule	56.2%
Lack of timely notification of volunteer opportunities	24.3%
School does not encourage involvement	10.8%
Family and health problems	14.9%
Lack of child or adult care services	14.5%
Transportation	10.8%
Involvement not appreciated	10.8%

Items parents perceive as impediments to parental involvement that are at least partially within the control of the schools are the processes by which schools notify parents of volunteer opportunities, the means by which the school encourages or enables interaction between parents and the school, and the approach of the school toward parental involvement.

In previous reports of the parent survey, analyses were performed relating parent satisfaction to school report card grades. Since report card grades were not available for 2015, teacher survey results were analyzed, and were related to parent satisfaction with the overall learning environment of the school. Five categories of parent satisfaction were created (quintiles), from lowest to highest, with each category containing approximately one-fifth of schools. For nearly all teacher survey items, as the level of parent satisfaction with the learning environment of their child's school increased, so did the percent of teachers who viewed the school favorably. The largest difference between teachers in schools with the highest and lowest parent satisfaction with the school learning environment was with respect to home and school relations, and the smallest difference between teachers in schools with the highest and lowest parent satisfaction ratings of the school learning environment was with respect to working conditions.

Percentage of Teachers Who Strongly Agree that they are Satisfied with the Overall Measure of Each Characteristic by Parental Satisfaction with the School Learning Environment:

Characteristic	Lowest Parent Satisfaction	2nd Quintile	3rd Quintile	4th Quintile	Highest Parent Satisfaction
Learning Environment	**45**	57	64	69	**77**
Home and School Relations	**25**	34	44	52	**65**
Social and Physical Environment	**44**	57	63	69	**77**
Working Conditions	**47**	55	60	65	**72**

See Tables 26, 27, 28, and 29

Results Regarding Bullying: Approximately 19 percent of parents reported that their child had been bullied. When bullying occurred, parents most frequently reported that it occurred in the classroom or in some other location in the school. Sporting events were the location with the fewest reports as a location for bullying, followed by online and/or texting during school hours.

Approximately 16 percent of teachers reported that they were bullied by another adult at their school and approximately 14 percent of teachers reported that they either agreed or strongly agreed that they had been bullied by a student at school.

PART ONE
Administration of the 2015 Parent and Teacher Surveys

The design and sampling methodology for the parent survey were established in 2001. The EOC contracted with the Institute of Families in Society at the University of South Carolina to design the survey and to recommend a medium for distributing the survey. To maintain complete anonymity and to maximize the return rate, the Institute recommended that the survey be mailed to a sample of parents along with a postage paid, return envelope. While the sampling methodology proposed by the Institute was implemented, the parent survey has never been mailed to parents due to budgetary restrictions. Instead, schools have been given the responsibility for distributing and collecting the forms. Generally, schools send the surveys home with students. Some schools have held parent meetings or special meetings at school during which the surveys were distributed.

Rather than surveying all parents of public school students, the parents of students in the highest grade at all elementary, middle and high schools are surveyed. In high schools and career centers, parents of all 11th graders are surveyed. In schools with a grade configuration that spans multiple levels, parents of children in multiple grades are surveyed. For example, in a school with a grade span of grades 6 through 10, parents of children in grades 8 and 10 are surveyed. For parents in schools with a grade span of K-12, parents of children in grades 5, 8 and 11 are surveyed. Parents in schools containing grades 2 or lower, which include primary schools, child development schools and schools with configurations like K, K-1, and K-2 are not surveyed. The parent survey is typically administered during the second semester of each school year.

A copy of the 2015 survey is in the Appendix A. The 2015 administration of the parent survey occurred over the following time period and involved the following actions.

March 4, 2015	All schools received survey forms.
April 3, 2015	Date for parent survey forms returned to school.
April 9, 2015	Last day for schools to mail completed forms to contractor.

Source: SC Department of Education

A school survey coordinator, a staff person designated by the school principal, distributed and collected the parent surveys at each school according to instructions provided by the South Carolina Department of Education (SCDE). According to SCDE, an independent contractor hired by the agency to mail to each school the following:

- ✓ An administrative envelope containing;
  1. A letter to the principal from the Education Oversight Committee (EOC),
  2. Two sets of instructions for administering the surveys,
  3. A page of shipping instructions, and
  4. One pre-addressed, bar-coded UPS shipping label (used to return completed surveys to contractor, freight prepaid).
- ✓ Parent survey envelopes. Each envelope contains a letter from the State Superintendent of Education and a parent survey form.
- ✓ Student survey forms.[1]

The name of each school was printed on the survey forms to assist parents who were completing surveys for multiple schools. Schools were also advised to "distribute the parent surveys as soon as

[1] "Administration of the 2015 Report Card Surveys," South Carolina Department of Education.

possible" after delivery. The cost of printing, shipping, processing and scanning the parent surveys was approximately $115,000.[2]

Each school's designated survey coordinator then distributed envelopes containing the parent survey and letter from the state Superintendent of Education to each classroom teacher within the designated grade being surveyed. Teachers gave each student an envelope and instructions to take the envelope home for their parents to complete and then return the completed survey to school in the sealed envelope. The envelopes were designed to maintain the confidentiality and anonymity of all parents. Parents were given the option of mailing the completed survey directly to SCDE with parents incurring the cost of the mailing or of returning the survey to the school. The school survey coordinator was expressly advised that mailing of the envelopes directly to the parents was allowed with all costs to be borne by the school. Information did not exist to document if any schools mailed the parent surveys to parents.

Upon receiving the completed parent surveys, the school survey coordinator then mailed the forms to the independent contractor for scanning and preparation of the data files. Individual school results were tabulated by SCDE. The overall parent satisfaction scores of three questions relating to the school's overall learning environment, home and school relations, and social and physical environment were printed on the 2015 annual school report cards. For each school, SCDE aggregated the responses to all survey questions and provided the data files to the district office.

The 2015 parent survey contained a total of fifty-eight questions. Forty-seven questions were designed to elicit information on parental perceptions and parental involvement patterns. For the first twenty-three questions, parents were asked to respond to individual statements using one of the following responses: Strongly Disagree, Disagree, Agree, Strongly Agree or Don't Know. These twenty-one questions focused on three key components: learning environment, home and school relations, and the physical and social environment of their child's school. These components and individual activities reflect the framework devised by Dr. Joyce Epstein of the National Network of Partnership Schools.

Parents were asked five questions about their participation in various parental involvement activities both in and outside of the school. Parents were also asked whether each of a list of seven items were potential barriers to their involvement in their child's education. New to the 2015 survey were three items focused on whether parents believed their child was bullied at school in the previous year, where the bullying occurred, and whether the bullying was verbal or physical. Finally, parents were asked to provide specific information about themselves, their child, and their household. Parents were asked four questions about their child: their child's grade in school, gender, race/ethnicity, and grades on his or her last report card. Four questions sought information about the parent: his or her gender, race/ethnicity, highest level of education and total yearly household income.

For this year, analyses of the parent survey were performed in conjunction with responses of teachers to the annual teacher survey. The teacher survey also includes items on the learning environment in the school, home and school relations, and the social and physical environment of the school. The teacher survey also includes items regarding teacher perceptions of their working conditions, including the physical conditions that impact their teaching, the non-academic tasks associated with being a teacher, the working conditions in their school, and other items. All items are presented with the following responses available: Strongly Disagree, Disagree, Agree, and Strongly Agree, and Don't Know.

The teacher surveys were administered online to all teachers in all grade levels. A link to the survey was available on the front page of the South Carolina Department of Education (SCDE) web-site from March 4, 2015 to April 3, 2015.

[2] Communication from South Carolina Department of Education to EOC staff.

PART TWO
Respondents of the 2015 Parent Survey

As reflected in Table 1, the total number of parent surveys returned in 2015 was 62,192, which was 2,899 (4.9 percent) more than the number returned in the prior year. This increase reverses a trend of decreasing parent responses from 2011 to 2014. The current year response total is15.6 percent lower than the highest response total (73,755), which was obtained in 2011.

Table 1
Total Number of Parent Surveys Returned

Year	Surveys
2015	62,192
2014	59,293
2013	66,787
2012	69,581
2011	73,755
2010	69,474

Using two methods of determining response rates and the total number of parent surveys returned, two response rates were calculated in Table 2. The first method compares the number of responses to the number of surveys distributed, and the second method compares the number of responses to the number of students in grades 5, 8, and 11 (grades 5 and 8 are typically the highest grades in elementary and middle school, and grade 11 is the high school grade targeted for administration of the parent survey). From these separate calculations, it appears that between 32.3 and 38.3 percent of all eligible parents surveyed responded to the 2015 parent survey. In the prior year using the same two methodologies, the response rate was between 31 and 37 percent.

Table 2
Determining the Response Rate

	Sample Size	Surveys Returned	Response Rate
Method 1: Surveys Distributed	192,663	62,192	32.3%
Method 2: ADM[6] of 5, 8 and 11th grades	162,254		38.3%

Parents completing the survey were asked seven questions about their child:

1. What grade is your child in? (3rd, 4th, 5th, 6th, 7th, 8th, 9th, 10th or 11th)
2. What is your child's gender?
3. What is your child's race/ethnicity?
4. What grades did your child receive on his/her last report card?
5. Has your child been bullied at school this year?
6. If yes, was your child bullied:
   - In Classroom
   - Other location at school
   - At sporting events
   - On-line/texting during school
   - On the bus
   - After school

7. If yes, was you child bullied
    - Physically
    - Verbally
    - Both

The following definition of bullying was provided on the survey:

> *Bullying is when 1 or more students tease, threaten, spread rumors about, hit, shove, or hurt another student over and over again physically. It is not bullying when 2 students of about the same strength or power argue or fight or tease each other in a friendly way.*

Parents were also asked four questions about themselves and their family:

1. What is your gender?
2. What is your race/ethnic group?
3. What is the highest level of education you have completed?
    - Attended elementary/high school
    - Completed high school/GED
    - Earned associate degree
    - Attended college/training program
    - Earned college degree
    - Postgraduate study/and/or degree
4. What is your family's total yearly household income?
    - Less than $15,000
    - $15,000 - $24,999
    - $25,000 - $34,999
    - $35,000 - $54,999
    - $55,000 - $75,000
    - More than $75,000

Responses to these questions revealed the following about the parents who completed the 2015 parent survey (Table 3).

Table 3
Respondents to the 2015 Parent Survey
(n=62,192)

Gender	
Male	**14.8%**
Female	**85.2%**
Race	
African-American	**30.1%**
Caucasian/white	**59.0%**
Hispanic	**6.4%**
All Other	**4.5%**
Education	
Attended elementary/high school	**9.8%**
Completed high school/GED	**11.2%**
Earned Associate Degree	**22.5%**
Attended college/training program	**22.3%**
Earned college degree	**20.4%**
Postgraduate study/and/or degree	**13.9%**
Household Income	
Less than $15,000	**12.6%**
$15,000 - $24,999	**13.2%**
$25,000 - $34,999	**14.0%**
$35,000 - $54,999	**13.4%**
$55,000 - $75,000	**16.4%**
More than $75,000	**30.4%**
Their Child Enrolled in:	
Grades 3-5	**44.9%**
Grades 6-8	**36.4%**
Grades 9-11	**18.8%**
Their Child's Gender:	
Male	**45.1%**
Female	**54.9%**
Their Child's Ethnicity:	
African-American	**30.6%**
Caucasian/White	**57.0%**
Hispanic	**6.6%**
All Other	**5.8%**
Their Child's Grades:	
All or mostly A's and B's	**63.2%**
All or mostly B's and C's	**25.9%**
All or mostly C's and D's	**9.0%**
All or mostly D's and F's	**1.9%**

Note: Percentages may not add up to 100% due to rounding.

As in prior years, the "typical" parent responding to the survey was a white female having attended or graduated from college. Over 60 percent of the respondents who answered the question about income reported earning over $35,000. The percentage of respondents that identified themselves as Hispanic has steadily increased from 5.0 percent in 2010 to 6.6 percent in 2015.

To determine if the survey responses were representative of elementary, middle and high school parents, the following analysis was done. First, 51,520 parents who returned the 2015 survey indicated that their child was in 5th, 8th, or 11th grade. Defining grade 5 as elementary schools, grade 8 as middle school and grade 11, high school, approximately 44 percent of parents who completed the survey were elementary school parents, 36 percent middle school, and 20 percent high school (Table 4). As compared to the prior year, the percentage of surveys reflecting the perceptions of elementary school parents declined by 2 percent, middle school parents remained the same, and the percentage of parents of high school students increased by 2 percent (from 18 to 20).

The representativeness of the 2015 parent surveys returned of the population of students was investigated by comparing the grade level and ethnicity of students enrolled in the 2014-15 academic year to the grade level and ethnicity of students as reported by parents in the 2015 parent survey. Considering only students in grades 5, 8, and 11, 44 percent of the parent surveys indicated their child was enrolled in grade 5, yet according to the 135-day Average Daily Membership (ADM) enrollment, only 34 percent of students are in grade 5. The percentage of parents who reported their child was enrolled in grade 8 is nearly identical to the percentage of student enrolled in grade 8 according to the ADM. The percentage of parents who reported their child was enrolled in grade 11 (20 percent) is much smaller than the percentage of students enrolled in grade 11 from the ADM (31 percent). Elementary school students are, then, over-represented in the parent surveys returned and high school students are under-represented in these data.

Table 4
Parental Respondents by Child's Grade

Grade of Child	Surveys Returned	% of Surveys from Grades 5, 8, & 11		2014-15 135-day ADM	% of ADMs for Grades 5, 8 & 11
Grade 5	22,586	44%		55,230	34%
Grade 8	18,660	36%		57,044	35%
Grade 11	10,274	20%		49,980	31%
TOTAL	51,520			162,254	

When asked about their child's race or ethnicity, 57.0 percent of the parents responded that their child's ethnicity was white, 30.6 percent African American and 6.6 percent Hispanic. With respect to the ethnicity of children in the public schools of South Carolina in 2014-15, parents whose children are African American were underrepresented by 6.7 percent, and parents whose children are Hispanic were underrepresented by 1.4 percent in the respondents (Table 5).

Table 5
Ethnicity of Children

	2015 Parent Survey	Student Enrollment All Public Schools 2014-15[3]	Difference
White	57.0%	52.0%	5.0%
African American	30.6%	37.3%	(6.7%)
Hispanic	6.6%	8.0%	(1.4%)
Other	5.8%	2.7%	3.1%

Note: “Other” includes American Indian/Alaskan, Asian, Hawaiian Native/Pacific Islander and Two or more races.

With respect to educational attainment, 34.1 percent of parents who responded to the survey in 2015 had earned a bachelor or postgraduate degree. For comparison purposes, the United States Census Bureau projected that 25.1 percent of persons 25 years old and over in South Carolina had earned a bachelor’s degree or higher in 2009.[4]

Regarding the annual household income of the respondents, in 2015 60.2 percent of the parents who completed the survey reported having an annual household income in excess of $35,000. For comparison purposes, according to the U.S. Census Bureau, the median household income in South Carolina from 2009-2013 was $44,779.[5]

Conclusions

- A total of 62,192 parent surveys were completed and returned in 2014, which was 2,899 (4.9 percent) more than the number returned in the prior year. This increase in response reversed a 3-year trend of declining parent response.
- Using two methods of calculating a response rate, one method that underestimated and one that overestimated the total number of parents eligible to take the survey, the response rate to the 2015 parent survey was between 32 and 38 percent, which is slightly higher than the response rate of 31 to 37 percent in 2014.
- An analysis of the respondents to the 2015 parent survey found that the survey responses typically overrepresented the perceptions of parents in elementary schools and underrepresented the perceptions of parents who have children in high school.
- Respondents typically obtained higher educational achievements and had greater median household incomes than the general population of South Carolina.
- The percentages of respondents by racial/ethnic group were within 7 percent of the make-up of the South Carolina population.
- African-American parents were more underrepresented in the 2015 survey (6.7 percent) than in the 2014 survey (2.9 percent).

[4] U.S. Census Bureau, “State and County Quick Facts” <http://quickfacts.census.gov/qfd/states/45000.html>, accessed April 13, 2015.

[5] Ibid.

PART THREE
Results for Items of the 2015 Parent Survey

The parent survey was designed to determine: (1) parent perceptions or satisfaction with their child's public school and (2) parental involvement efforts in public schools. The following is an analysis that documents the actual parent responses to questions focusing on parental satisfaction and parental involvement.

Parent Perceptions of Their Child's School

The information below summarizes the results of the 2015 parent survey. At the school level, responses to these questions can reveal the strengths and weaknesses of parental involvement initiatives at the individual school site. Statewide, the data provide policymakers information on the overall effectiveness of policies and programs in promoting parental involvement. The following analysis focuses on parent perceptions or satisfaction with the learning environment, home and school relations, and the social and physical environment of their children's schools. In analyzing responses, "significant change" is defined as a change of three percent or more in satisfaction.

A. Learning Environment

Five questions in the parent survey ask parents to reflect upon the learning environment of their child's school. Questions 1 through 4 are designed to elicit parental agreement with specific aspects of the learning environment at their child's school, focusing on homework, expectations, and academic assistance. Question 5 offers parents the opportunity to report on their overall satisfaction with the learning environment at their child's school. For each school with a sufficient number or parent survey responses, the aggregate parental responses to question 5 are included on the annual school report.

Table 6 summarizes the total responses to these five questions for all parents who completed the 2015 parent survey. Overall, 87.6 percent of parents responded that they were satisfied with the learning environment of their child's school. The percentage of parents who disagreed or strongly disagreed was highest for questions 4 and 5. Approximately 17 percent of parents either did not believe or did not know if their child received extra help when needed.

Table 6
Parent Responses to the 2015 Learning Environment Questions
(Percentage of Parents with each Response)

Question	Agree or Strongly Agree	Disagree or Strongly Disagree	Don't Know
1. My child's teachers give homework that helps my child learn.	**89.2**	**8.3**	**2.5**
2. My child's school has high expectations for student learning.	**92.2**	**5.8**	**2.0**
3. My child's teachers encourage my child to learn.	**91.8**	**5.2**	**3.0**
4. My child's teachers provide extra help when my child needs it.	**82.8**	**11.0**	**6.2**
5. I am satisfied with the learning environment at my child's school	**87.6**	**10.7**	**1.7**

Table 7 compares the percentage of parents who responded that they agreed or strongly agreed to these questions each year from 2011 through 2015. The pattern over time is high parental satisfaction with the learning environment, with the highest levels of parental satisfaction for all items in 2015.

Table 7
Percentage of Parents Who Agree or Strongly Agree they are Satisfied with each Learning Environment Question: 2011 through 2015

Learning Environment Questions	**2015**	2014	2013	2012	2011
1. My child's teachers give homework that helps my child learn.	**89.2**	88.9	89.6	89.9	86.7
2. My child's school has high expectations for student learning.	**92.2**	91.6	91.7	91.7	88.9
3. My child's teachers encourage my child to learn.	**91.8**	91.2	91.5	91.8	88.7
4. My child's teachers provide extra help when my child needs it.	**82.8**	81.9	81.7	81.9	78.7
5. I am satisfied with the learning environment at my child's school	**87.6**	86.7	87.0	87.2	84.3

Parents of elementary school students view the learning environment of the school more favorably (90.2 percent) than do parents of either middle (84.9 percent) or high school (85.3 percent) students (Table 8). The difference between the parent responses for parents of middle and high school students are not large enough to suggest these groups differ in their perceptions of their child's school. Parents of elementary school students do appear to view the learning environment of their child's school most favorably.

Table 8
I am Satisfied With the Learning Environment at My Child's School.
(Percentage of Parents by School Type: Elementary, Middle or High School)

School Type	Number of Responses	Agree or Strongly Agree	Disagree or Strongly Disagree
Elementary	29,675	90.2	8.6
Middle	21,076	84.9	13.1
High	9,698	85.3	12.2

B. Home and School Relations

The next eleven questions on the parent survey determine parent perception of home and school relations by focusing on the relationship between the parent and their child's teacher and between the parent and the school. Question 11 offers parents the opportunity to report on their overall satisfaction with home and school relations at their child's school. For each school with a sufficient number of parent responses, the aggregate parental responses to question 11 are included on the annual school report card.

Table 9 summarizes the total responses to these eleven questions for all parents who completed the 2015 parent survey.

Table 9
Parent Responses to the 2015 Home and School Relations Questions
(Percentage of Parents with each Response)

Home and School Relations Questions	Agree or Strongly Agree	Disagree or Strongly Disagree	Don't Know
1. My child's teachers contact me to say good things about my child	**58.1**	**39.8**	**2.1**
2. My child's teachers tell me how I can help my child learn.	**64.1**	**33.4**	**2.5**
3. My child's teachers invite me to visit my child's classrooms during the school day.	**49.5**	**45.5**	**5.0**
4. My child's school returns my phone calls or e-mails promptly.	**81.9**	**12.8**	**5.3**
5. My child's school includes me in decision-making.	**69.6**	**24.2**	**6.2**
6. My child's school gives me information about what my child should be learning in school.	**76.9**	**20.9**	**2.2**
7. My child's school considers changes based on what parents say.	**51.6**	**24.9**	**23.5**
8. My child's school schedules activities at times that I can attend.	**78.8**	**16.9**	**4.3**
9. My child's school treats all students fairly.	**70.2**	**16.6**	**13.1**
10. My principal at my child's school is available and welcoming.	**81.8**	**9.7**	**8.5**
11. I am satisfied with home and school relations at my child's school	**73.1**	**14.4**	**12.5**

Overall, 73.1 percent of parents were satisfied with home and school relations at their child's school, which is 1.4 percent more than the percentage in 2014. An examination of questions 1 through 10, which ask parents more specific questions about their personal experiences at their child's school, reveals the following, which is consistent with results of the 2014 survey:

- Parents overwhelmingly agreed that the principal at their child's school was available and welcoming.

- Slightly more than 80 percent of the parents agreed that their child's school returned phone calls or e-mails promptly and scheduled activities at times that parents could attend.

- Approximately four out of ten parents disagreed or strongly disagreed that their child's teachers contacted them to say good things about their child or invited the parents to visit the classroom during the school day.

- Approximately one third of the parents disagreed that their child's teachers told them how to help their child learn.

- Approximately one-fourth of parents disagreed or strongly disagreed that their child's school included parents in decision-making or considered changes based on parental input.

- Approximately one-half of parents disagreed, strongly disagree, or did not know if their child's school considered changes based on parental input.

- Approximately one in three parents did not believe or did not know if students were treated fairly at their child's school.

As documented in Table 10, the trend is that parental satisfaction with home and school Relations increased from 2011 through 2013, declined dramatically in 2014, and changed little from 2014 to 2015. The dramatic decline in satisfaction from 2013 to 2014 is not accompanied by a corresponding increase in the percentage of parents expressing dissatisfaction with home and school relations. Instead, there was a substantial increase from 2013 to 2014 in the percentage of parents who indicated they did not have an opinion of the home and school relations. The percentage of parents who indicated they did not have an opinion did not change dramatically from 2014 to 2015.

Table 10
Percentage of Parents Who Agree or Strongly Agree they are Satisfied with Home and School Relations: 2011 through 2015

	2015	2014	2013	2012	2011
Agree or Strongly Agree	**73.1**	71.7	83.3	82.9	80.2
Disagree or Strongly Disagree	**14.4**	14.6	13.3	13.7	13.9
Don't Know	**12.5**	13.7	3.4	3.4	5.9

The pattern of parental satisfaction with home and school relations is similar to the pattern of parental satisfaction with the learning environment (Table 11). The percentages of parents of students in middle school and high school who view the home and school relations favorably (68.8 and 70.3 percent, respectively), are nearly the same. Both, however, are lower than the percentage of parents of students in elementary school who view home and school relations favorably (77.5 percent).

Table 11
I am Satisfied with Home and School Relations at My Child's School.
(Percentage of Parents by School Type: Elementary, Middle or High School)

School Type	Agree or Strongly Agree	Disagree or Strongly Disagree
Elementary	77.5	10.9
Middle	68.8	18.0
High	70.3	17.2

C. Social and Physical Environment

Five questions on the parent survey focus on the social and physical environment of schools. These questions are designed to elicit parent perceptions of the cleanliness, safety, and student behavior at their child's school. Question 5 asks parents to report on their overall satisfaction with the social and physical environment of their child's schools. For each school with a sufficient number of parent responses, the aggregate parental responses to question 5 are included on the annual school report card.

Table 12 summarizes the total responses to these five questions for all parents who completed the 2015 parent survey. Nine in ten parents agreed or strongly agreed that their child's school was kept neat and clean and that their child felt safe at school. On the other hand, over one out of three parents either did not believe or did not know whether students at their child's school were well behaved, and 15.5 percent of parents did not know or did not believe that their child's teachers cared about their child as an individual.

Table 12
Parent Responses to the 2015 Social and Physical Environment Questions
(Percentage of Parents with each Response)

Social and Physical Environment Questions	Agree or Strongly Agree	Disagree or Strongly Disagree	Don't Know
1. My child's school is kept neat and clean.	**90.5**	**6.1**	**3.4**
2. My child feels safe at school.	**89.1**	**8.2**	**2.6**
3. My child's teachers care about my child as an individual.	**84.6**	**8.5**	**7.0**
4. Students at my child's school are well behaved.	**64.9**	**21.1**	**14.0**
5. I am satisfied with the social and physical environment at my child's school.	**85.3**	**10.8**	**3.9**

Table 13 presents the 2015 results of the South Carolina parent survey with the results of parent surveys administered since 2011. The data document that parental responses to the five questions regarding the social and physical environment of their child's school are consistent with the prior year's results. Over time, parent satisfaction with the social and physical environment of their child's schools as reflected in the responses to these five questions has generally increased. The only question for which parental satisfaction declined was with respect to student safety, which decreased by 2.1 percent.

Table 13
Percentage of Parents Who Agree or Strongly Agree they are
Satisfied with each Social and Physical Environment Question: 2011 through 2015

Social and Physical Environment Questions	**2015**	2014	2013	2012	2011
1. My child's school is kept neat and clean.	**90.5**	90.6	91.5	91.3	90.0
2. My child feels safe at school.	**89.1**	91.2	91.0	90.9	89.7
3. My child's teachers care about my child as an individual.	**84.6**	83.8	83.7	84.1	81.1
4. Students at my child's school are well behaved.	**64.9**	64.8	64.0	63.7	61.2
5. I am satisfied with the social and physical environment at my child's school	**85.3**	84.4	84.3	84.1	82.4

Data presented in Table 14 demonstrate that the differences in parental satisfaction in the social and physical environment of their child's school by school type are consistent with results for both the learning environment and home and school relations. The percentage of parents of elementary school students express more satisfaction (89.2 percent) than either the parents of middle school students (81.7 percent) or high school students (80.4 percent). The difference between the percentages for parents of middle school and high school parents are not large enough to infer that these parents view the school differently.

Table 14
I am Satisfied with the Social and Physical Environment at My Child's School.
(Percentage of Parents by School Type: Elementary, Middle or High School)

Type	Agree or Strongly Agree	Disagree or Strongly Disagree
Elementary	89.2	8.0
Middle	81.7	13.9
High	80.4	13.9

D. Parental Involvement

According to the National Network of Partnership Schools, founded and directed by Dr. Joyce Epstein at Johns Hopkins University, there are six types of successful partnerships between the school, family and community:[6]

- Type 1. Parenting – Assist families with parenting skills and setting home conditions to support children as students. Also, assist schools to better understand families.

- Type 2. Communicating – Conduct effective communications from school-to-home and home-to-school about school programs and student progress.

- Type 3. Volunteering – Organize volunteers and audiences to support the school and students. Provide volunteer opportunities in various locations and at various times.

- Type 4. Learning at Home – Involve families with their children on homework and other curriculum-related activities and decisions.

- Type 5. Decision Making – Include families as participants in school decisions, and develop parent leaders and representatives.

- Type 6. Collaborating with the family – Coordinate resources and services from the community for families, students, and the school, and provide services to the community.

In addition to determining parent satisfaction with their child's school, the annual survey of parents in South Carolina includes questions designed to elicit information on the level of parental involvement in schools. The questions focus on the first five types of parental involvement. It should be reiterated that parents self-report their involvement.

First, parents were asked to specifically respond to eight questions relating to their involvement in their child's school. These questions focus on the following types of parental involvement: parenting, volunteering and decision making. Parents were asked specifically to respond to these eight questions in one of four ways:

- I do this.
- I don't do this but would like to.
- I don't do this and I don't care to.
- The school does not offer this activity/event.

The responses are reflected in Table 15 with the middle column highlighting the percentage of parents who expressed an interest in becoming involved in these school activities. These parents want to be

[6] Epstein, et. al. 2002. *School, Family, and Community Partnerships: Your Handbook for Action, Second Education*. Thousand Oaks, CA: Corwin Press, Inc. <http://www.csos.jhu.edu/P2000/nnps_model/school/sixtypes.htm>.

involved but either have personal barriers preventing their involvement or face obstacles at the school level. At the school level, parents responding "I don't do this but would like to" are the parents for whom school initiatives to improve parental involvement should be focused.

Table 15
Percent of Parents Providing Each Response to
Parental Involvement Questions Regarding Activities at the School

Parental Involvement Question	I do this	I don't but would like to	I don't and don't care to	Activity/event not offered
Attend Open Houses or parent-teacher conferences	79.9	**15.0**	4.3	0.9
Attend student programs or performances	80.4	**14.7**	3.7	1.2
Volunteer for the school	34.1	**38.0**	24.7	3.2
Go on trip with my child's school	35.3	**42.3**	17.0	5.4
Participate in School Improvement Council Meetings	12.3	**43.3**	39.0	5.4
Participate in Parent-teacher Student Organizations	29.2	**35.1**	32.7	2.9
Participate in school committees	15.3	**37.8**	40.0	6.9
Attend parent workshops	24.5	**38.5**	22.5	14.6

Based on the responses in Table 15 and the six types of involvement, there are significant opportunities for improving parental involvement in South Carolina's public schools.

- Decision-Making – Substantially fewer parents report being involved in the School Improvement Council and school committees than in any other activity. Slightly less than one-third of parents report participating in Parent-Teacher-Student Organizations. Decision making, including parents and families in school decisions, and developing parent leaders and representatives are areas for growth where parents want to be involved in these decision-making organizations.

- Volunteering – Approximately 34 percent of the parents responded that they volunteered while 38 percent wanted to volunteer.

- Parenting - Over three-fourths of the parents attended open houses, parent-teacher conferences or student programs, all activities that support their children. Approximately one-fourth reported attending parent workshops while approximately 15 percent contend that such workshops were not provided at their child's school.

Parents were asked five questions about their involvement with their child's learning, both at the school site and at home. Parents could respond in one of three ways:

- I do this
- I don't do this but would like to
- I don't do this and I don't care to

Table 16 summarizes parental responses to these five questions.

Table 16
Percent of Parents Providing Each Response to
Parental Involvement Questions Regarding Their Child's Learning

	I do this	I don't but would like to	I don't and don't care to
Visit my child's classroom during the school day	28.1	51.7	20.2
Contact my child's teachers about my child's school work.	75.3	18.7	6.0
Limit the amount of time my child watches TV, plays video games, surfs the Internet	83.0	9.3	7.7
Make sure my child does his/her homework	94.7	3.5	1.8
Help my child with homework when he/she needs it.	93.2	5.2	1.6

Clearly, parents overwhelmingly report being involved in activities and decisions to support their child's learning. Over 93 percent of parents reported helping their child with his or her homework while 83.0 percent report limiting television and other distractions at home. Over one-fourth of parents responded that they visited their child's classroom during the day while a majority wanted to become involved in this way. These responses are similar to parent responses in prior years.

There are obstacles that impede parental involvement in schools. These obstacles may include lack of transportation, family responsibilities, and work schedules. Schools may not encourage or facilitate parental involvement at the school level. The annual parent survey asks parents to respond "true" or "false" to seven questions on factors that impact their involvement. The results from 2011 through 2015 are included in Table 17. Consistently across years, work schedule is the most common obstacle to parent involvement. At the individual school, the responses to these questions may assist principals and teachers in scheduling parental involvement activities or even parent-teacher conferences at times and places convenient for both parents and teachers.

Table 17
Percentage of Parents Experiencing Each Impediment to Involvement in Schools

	2015	2014	2013	2012	2011
Lack of transportation reduces my involvement	**10.8**	12.2	11.6	11.6	11.5
Family health problems reduce my involvement.	**14.9**	15.5	14.6	14.4	14.3
Lack of available care for my children or other family members reduces my involvement.	**14.5**	14.8	14.1	14.7	14.5
My work schedule makes it hard for me to be involved.	**56.2**	57.1	54.6	53.8	54.4
The school does not encourage my involvement.	**16.2**	17.5	16.1	15.7	16.2
Information about how to be involved either comes too late or not at all.	**24.3**	25.5	23.7	23.5	24.6
I don't feel like it is appreciated when I try to be involved.	**10.8**	11.9	11.3	10.6	11.4

Finally, parents were also asked several questions about their child's school and its efforts at increasing parental involvement. Across these questions and across time, two-thirds or more of parents consistently rated the efforts of their child's school at parental involvement efforts as good or very good (Table 18). Fewer than 10 percent of parents have provided unfavorable responses regarding their child's school for any of these questions over the past three years.

Table 18
Percent of Parents Providing Each Response to
Parental Involvement Questions Regarding School Effort: 2013-2015

	Very Good or Good			**Bad or Very Bad**			**Okay**		
Question:	**2015**	2014	2013	**2015**	**2014**	2013	**2015**	2014	2013
School's overall friendliness.	**80.9**	80.6	79.3	**2.2**	**1.6**	2.2	**16.9**	16.9	18.4
School's interest in parents' ideas and opinions.	**62.6**	62.5	63.4	**7.4**	**8.1**	7.6	**30.0**	29.4	30.1
School's effort to get important information from parents.	**70.8**	68.6	67.4	**6.3**	**7.5**	7.6	**22.9**	24.0	25.1
The school's efforts to give important information to parents.	**75.5**	73.9	73.1	**5.3**	**6.3**	6.1	**19.3**	19.8	20.8
How the school is doing overall.	*	76.9	75.8	*	**3.6**	3.2	*	19.5	21.0

* Not included in 2015 survey.

E. Bullying

Three new questions on the parent survey for 2015 addressed the topic of bullying. The first asked the parent if their child had been bullied at school. If a parent responded yes to the first question, they were asked to respond to two additional questions. The second question asked parents where their child was bullied, with the following options provided:

In classroom
Other location at school
At sporting events
On-line/texting during school
On the bus
After school

The final question asked whether their child was bullied physically, verbally, or both.

A total of 43,455 (72.1 percent) of parents indicated that their child was not bullied at school, while 11,583 (19.1 percent) parents indicated that their child was bullied at school, and 5,273 (8.7 percent) parents were not sure whether their child was bullied at school. Table 19 presents a summary of the locations in which children were bullied, ordered by frequency of occurrence. Classrooms were the location parents reported their child was bullied in most frequently (11.9 percent), followed by some other location at school (9.4 percent). Although only 5.1 percent of parents indicated that their child was bullied on the bus, this should not be interpreted as the percentage of bus riding children who were bullied, because we do not know whether all children of responding parents rode the bus. The percentage of parents who reported their child was bullied at sporting events was the smallest (0.8 percent), and the percentage of parents who reported their child was bullied online was only 1.8 percent.

Table 19
Percent of Parents Indicating Their Child was Bullied by Location

Location of Bullying	Number	Percent
In classroom	7,413	11.9
Other location at school	5,869	9.4
On the bus	3,149	5.1
After school	1,750	2.8
On-line/texting during school	1,129	1.8
At sporting events	469	0.8

Individual students may have been bullied in more than one of these locations. Table 20 presents a summary of the number of different locations where parents reported that their child had been bullied. Most parents who indicated their child was bullied also indicated that bullying occurred in only one location.

Table 20
Number of Locations in Which Parents Reported Their Child Being Bullied

Number of Locations	Number of Parents	Percentage of Percent
0	49,842	80.1
1	7,302	11.7
2	3,301	5.3
3	1,279	2.1
4	345	0.6
5	86	0.1
6	38	0.1

Conclusions

- In 2015 parental satisfaction in all areas assessed by the survey: Learning Environment (87.6 percent), Home and School Relations (73.1 percent), and the Social and Physical Environment (85.3 percent) is similar to the levels reported in 2014.
- Parental satisfaction with the Home and School Relations for their child's school in 2015 (73.1 percent) increased only slightly from 2014 (71.7). The decrease from the 2013 level of satisfaction (83.3 percent) was not recovered. The percentage of parents who did not indicate a level of satisfaction with home and school relations in 2015 (12.5 percent) did not differ markedly from 2014 (13.7 percent), both of which are dramatic increases from 2013 (3.4 percent).
- Parents of elementary school students are more satisfied than parents of either middle or high school students, which do not differ from one another in their levels of satisfaction.
- Parental work schedule continues to be the largest impediment to parental involvement in school activities.
- The percentage of parents who reported that their child was bullied at school was 19.1, with the most frequent location of the bullying being in the classroom.

PART FOUR
Results of the 2015 Teacher Survey

Teacher Survey Methodology

The teacher surveys were administered online to all teachers in all grade levels. A link to the survey was available on the front page of the South Carolina Department of Education (SCDE) web-site. The teacher survey was available during the same period of time that the parent survey was available, from March 4, 2015 to April 3, 2015.

The 2015 teacher survey contained a total of 81 questions. Items included in the teacher survey are included in Appendix B. Seventy-two questions were designed to elicit information on teacher perceptions with respect to four aspects of their school. Three of these were in common with the parent survey, though the content of specific items differed from the parent survey: learning environment, home and school relations, and the physical and social environment of their school. The last aspect of the school assessed on the teacher survey was the professional working environment of the school. For each of these areas, teachers were asked to respond to individual statements using one of the following responses: Strongly Disagree, Disagree, Agree, Strongly Agree, or Don't Know.

Additional questions obtained the race, gender, teacher preparation, the highest degree obtained, whether the teacher was national board certified, the total number of years of experience, and the number of years spent at the current school. Teachers were also asked to identify their current school.

Responses to the parent survey were returned to their child's school and from the school were returned to the scoring contractor. In this process, school identification was also made for the parent surveys. Summary results of the parent survey for each school could then be associated with either the individual results of each teacher from the same school or summary results of teachers for the same school.

Method of Analysis

In previous years, the results of the parent survey were compared to the absolute ratings of the school. Since there were no absolute ratings in 2015, this was not possible. Instead, this analysis used the parent survey item that addressed the overall learning environment of the school to create five groups (quintiles) of schools based on parent perceptions of the learning environment in the school.

To accomplish this, the mean score for the overall learning environment of the school were computed, then schools were ordered from high to low based on this mean. These mean scores were categorized into quintiles. Quintiles divide a set of ordered scores into five groups, with each group containing as near to 20 percent of the scores as possible. The lowest 20 percent of school learning environment scores are in the 1st quintile, and represent the schools with the lowest level of parent satisfaction with the school learning environment. The next 20 percent are in the 2nd quintile, etc. The highest 20 percent of learning environment scores are in the 5th quintile, and represent the schools with the highest level of parent satisfaction with the school learning environment.

Teacher Perceptions of Their School

Responses to the items that describe the teacher respondents are presented in Table 21. The overwhelming majority of teachers responding in 2015 were female (82.8 percent), white (81.0 percent), and have a Master's degree (62.6 percent). Approximately 14.3 percent of responding teachers are national board certified. The largest percentage of responding teachers had between 7 and 15 years of experience as a teacher. With respect to the number of years teachers had spent at their current school,

teachers most frequently reported being at their school from 1 to 3 years, followed by 7 to 15 years. Teachers at a school from 1 to 3 years would include newly hired teachers as well as more experienced teachers who chose to change schools for some reason.

Table 21

Respondents to the 2015 Teacher Survey
(n=45,177)

Gender:	
Male	**17.1%**
Female	**82.8%**
Race:	
African-American	**13.2%**
Caucasian/White	**81.0%**
Hispanic	**1.0%**
All Other	**4.8%**
National Board Certified:	
Yes	**14.3%**
No	**85.7%**
Years of Experience as a Teacher:	
1-3 years	**14.0%**
4-6 years	**10.9%**
7-15 years	**33.9%**
15-25 years	**25.6%**
26 or more years	**15.6%**
Years Teaching at Current School:	
1-3 years	**35.0%**
4-6 years	**16.5%**
7-15 years	**32.6%**
15-25 years	**11.4%**
26 or more years	**4.5%**
Teacher Preparation:	
Bachelor's degree program	**30.6%**
5th year program	**57.0%**
Master's degree program	**6.6%**
Alternative Certification	**5.8%**
Highest Degree:	
Bachelor's	**31.3%**
Master's	**62.6%**
Doctorate	**1.6%**
Other	**4.6%**

A. Learning Environment

The responses of teachers to selected questions regarding the learning environment of their school are presented in Table 22. Overall, 89.3 percent of teachers agreed or strongly agreed that they were satisfied with the learning environment of their school. Teachers responded most favorably to the first nine items of the survey, which primarily assess whether they feel the instruction provided students at their school is effective. Two additional items that had high percentages of teachers that agreed or strongly agreed were that their school has high expectations of teachers (94.6 percent), and that teachers respect one another (93.2 percent). The item that the lowest percentage of teachers agreed or strongly agreed with was item 10, that the level of staff and teacher morale is high (78.2 percent).

Table 22
Percent of Teachers with each Response to Learning Environment Questions

Social and Physical Environment Questions	Agree or Strongly Agree	Disagree or Strongly Disagree	Don't Know
1. My school provides challenging instructional programs for students.	96.9	2.6	0.2
2. Teachers at my school effectively implement the state curriculum standards.	97.9	1.1	0.7
3. Teachers at my school focus instruction on understanding, not just memorizing facts.	96.7	2.2	0.5
4. Teachers at my school have high expectations for students' learning.	96.5	2.5	0.6
5. There is a sufficient amount of classroom time allocated to instruction in essential skills.	92.6	6.2	0.5
6. Student assessment information is effectively used by teachers to plan instruction.	95.4	3.5	0.6
7. Effective instructional strategies are used to meet the needs of low achieving students.	93.6	5.1	0.5
8. My school offers effective programs for students with disabilities.	92.7	5.5	1.2
9. Instructional strategies are used to meet the needs of academically gifted students.	93.4	5.0	1.3
10. The level of teacher and staff morale is high at my school.	78.2	21.2	0.1
11. Teachers respect each other at my school.	93.2	5.9	0.2
12. Teachers at my school are recognized and appreciated for good work.	84.6	13.9	0.2
13. Students at my school are motivated and interested in learning.	84.8	14.3	0.1
14. There are sufficient materials and supplies available for classroom and instructional use.	87.8	11.1	0.3
15. Our school has a good selection of library and media material.	90.8	6.7	1.7
16. Our school has sufficient computers for instructional use.	78.6	19.9	0.5

17. Computers are used effectively for instruction at my school.	88.0	10.5	0.5
18. There are relevant professional development opportunities offered to teachers at my school.	89.0	10.0	0.4
19. The school administration communicates clear instructional goals for the school.	91.1	8.1	0.2
20. The school administration sets high standards for students.	91.9	7.1	0.2
21. The school administration has high expectations for teacher performance.	94.6	3.5	0.3
22. The school administration provides effective instructional leadership.	87.3	11.3	0.2
23. Student assessment information is used to set goals and plan programs for my school.	93.8	4.7	0.5
24. Teacher evaluation at my school focuses on instructional improvement.	91.1	7.2	0.8
25. School administrators visit classrooms to observe instruction.	91.1	7.2	0.5
26. The school administration arranges for collaborative planning and decision making.	89.1	9.5	0.4
27. I am satisfied with the learning environment in my school.	**89.3**	**9.8**	**0.2**

B. Home and School Relations

Table 23 presents the results of questions of the teacher survey that address home and school relations. Overall, 81.9 percent of teachers are satisfied with home and school relations. The items with the largest percentage of teachers agreeing or strongly agreeing were the first two items, which indicate that parents are aware of school policies and school activities. Two additional items with large percentage of teachers that agree or strongly agree are that parents support instructional decisions regarding their children (88.3 percent) and understand the school's instructional programs (88.2 percent). The item with the smallest percentage of teachers that agree or strongly agree is that parents participate as volunteers in the classroom (64.5 percent). As previously reported, 34.1 percent of parents indicated that they volunteered at school (Table 15), while 51.7 percent of parents indicated that they would like to visit their child's classroom during the school day (Table 16).

Table 23
Percent of Teachers with each Response to Home and School Relations Questions

Home and School Relations Questions	Agree or Strongly Agree	Disagree or Strongly Disagree	Don't Know
1. Parents at my school are aware of school policies.	93.9	5.5	0.4
2. Parents at my school know about school activities.	95.3	4.0	0.4
3. Parents at my school understand the school's instructional programs.	88.2	10.5	0.6
4. Parents at my school are interested in their children's schoolwork.	82.2	17.1	0.4
5. Parents at my school support instructional decisions regarding their children.	88.3	10.6	0.6
6. Parents attend conferences requested by teachers at my school.	84.7	14.0	1.0
7. Parents at my school cooperate regarding discipline problems.	86.6	12.1	0.8
8. Parents attend school meetings and other school events.	78.3	20.7	0.6
9. Parents participate as volunteer helpers in the school or classroom.	64.5	31.7	3.5
10. Parents are involved in school decisions through advisory committees.	73.9	13.4	2.8
11. I am satisfied with home and school relations.	**81.9**	**17.2**	**0.4**

C. Social and Physical Environment

Table 24 presents the results of questions of the teacher survey that address the social and physical environment of the school, including three items on bullying. Overall, 91.0 percent of teachers were satisfied with the social and physical environment of their school. The three items with the largest percentage of teachers that agree or strongly agree are items that indicated the teachers' level of safety at the school. Sixteen (16) percent of teachers indicated that they have been bullied by another adult at their school, and fourteen (14) percent of teachers indicated that they have been bullied by a student at their school. Eighty-eight (88) percent of teachers indicated that they have been provided professional guidance on how to assist in preventing and/or dealing with bullying.

Table 24
Percent of Teachers with each Response to Social and Physical Environment Questions

Social and Physical Environment Questions	Agree or Strongly Agree	Disagree or Strongly Disagree	Don't Know
1. The grounds around my school are kept clean.	94.8	4.5	0.5
2. The hallways at my school are kept clean.	95.2	3.9	0.6
3. The bathrooms at my school are kept clean.	88.9	9.7	0.6
4. The school building is maintained well and repaired when needed.	90.8	8.1	0.5
5. There is sufficient space for instructional programs at my school.	86.9	11.8	0.5
6. Students at my school behave well in class.	85.8	13.6	0.2
7. Students at my school behave well in the hallways, in the lunchroom, and on school grounds.	82.5	16.6	0.6
8. Rules and consequences for behavior are clear to students.	86.8	12.7	0.2
9. The rules for behavior are enforced at my school.	93.7	5.6	0.3
10. I have been bullied by an adult at this school.	15.9	79.3	4.6
11. I have been bullied by a student at this school.	14.2	81.0	5.8
12. My school or district provides me with training to assist in preventing and/or dealing with bullying	88.4	11.0	0.6
13. The rules about how students should behave in my school are fair.	84.0	14.8	0.2
14. I feel safe at my school before and after school hours.	95.3	3.2	0.6
15. I feel safe at my school during the school day.	96.3	2.3	0.5
16. I feel safe going to or coming from my school.	96.6	1.6	0.6
17. Students from different backgrounds get along well at my school.	93.5	5.2	0.4
18. Teachers and students get along well with each other at my school.	95.5	3.5	0.2
19. Teachers at my school collaborate for instructional planning.	79.8	6.7	0.4
20. I am satisfied with the social and physical environment at my school.	**91.0**	**8.3**	**0.1**

D. Teacher Working Conditions

Overall, 89.4 percent of teachers were satisfied with the working conditions at their school. Teachers were most satisfied with their familiarity with local, state, and national policies that affect teaching and learning (item 6), and with communication technology (item 3). Also noteworthy is that 91.9 percent of teachers agreed or strongly agreed that their decisions regarding instruction and student progress were supported, and that 91.7 percent of teachers agreed or strongly agreed that teachers at their school were encouraged to develop innovative solutions to problems. The items with the lowest percentages of teachers who agreed or strongly agreed dealt with class size and non-instructional duties that interfered with teaching.

Table 25
Percent of Teachers with each Response to Working Conditions Questions

Working Conditions Questions	Agree or Strongly Agree	Disagree or Strongly Disagree	Don't Know
1. I have sufficient space in my classroom to meet the educational needs of my students.	87.0	10.9	1.9
2. My non-instructional duties do not interfere with my essential role of educating students.	82.2	16.5	1.0
3. I have access to reliable communication technology, including phone, fax, and e-mail.	96.4	3.0	0.2
4. I feel supported by administrators at my school.	88.6	10.9	0.2
5. The faculty and staff at my school have a shared vision.	90.8	8.3	0.2
6. I am familiar with local, state, and national policies and how they affect teaching and learning.	97.1	2.1	0.3
7. Local, state, or national policies assist me in meeting the educational needs of my students.	88.0	10.6	0.8
8. The school leadership makes a sustained effort to address teacher concerns.	86.3	13.0	0.3
9. My decisions in areas such as instruction and student progress are supported.	91.9	7.1	0.6
10. Teachers at my school are encouraged to develop innovative solutions to problems.	91.7	7.6	0.3
11. I feel comfortable raising issues and concerns that are important to me.	82.8	16.7	0.2
12. Sufficient resources are available to allow teachers to take advantage of professional development activities.	88.8	10.0	0.4
13. My class sizes allow me to meet the educational needs of my students.	79.9	17.5	2.2
14. I am satisfied with my current working conditions.	**89.4**	**10.0**	**0.2**

E. Results of the Teacher Survey Related to Parent Perceptions of the School Learning Environment

The relationship between teacher perceptions of the overall learning environment of the school and parent perceptions of the overall learning environment of the school is presented in Table 26. The numbers in parentheses in the table are the percentages of teachers within each column of the table. Each column represents a different group of schools based on the parent perceptions of the learning environment in the school.

Consider the pattern of the percentages of teachers responding that they strongly agreed that they were satisfied with the overall learning environment of the school as the parent rating increased from lowest to highest. The percentage of parents who strongly agreed increased with each increase in the parental perception of the school. Now considering all other rows in the table, within each row, the percentage of teachers decreases as the parental perception of the school becomes more favorable. From this perspective, the one row that defines the pattern of teacher perception of the school learning environment in relation to parental perception of the school learning environment is the row of the table associated with teachers who strongly agreed that they were satisfied with the learning environment. This one row includes 62 percent of teacher responses, so it represents the opinions of the majority of teachers. The simple summary of this relationship is that as parents perceive the school more favorably, so do teachers.

Table 26
Teacher Perceptions of School Learning Environment by Parental Perception of the School Learning Environment

Teacher Rating of the Learning Environment	Parent Rating of the School Learning Environment					All Teacher Responses
	1st Fifth (Lowest)	2nd Fifth	3rd Fifth	4th Fifth	5th Fifth (Highest)	
Strongly Disagree	608* (7**)	336 (4)	245 (3)	166 (2)	74 (1)	1429 (3)
Disagree	918 (11)	665 (8)	505 (6)	307 (4)	170 (2)	2565 (6)
Agree	2993 (37)	2801 (32)	2394 (28)	2160 (25)	1408 (19)	11756 (28)
Strongly Agree	3656 (45)	4961 (56)	5502 (64)	5891 (69)	5619 (77)	25629 (62)

* Number of Teacher Responses

** Percent of Responses within each column (may not add to 100 due to rounding).

A similar relationship exists between teacher perceptions of home and school relations of the school with parental perceptions of the school learning environment (Table 27). The percentage of teachers who strongly agreed that they were satisfied with the home and school relations at their school increased as the parental perceptions of the learning environment of the school became more favorable. The pattern of percentages of teachers within any other row of the table consistently decreases as the parental perception the school learning environment becomes more favorable. The row of the table with the largest percentage of teacher responses (strongly agree), most clearly defines the relationship between

teacher perceptions of home and school relations and parental perceptions of the school learning environment; as parents view the learning environment of the school more favorably, teachers view the home and school relations for their school more favorably.

Table 27
Teacher Perceptions of Home and School Relations by Parental Perception of the School Learning Environment

Teacher Rating of Home and School Relations	Parent Rating of the School Learning Environment					All Teacher Responses
	Lowest Quintile	2nd Quintile	3rd Quintile	4th Quintile	Highest Quintile	
Strongly Disagree	902* (11**)	455 (5)	303 (3)	195 (2)	91 (1)	1946 (5)
Disagree	1875 (23)	1368 (16)	954 (11)	644 (8)	316 (4)	5157 (12)
Agree	3395 (42)	3930 (45)	3631 (42)	3251 (38)	2114 (29)	16321 (39)
Strongly Agree	2002 (24)	3019 (34)	3789 (44)	4435 (52)	4756 (65)	18001 (43)

* Number of Teacher Responses
** Percent of Responses within each column (may not add to 100 due to rounding).

The same relationship occurs between teacher perceptions of the social and physical environment of the school with parental perceptions of the school learning environment (Table 28). The percentage of teachers who strongly agreed that they were satisfied with the social and physical environment of their school increased as the parental perceptions of the learning environment of the school became more favorable. The pattern of percentages of teachers within any other row of the table consistently decreases as the parental perception the school learning environment becomes more favorable. The row of the table with the largest percentage of teacher responses (strongly agree), most clearly defines the relationship between teacher perceptions of the social and physical environment of their school and parental perceptions of the school learning environment; as parents view the learning environment of the school more favorably, teachers view the social and physical environment of their school more favorably.

Table 28
Teacher Perceptions of the Social and Physical Environment of Their School by Parental Perception of School Learning Environment

Teacher Rating of the Social and Physical Environment	Parent Rating of the School Learning Environment					All Teacher Responses
	Lowest Quintile	2nd Quintile	3rd Quintile	4th Quintile	Highest Quintile	
Strongly Disagree	457* (6**)	262 (3)	187 (2)	121 (1)	78 (1)	1105 (3)
Disagree	862 (11)	576 (7)	446 (5)	275 (3)	154 (2)	2313 (6)
Agree	3254 (40)	2907 (33)	2582 (30)	2240 (26)	1482 (20)	12465 (30)
Strongly Agree	3640 (44)	5054 (57)	5472 (63)	5589 (69)	5577 (76)	24702 (60)

* Number of Teacher Responses
** Percent of Responses within each column (may not add to 100 due to rounding).

The relationship between teacher perceptions of the working conditions in their school and parental perceptions of the school learning environment is parallel to three previous results: as parent perception of the school learning environment increases, teacher perceptions of their working conditions increase (Table 29).

Table 29
Teacher Perceptions of their Working Conditions by Parental Perception of School Learning Environment

Teacher Rating of their Working Conditions	Parent Rating of the School Learning Environment					All Teacher Responses
	Lowest Quintile	2nd Quintile	3rd Quintile	4th Quintile	Highest Quintile	
Strongly Disagree	571* (7**)	359 (4)	295 (3)	198 (2)	125 (2)	1548 (4)
Disagree	776 (9)	636 (7)	559 (6)	420 (5)	203 (3)	2594 (6)
Agree	3038 (37)	2930 (33)	2652 (30)	2414 (28)	1698 (23)	12732 (31)
Strongly Agree	3836 (47)	4883 (55)	5192 (60)	5514 (65)	5277 (72)	24702 (59)

* Number of Teacher Responses
** Percent of Responses within each column (may not add to 100 due to rounding).

This same pattern exists between most of the items of the parent survey and parental perceptions of the school learning environment.

F. Results of Teacher Survey Questions on Bullying

The teacher survey asked teachers if they had been bullied by a student or by an adult. Table 30 presents a summary of the responses to these questions. Sixteen (16) percent of teachers agreed that they were bullied by another adult at their school, while 14 percent of teachers indicated that they were bullied by a student at their school. Although not presented here, these results are consistent for teachers in elementary, middle, and high school.

Table 30
Teacher Perceptions of Being Bullied

Teacher Response	Bullied by Another Adult at School (%)	Bullied by a Student at School (%)
Strongly Disagree	31,318 (70)	32,903 (73)
Disagree	4,033 (9)	3,345 (7)
Agree	2,106 (5)	1,614 (4)
Strongly Agree	5,003 (11)	4,288 (10)
Don't Know	2,084 (5)	2,627 (6)

* Number of Teacher Responses
** Percent of Responses within each column (may not add to 100 due to rounding).

Conclusions

- Overall, 89.3 percent of teachers were satisfied with the learning environment of their school, as compared to 87.6 percent of parents.
- Overall, 81.9 percent of teachers were satisfied with home and school relations, as compared to 73.1 percent of parents.
- Overall, 91.0 percent of teachers were satisfied with the social and physical environment of their school, as compared to 85.3 percent of parents.
- For nearly all items, as parent satisfaction with the learning environment of their child's school increased, teacher satisfaction with the learning environment, home and school relations, and the social and physical environment of their school increased.
- The percentage of teachers who reported that they were bullied by a student at their school was approximately 14 percent.
- The percentage of teachers who reported that they were bullied by another adult at their school is approximately 16 percent.

APPENDIX A

The 2015 Parent Survey

4803036903

South Carolina
Parent Survey

DIRECTIONS

- Correct Mark:
- Incorrect Mark:
- Erase completely to change.
- Use a No. 2 pencil only.
- Fill in bubble completely.
- Do not fold or staple.

Parents in South Carolina who have children in selected grades are being asked to complete this survey. This survey asks you how you feel about your child's school. Since this survey will be used to help make your child's school a better place, it is very important to tell us exactly what you think. Your answers will be kept private. The school will get a summary of the survey results.

Directions: Read each statement. Decide if you agree, mostly agree, mostly disagree or disagree with the statement. Then darken the bubble beside each statement. Do not write your name or address on this survey.

Learning Environment	Strongly Disagree	Disagree	Agree	Strongly Agree	Don't Know
1. My child's teachers give homework that helps my child learn.					
2. My child's school has high expectations for student learning.					
3. My child's teachers encourage my child to learn.					
4. My child's teachers provide extra help when my child needs it.					
5. I am satisfied with the learning environment at my child's school.					
Home-School Relations					
1. My child's teachers contact me to say good things about my child.					
2. My child's teachers tell me how I can help my child learn.					
3. My child's teachers invite me to visit my child's classrooms during the school day.					
4. My child's school returns my phone calls or e-mails promptly.					
5. My child's school includes me in decision-making.					
6. My child's school gives me information about what my child should be learning in school.					
7. My child's school considers changes based on what parents say.					
8. My child's school schedules activities at times that I can attend.					
9. My child's school treats all students fairly.					
10. The principal at my child's school is available and welcoming.					
11. I am satisfied with home-school relations at my child's school.					
Social and Physical Environment					
1. My child's school is kept neat and clean.					
2. My child's teachers care about my child as an individual.					
3. Students at my child's school are well-behaved.					
4. My child feels safe at school.					
5. My child's teachers and school staff prevent or stop bullying at school.					
6. My child's school has an anti-bullying program to prevent or deal with bullying.					
7. I am satisfied with the social and physical environment at my child's school.					

Please tell us if you do the following:	I do this	I don't do this, but I would like to	I don't do this, and I don't care to	The school does not offer this activity/event
1. Attend Open Houses or parent-teacher conferences.				
2. Attend student programs or performances.				
3. Volunteer for the school (bake cookies, help in office, help with school fundraising, etc.).				
4. Go on trips with my child's school (out-of-town band contest, field trip to the museum, etc.).				
5. Participate in School Improvement Council meetings.				
6. Participate in Parent-Teacher-Student Organizations (PTA, PTO, etc.).				
7. Participate in school committees (textbook committee, spring carnival committee, etc.).				
8. Attend parent workshops (how to help my child with school work, how to talk to my child about drugs, effective discipline, etc.).				

Go on to next page.

The 2015 Parent Survey

4860036906

Please tell us if you do the following:

	I do this	I don't do this, but I would like to	I don't do this, and I don't care to
1. Visit my child's classrooms during the school day.	○	○	○
2. Contact my child's teachers about my child's school work.	○	○	○
3. Limit the amount of time my child watches TV, plays video games, surfs the Internet, etc.	○	○	○
4. Make sure my child does his/her homework.	○	○	○
5. Help my child with homework when he/she needs it.	○	○	○

Please mark if each of the following are True or False:

	True	False
1. Lack of transportation reduces my involvement.	○	○
2. Family health problems reduce my involvement.	○	○
3. Lack of available care for my children or other family members reduces my involvement.	○	○
4. My work schedule makes it hard for me to be involved.	○	○
5. The school does not encourage my involvement.	○	○
6. Information about how to be involved either comes too late or not at all.	○	○
7. I don't feel like it is appreciated when I try to be involved.	○	○

Please rate your school on:

	Very Good	Good	Okay	Bad	Very Bad
1. The school's overall friendliness.	○	○	○	○	○
2. The school's interest in parents' ideas and opinions.	○	○	○	○	○
3. The school's efforts to get important information from parents.	○	○	○	○	○
4. The school's efforts to give important information to parents.	○	○	○	○	○

Please answer the following questions about your child:

1. What grade is your child in? ○ 3rd ○ 4th ○ 5th ○ 6th ○ 7th ○ 8th ○ 9th ○ 10th ○ 11th
2. What is your child's gender? ○ Male ○ Female
3. What is your child's race/ethnicity? ○ African American/Black ○ Hispanic ○ Asian American/ Pacific Islander ○ Caucasian/White ○ Native American ○ Other
4. What grades did your child receive on his/her last report card? ○ All or mostly A's and B's ○ All or mostly Cs and D's ○ All or mostly Bs and C's ○ All or mostly Ds and F's
5. Has your child been bullied at school this year? ○ Yes ○ No ○ Don't know
6. If yes, was your child bullied: (Check all that apply) ○ In classroom ○ Other location at school ○ At sporting events ○ On-line/texting during school ○ On the bus ○ After school
7. If yes, was your child bullied: (Check all that apply) ○ Physically ○ Verbally ○ Both

Bullying is when 1 or more students tease, threaten, spread rumors about, hit, shove, or hurt another student over and over again physically. It is not bullying when 2 students of about the same strength or power argue or fight or tease each other in a friendly way.

Please answer the following questions about yourself. We are asking these questions because we want to be sure that schools are including all parents. For each question, mark only one answer. Your answers will be kept private.

1. What is your gender? ○ Male ○ Female
2. What is your race/ethnicity? ○ African American/Black ○ Hispanic ○ Asian American/ Pacific Islander ○ Caucasian/White ○ Native American ○ Other
3. What is the highest level of education you have completed?
   ○ Attended elementary/high school ○ Earned Associate Degree ○ Earned college degree
   ○ Completed high school/GED ○ Attended college/training program ○ Postgraduate study and/or degree
4. What is your family's total yearly household income?
   ○ Less than $15,000 ○ $25,000 - $34,999 ○ $55,000 - $75,000
   ○ $15,000 - $24,999 ○ $35,000 - $54,999 ○ More than $75,000

Thank you very much for completing this survey!

DO NOT MARK IN THIS AREA

APPENDIX B

The 2015 Teacher Survey

Response options provided were:

Strongly Disagree
Disagree
Agree
Strongly Agree
Don’t Know

Learning Environment

1. My school provides challenging instructional programs for students.
2. Teachers at my school effectively implement the State Curriculum Standards.
3. Teachers at my school focus instruction on understanding, not just memorizing facts.
4. Teachers at my school have high expectations for students' learning.
5. There is a sufficient amount of classroom time allocated to instruction in essential skills.
6. Student assessment information is effectively used by teachers to plan instruction.
7. Effective instructional strategies are used to meet the needs of low achieving students.
8. My school offers effective programs for students with disabilities.
9. Instructional strategies are used to meet the needs of academically gifted students.
10. The level of teacher and staff morale is high at my school.
11. Teachers respect each other at my school.
12. Teachers at my school are recognized and appreciated for good work.
13. Students at my school are motivated and interested in learning.
14. There are sufficient materials and supplies available for classroom and instructional use.
15. Our school has a good selection of library and media material.
16. Our school has sufficient computers for instructional use.
17. Computers are used effectively for instruction at my school.
18. There are relevant professional development opportunities offered to teachers at my school.
19. The school administration communicates clear instructional goals for the school.
20. The school administration sets high standards for students.
21. The school administration has high expectations for teacher performance.
22. The school administration provides effective instructional leadership.
23. Student assessment information is used to set goals and plan programs for my school.
24. Teacher evaluation at my school focuses on instructional improvement.
25. School administrators visit classrooms to observe instruction.
26. The school administration arranges for collaborative planning and decision making.
27. I AM SATISFIED WITH THE LEARNING ENVIRONMENT IN MY SCHOOL.

APPENDIX B

The 2015 Teacher Survey

Social and Physical Environment of the School

1. The grounds around my school are kept clean.
2. The hallways at my school are kept clean.
3. The bathrooms at my school are kept clean.
4. The school building is maintained well and repaired when needed.
5. There is sufficient space for instructional programs at my school.
6. Students at my school behave well in class.
7. Students at my school behave well in the hallways, in the lunchroom, and on school grounds.
8. Rules and consequences for behavior are clear to students.
9. The rules for behavior are enforced at my school.
10. The rules about how students should behave in my school are fair.
11. I have been bullied by an adult at this school.
12. I have been bullied by a student at this school.
13. My school or district provides me with training to assist in preventing and/or dealing with bullying.
14. I feel safe at my school before and after school hours.
15. I feel safe at my school during the school day.
16. I feel safe going to or coming from my school.
17. Students from different backgrounds get along well at my school.
18. Teachers and students get along well with each other at my school.
19. Teachers at my school collaborate for instructional planning.
20. I AM SATISFIED WITH THE SOCIAL AND PHYSICAL ENVIRONMENT AT MY SCHOOL.

Home and School Relations

1. Parents at my school are aware of school policies.
2. Parents at my school know about school activities.
3. Parents at my school understand the school's instructional programs.
4. Parents at my school are interested in their children's schoolwork.
5. Parents at my school support instructional decisions regarding their children.
6. Parents attend conferences requested by teachers at my school.
7. Parents at my school cooperate regarding discipline problems.
8. Parents attend school meetings and other school events.
9. Parents participate as volunteer helpers in the school or classroom.
10. Parents are involved in school decisions through advisory committees.
11. I AM SATISFIED WITH HOME AND SCHOOL RELATIONS.

APPENDIX B

The 2015 Teacher Survey

Teacher Working Conditions

1. I have sufficient space in my classroom to meet the educational needs of my students.
2. My non-instructional duties do not interfere with my essential role of educating students.
3. I have access to reliable communication technology, including phone, fax, and e-mail.
4. I feel supported by administrators at my school.
5. The faculty and staff at my school have a shared vision.
6. I am familiar with local, state, and national policies and how they affect teaching and learning.
7. Local, state, or national policies assist me in meeting the educational needs of my students.
8. The school leadership makes a sustained effort to address teacher concerns.
9. My decisions in areas such as instruction and student progress are supported.
10. Teachers at my school are encouraged to develop innovative solutions to problems.
11. I feel comfortable raising issues and concerns that are important to me.
12. Sufficient resources are available to allow teachers to take advantage of professional development activities.
13. My class sizes allow me to meet the educational needs of my students.
14. I AM SATISFIED WITH MY CURRENT WORKING CONDITIONS.

Additional Questions:

1. Gender
2. Race
3. National Board Certification
4. Years of Experience as a Teacher
5. Years Teaching at Current School
6. Teacher Preparation
7. Highest Degree Obtained.

The Education Oversight Committee does not discriminate on the basis of race, color, national origin, religion, sex, or handicap in its practices relating to employment or establishment and administration of its programs and initiatives. Inquiries regarding employment, programs and initiatives of the Committee should be directed to the Executive Director 803.734.6148.

FYI

By Peter Elkind

Photograph by Sam Kaplan for Fortune; Styling by Brian Byrne

When Exxon Mobil, GE, Intel, and others pushed for the education standards, they incurred the wrath of Tea Party conservatives and got a painful lesson in modern politics.

In February 2014, two of the world's richest men, Bill Gates and Charles Koch, dined together at a West Coast restaurant.

They made quite the odd couple: the Seattle Microsoft MSFT -1.06% co-founder, now devoting his time and fortune to changing the world, and the Kansas industrialist, still running his private conglomerate while working to shrink government to the size of a pea.

The two discussed many subjects and even touched, diplomatically, on topics they disagree about, such as climate change. There was a second sensitive subject that Gates broached, and it didn't come up by chance. His team at the Gates Foundation had engaged in a process it calls a "faction analysis" and identified Koch as a key opponent on a crucial issue. Gates had a mission that night: He wanted to persuade Koch to change his mind about Common Core.

The two men were bankrolling opposite sides in a raging war over the future of American education. Through his charitable foundation, Gates has spent more than $220 million on the Common Core education standards, aimed at boosting the dismal performance of American children. Starting in 2010, 45 states adopted the benchmarks—which spell out what kids from kindergarten through high school should learn in reading and math—with little controversy. But a backlash ensued, and by early 2014 the standards were under fierce political attack, facing repeal in many states. Koch and his brother David were sponsoring several Tea Party–aligned groups that were fueling the rebellion.

The ABCs of Common Core

- Replacing a hodgepodge of separate guidelines in 50 states, Common Core aims to provide a rigorous, more focused nationwide blueprint for what students should know at each grade in math and English language arts.

- It seeks to improve the U.S.'s poor educational performance relative to other countries', making graduates "college- and career-ready" and assuring the nation's economic competitiveness.

- It's paired with new standardized tests to measure progress in meeting the new standards and to allow comparisons among states.

- Though developed without government funds, Common Core received a boost from Race To The Top, a competition for federal education grants, prompting unexpectedly rapid adoption.

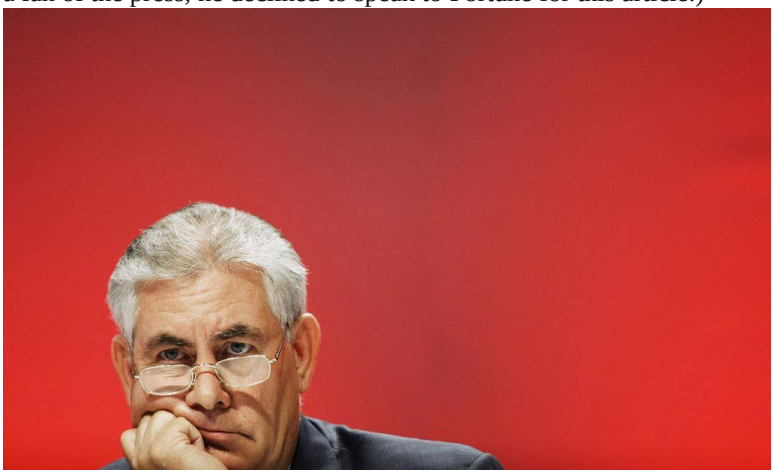
Now Gates tried to convince his dinner companion that opposing Common Core was bad both for Koch Industries, which employs 60,000 Americans, and the rest of U.S. business. "Bill talked to Koch to understand what his concerns were and to explain what he thought was the potential and the promise for the Common Core," says Allan Golston, president of the U.S. program for the Bill & Melinda Gates Foundation. "When someone doesn't believe in what you're doing, it's important to engage with them."

But Koch wasn't willing to engage with Gates on the issue. Instead, like a senator politely brushing off a constituent, he gave Gates the name of one of his staffers who focuses on the subject and suggested Gates call the staffer. The Microsoft billionaire left empty-handed. (Both men declined to discuss their dinner.)

This extraordinary tête-à-tête is just one example of how the war over Common Core has personally engaged—and bedeviled—some of America's most powerful business leaders. Hugely controversial, it has thrust executives into the uncomfortable intersection of business and politics.

In truth, Common Core might not exist without the corporate community. The nation's business establishment has been clamoring for more rigorous education standards—ones that would apply across the entire nation—for years. It views them as desperately needed to prepare America's future workforce and to bolster its global competitiveness. One measure of the deep involvement of corporate leaders: The Common Core standards were drafted by determining the skills that businesses (and colleges) need and then working backward to decide what students should learn.

Organizations such as the Business Roundtable have devoted considerable effort to the initiative. The education chair for that association of CEOs, Exxon Mobil XOM 0.43% chief Rex Tillerson, has played a particularly prominent role. A stern, commanding figure with an Old Testament glare and a chewy Texas drawl, Tillerson is an unlikely person to lead a campaign of persuasion. (Never a fan of the press, he declined to speak to Fortune for this article.)



Rex Tillerson, the CEO of Exxon Mobil, has led the Business Roundtable's advocacy for common core. The staunch Republican has found himself accused of promoting big government.Photo: Susana Vera— Reuters

But Tillerson has taken on the challenge with trademark intensity. He has pressed other CEOs to join the cause, spread the word by appearing at education summits, underwritten TV advertisements, and personally called legislators in multiple states to press for their support. His company went so far as to cut off campaign contributions to some politicians—even those who support the oil and gas industry—who spurned Tillerson's entreaties on Common Core.

Other companies have been much more timid or have retreated in the face of the controversy. For example, General Electric GE 0.99% —once among Common Core's biggest supporters—has fled the fight after becoming a Tea Party target. "There's a somewhat unwritten rule that if you're a CEO, you only get your business involved in an issue that rewards your company in some fashion," says former Intel chief Craig Barrett. Education reform is "such a hot topic," especially as Common Core made it "more of a tar baby," that "it's sometimes difficult to get people enthusiastic," he adds. "A lot of people just sit on the sidelines." Adds Barrett, with exasperation: "It's turned into a political food fight instead of an education discussion … The hope is that rational minds will prevail."

This is a story about the role Big Business has played in the war over Common Core: how a handful of executives helped turn a decades-old ambition for education reform into reality, their fumbling bewilderment at finding themselves assailed by opposition they didn't expect or understand, and how they've regrouped and rallied to defend what they wrought. It's a high-stakes conflict that has generated breathtaking political flip-flops (see "The Flip-floppers and the Wafflers") and upended traditional alliances, turning natural bedfellows into bitter enemies. It has seen some of the nation's foremost capitalists accused of promoting an "immoral," "freedom-robbing," "socialist agenda," aimed at turning America's children into "mindless drones for the corporate salt mines."

Along the way it has reinforced the depth of a growing divide that once would have seemed inconceivable: the gap between America's most ardent conservatives and Big Business, which the former increasingly view as part of an undifferentiated "establishment" and hence nearly as odious as government. For now, Common Core has established itself in the vast majority of states and seems to be taking root. But the conflict over this issue shares many traits with the crusades over Obamacare, and one of them is this: Its most fervent opponents show not the slightest sign of relenting.

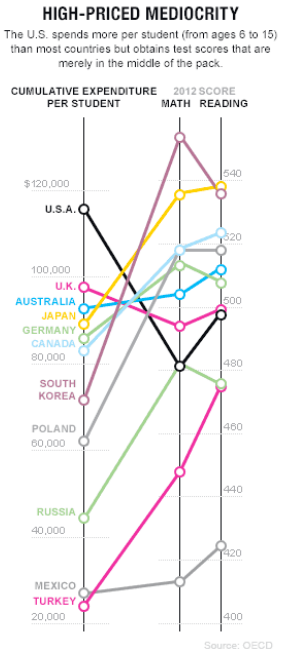
Protesters outside a State Board of Education meeting in Irvine, Calif. in 2014.Photo: Anna Reed—The Orange County Register/Zumapress.com

For decades, CEOs have bemoaned the state of U.S. education—with justification. American 15-year-olds ranked 27th out of 34 industrialized countries in math, and 17th in reading in the most recent international tests. Colleges complain that significant percentages of their entering freshmen require a remedial course. Businesses say they can't find enough skilled U.S. workers.

But the executive mind-set on the issue—favoring consistency, efficiency, and accountability—has clashed with the American tradition of local control. "Why on earth can't we insist on universal standards at least for 9-year-olds?" asked Alcoa AA 1.86% CEO Paul O'Neill at a 1996 education summit convened by then-IBM IBM -0.74% CEO Lou Gerstner and attended by business leaders and 43 governors. "Can't a 9-year-old multiply by nine and get the same answer in all 50 states?"

To CEOs, the issue has always been a no-brainer. In an increasingly global economy, what sense does it make for America to have 50 different sets of education standards? Gerstner helped establish a nonprofit called Achieve Inc. in 1996 to promote education reform. With a board filled with governors and CEOs, the group served, over the next two decades, as a sort of lab for the national standards movement.

The modern era of U.S. education reform actually dates back to 1983, when a commission convened under Ronald Reagan produced a landmark report titled A Nation at Risk. "A rising tide of mediocrity," it warned, "threatens our very future." In response, William Bennett, Reagan's education secretary, promoted mastery of a "common core of worthwhile knowledge, important skills, and sound ideas." (Bennett went so far as to design a full K-12 curriculum.)



Conservatives cheered. But the idea was strictly voluntary—"It remains a matter best left for final decision to state, local, and private authorities," Bennett noted—and it didn't get far.

Uncle Sam's role in education is an exquisitely sensitive issue: Federal law bars the government from dictating education standards or classroom curriculums. And state and local officials—who provide about 90% of public education funding—prize their control.

Reform efforts have needed to dance around that by seeking to persuade 50 states to embrace change voluntarily. "You really can't work this issue on a national level," says Gerstner. "You've got to work it state by state, city by city. It's messy. Unfortunately, it doesn't yield completely to reason, which businesspeople like."

Every president since Reagan has flailed at this issue. George H.W. Bush and Bill Clinton convened special education panels and launched commissions in unsuccessful attempts to establish voluntary national standards. Finally, the Clinton administration was able to pass a watered-down initiative that required states to adopt standards and tests—but left them entirely up to individual states.

Most established tragically low expectations. President George W. Bush's 2002 education reform, "No Child Left Behind," only worsened this problem. It set the impossible requirement that 100% of students be "proficient" in reading and math by 2014, and punished schools that weren't making adequate progress.

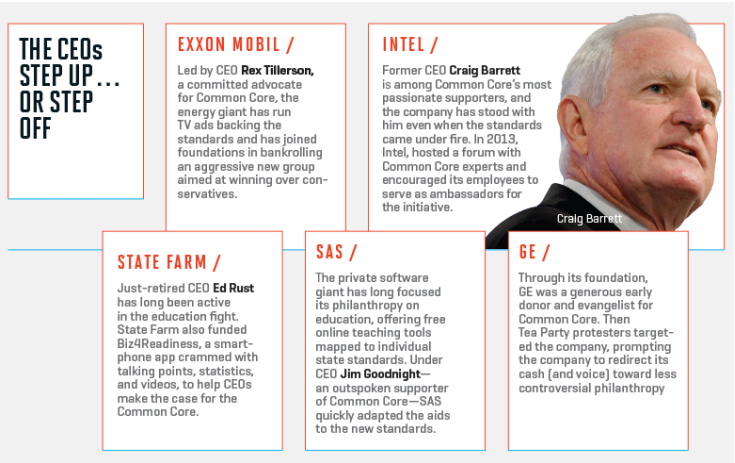


Photo: Nelson Ching—Bloomberg via Getty Images

To bring themselves closer to 100%, many states simply lowered the score needed to pass their tests. The result: In 2007, Mississippi judged 90% of its fourth graders "proficient" on the state's reading test, yet only 19% measured up on a standardized national exam given every two years. In Georgia, 82% of eighth-graders met the state's minimums in math, while just 25% passed the national test. A yawning "honesty gap," as it came to be known, prevailed in most states.

Finally, in April 2009, organizations representing state governors and education chiefs agreed to develop a single set of rigorous standards: the Common Core State Standards Initiative. The ambition was to make all children "college- and career-ready," with the same expectations in Mississippi as in Massachusetts. The standards would spell out what students should learn at each grade level, without dictating curriculum or how it would be taught. They would be accompanied by tough new standardized tests to measure progress in meeting the benchmarks. The tests were the hammer to drive improvement and provide accountability. The goal was universal acceptance. This would allow comparisons among states, help the children who move annually to a new state stay on track, and permit sharing of education ideas, textbooks, and teaching materials.

Promoters of the Common Core took three big steps to smooth adoption. First, they developed standards for only English language arts and math, avoiding the ideological land mines in teaching history and science, such as slavery, evolution, and global warming.

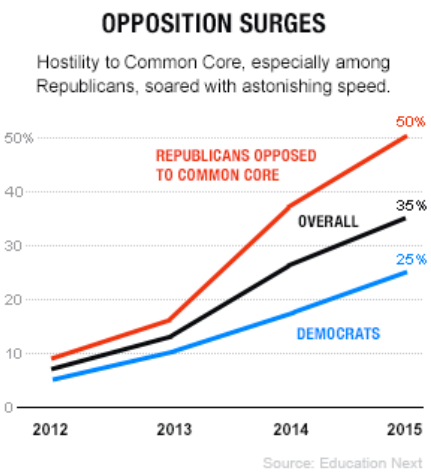
Second, they enlisted Bill Gates, whose foundation had already sunk hundreds of millions into other education initiatives. The Gates Foundation would help bankroll virtually every aspect of Common Core's development, promotion, and implementation. "This is like having a common electrical system," Gates told the Wall Street Journal in 2011. "It just makes sense to me." His spending would be critical—but it would later feed a view among some that one rich man shouldn't have so much say over a national policy.

In the short term, though, Gates' millions helped make possible the third (and most important) step: writing the new standards without a penny from Uncle Sam. "State-led initiative" became advocates' mantra for describing Common Core. "It had been crafted as a local-control issue, and we wanted to keep it that way," says former Intel INTC 1.49% CEO Barrett. "All the groundwork had been done very carefully."

When it came time to draft the provisions, career readiness was a central focus. The writers spent their first two months learning what colleges and businesses wanted high school graduates to know by the time they arrived on their doorstep. From there, the writers "back mapped," crafting grade-by-grade benchmarks to get them there. The resulting standards were then reviewed by teachers' unions, state education officials, academic groups, feedback panels, and independent validation committees. Two drafts were published online, generating 10,000 public comments and prompting further revisions.

Like the CEOs, federal education officials always knew this was treacherous terrain. "There was definitely discussion about whether the feds should get involved because of the potential for political backlash," says Joanne Weiss, a former top deputy in the U.S. Department of Education under Obama. Now a consultant, Weiss acknowledges that the administration walked a tightrope. "The department tried to get involved in a way that just handed money back to the states, that let them do their own thing. We wanted to be both supportive and arm's length—admittedly a hard balancing act." Common Core's architects also worked to avoid any federal taint.

For a time, it all worked according to plan—in fact, far better than anyone had imagined. The new standards rolled out to general praise from educators and endorsements from business groups. The most detailed appraisal (funded with $959,116 in Gates Foundation money) was conducted by the Thomas B. Fordham Institute, a right-leaning Washington think tank. Its 370-page analysis found the Core standards "clearly superior" to those in place in "the vast majority of states."



Forty-five states, more than half of them led by Republican governors, adopted Common Core by the end of 2011—remarkably short order. The only holdouts were Virginia, Alaska, Texas, and Nebraska; Minnesota took up the standards, but only for English.

The Obama administration tried to tiptoe. It didn't attempt to mandate implementation, but it strongly encouraged it. The administration accelerated the process by launching Race to the Top, a competition among the states for $4.35 billion in federal grants. Applicants received 70 points (out of a possible 500) for approving "enhanced standards and high-quality assessments" (most obviously Common Core) by August 2010. In the midst of a deep recession, the cash promoted a quick embrace. The federal Education Department also provided $350 million in grants to two consortia set up by the states to develop the new Common Core tests.

In the 45 states, adoption of the standards, which typically required just a public meeting and approval by the state education board, stirred little notice. "Zero," recalls Tony Bennett, the elected superintendent of public instruction in Indiana when the state signed on. "No controversy. No criticism."

Victory in hand, Common Core advocates turned their energies toward the task of implementation. They didn't foresee that a deep well of opposition was about to erupt. "In a sense the early days almost went too easy for us," Gates would later say. "Everything seemed to be on track … We didn't realize the issue would be confounded."

The "confounders" would turn out to be just the sort of people who today cause fits for billionaires and CEOs used to exercising power through traditional channels: passionate regular folks linked to activist networks with a firm grasp on how to maximize the power of the Internet and social media. Gayle Ruzicka, who volunteers as Utah state president for Phyllis Schlafly's Eagle Forum, had long been battling to preserve local control of schools. Indeed, Ruzicka takes "local control" far beyond where most parents would: She homeschooled all 12 of her children. Ruzicka was deeply concerned by what, in late 2010, she began to hear about Common Core. To her it sounded like, as she puts it, "a backdoor way in to national standards."

States had adopted Common Core, Ruzicka says, "before parents even knew what happened." In retrospect, approving an education transformation without building parental support would turn out to be a huge mistake. It meant that the opposition would mass and organize before many potential allies of the standards even realized they needed to be defended. Ruzicka began gearing up to fight it. She coined a phrase that crystallized her view of the problem with devastating rhetorical force: "Obamacore."

William Bennett

The Common Core standards are "excellent" and a "conservative idea."

Photo: Alex Wong—Getty Images

Ruzicka wasn't alone. In the fall of 2011, an Indiana mom named Heather Crossin became alarmed about changes in how her 8-year-old daughter was being taught math. Her third-grade homework didn't ask her just to solve three times nine—it demanded that she explain the reasoning behind her answer. Crossin was at a loss to help. The principal at her child's school blamed the changes on Common Core.

Crossin, who once served as a legislative assistant to Republican Rep. Dan Burton, began organizing Hoosiers Against Common Core. She approached local Tea Party groups and welcomed the help of national organizations opposed to the standards, including the American Principles Project, where a man named Emmett McGroarty served as education director and as a key figure in the fight. They fed Crossin's group anti–Common Core "white papers," set up its website, helped plot strategy and write leaflets, and even flew in for local rallies and media interviews.

The grass-roots moms' rebellion, fanned through social media and the Tea Party network, quickly gained momentum in multiple states. Says Business Roundtable vice president Dane Linn, then education policy chief for the National Governors Association: "We heard the rumblings in the states. It was like prairie fire after prairie fire."

One of the first shocks for supporters of Common Core came in Indiana. There opponents targeted superintendent Bennett, a conservative Republican whose advocacy of school vouchers, charter schools, tough teacher evaluations, and Common Core had made him a darling of national reformers. In November 2012, Bennett was defeated in a reelection bid by a massively outspent Democratic opponent, a former teacher who had voiced skepticism about the standards. In the same election, Tea Party Republican Mike Pence succeeded Mitch Daniels, the term-limited GOP governor who had backed the standards. A few months later, Indiana delayed its implementation of Common Core.

Tea Party groups soon made Common Core a national rallying cry. In 2013, Glenn Beck took up the cause, declaring it "Communism—we are dealing with evil." That April the Republican National Committee passed a resolution condemning Common Core as an "inappropriate overreach to standardize and control the education of our children so they will conform to a preconceived 'normal.' "

In the hands of opponents, the "state-led" plan, commissioned and adopted voluntarily by nearly all the nation's governors and school chiefs, was recast as a "national takeover of schooling" developed "behind closed doors" by "private trade groups" and, of course, Barack Obama. Indeed, the president's perceived imprimatur was the chief cause of opposition, in the view of some supporters. "If we had a Republican president, I don't think we would have had this backlash," says Cheryl Oldham, a former George W. Bush administration official who is now vice president of education policy at the U.S. Chamber of Commerce. "It was because it was viewed to have been something that was Obama-led and -driven and forced on everyone. That just fueled a lot of the pushback."

Critics claimed that Obama's Race to the Top funding had "forced" states to adopt Common Core. In fact, the federal government has a long history of granting states money to write standards. And the agreement that governors signed to develop Common Core explicitly welcomed Uncle Sam's cash, acknowledging an "appropriate federal role in supporting this state-led effort" with incentives. Obama would give critics further ammunition by repeatedly praising the higher standards, even as he took pains to note that they'd been developed "not by Washington."

To be sure, there were legitimate questions: Was the implementation schedule too rushed for such a massive classroom change? Were the standards too tough—or not rigorous enough? Some people were suspicious of business's support of the standards, charging that they were aimed at turning out corporate "drones" and "minimally educated worker bees." Did the English benchmarks emphasize analysis of "informational texts" too much, at the expense of literature? Would it all be too costly? Some questioned the premise that smarter standards would boost learning at all.

Gayle Ruzicka

"Once the standards become national like Common Core, parents have no say ... It's an agenda to control education from the top. People with money, like Gates, want to have a lot of control over what's being taught."

Photo: Rick Bowmer—AP

Teachers' union leaders—who had endorsed Common Core at the outset—complained bitterly about its rollout, especially objecting to the immediate use of new standardized tests in their performance evaluations. This criticism of "high-stakes testing" would later bring Common Core under assault from both ends of the political spectrum.

But it wasn't wonky details that threatened to unravel the initiative. It was the most extreme claims, which spread like wildfire. Schlafly called Common Core an Obama scheme—in collaboration with book publishers and Gates—to "dumb down" schoolchildren, "indoctrinate them to accept the left-wing view of America," engage in "active promotion of gay marriage," and "dismantle moral society." Bloggers warned that Common Core would allow the federal government to engage in wholesale data collection on schoolchildren—including iris scans—then sell the information "to the highest bidders." Parents charged that Common Core forced 10th-graders to read pornography out loud in class and required graphic sex-ed instruction. One Florida legislator asserted that the state's Common Core testing will "attract every one of your children to become as homosexual as they possibly can." Never mind that none of those assertions were true.

Common Core was becoming politically radioactive for Republicans. “All of a sudden in 2013, you saw these Common Core repeal bills getting introduced all over the place,” says Fordham Institute president Mike Petrilli. “Those of us for it were caught pretty flat-footed. We realized this thing was at risk.” If somebody didn’t fight back, it appeared, Common Core might go down in flames.

“It is utterly distressing to me to sit and watch these political debates around a subject that is so vitally important to our children, to the future of our country, and competitively,” fumed the silver-haired man in the dark suit and gold tie, waving his arms in exasperation. “And I’m going to tell you, I’m extraordinarily disappointed in my home state. I’ve spent many hours on the telephone during the last legislative session. To no avail. Could not make a dent. So the political forces around this are powerful. But they have to be taken on.”

It was a strange thing indeed to hear Rex Tillerson, CEO of Texas-based Exxon Mobil, bemoaning his impotence at a 2014 panel discussion in Washington, D.C. But such is the frustration of serving on the frontline in this war. Like other CEOs engaged in education reform, Tillerson sees high national standards as a “business imperative.” Companies simply can’t find enough skilled American workers.

But Tillerson articulates his view in a fashion unlikely to resonate with the average parent. “I’m not sure public schools understand that we’re their customer—that we, the business community, are your customer,” said Tillerson during the panel discussion. “What they don’t understand is they are producing a product at the end of that high school graduation.”

Bill Gates

“We view [Common Core] as quite fundamental ... The rollout of Common Core—there’s a lot that’s been learned ... including a mea culpa where in some places our foundation and perhaps others were naive ...”

Photo: Ringo Chiu—Zumapress.com

The Exxon CEO didn’t hesitate to extend his analogy. “Now is that product in a form that we, the customer, can use it? Or is it defective, and we’re not interested?” American schools, Tillerson declared, “have got to step up the performance level—or they’re basically turning out defective products that have no future. Unfortunately, the defective products are human beings. So it’s really serious. It’s tragic. But that’s where we find ourselves today.”

Exxon Mobil’s philanthropy has long been focused on math and science education. Tillerson himself became deeply engaged in the Common Core fight in early 2012, when he became chairman of the education and workforce committee for the Business Roundtable, the powerful Washington, D.C., trade group for 202 big-company CEOs.

But while opponents like Ruzicka and Crossin harnessed the power of the web, Tillerson’s team turned to an older, more genteel form of media—the kind that is better at reaching silver-haired CEOs than, say, blogger moms. In April 2012, Exxon Mobil ran an advertisement during the CBS telecast of the Masters golf tournament. Common Core is “unlocking a better way to prepare our children for college and careers,” the ad argued. The tagline: “Join Exxon Mobil in supporting the Common Core State Standards Initiative.”

It’s hard to say whether the Exxon ad had any impact when it first appeared. But by the time it aired again a year later, it generated a reaction—a deeply hostile one. Glenn Beck responded with a 12-minute polemic, and emails—99% critical, according to ExxonMobil Foundation executive director Pat McCarthy—cascaded in. “Big Businesses Whore for Common Core,” headlined one blog post discussing the ads. Critics began urging a company boycott. Wrote one: “Cut the gas cards up … this is disgusting.”

Even the government expressed frustration. In May 2013, Secretary of Education Arne Duncan scolded executives at a U.S. Chamber of Commerce event for failing to do more to defend Common Core: “I don’t understand why the business community is so passive when these kinds of things happen.”

Many companies stayed on the sidelines. Worse, one staunch supporter—GE (see below)—abandoned the fight.

As the threat grew during 2013, Common Core’s supporters struggled with how to fight back. “Our responses initially were fact-based,” says Achieve’s president, Mike Cohen. “But the opposition’s appeals were more emotional than that. It turned out facts didn’t often matter.”

Everyone wanted to coordinate strategy; supporters considered a national advocacy campaign, including TV ads. But advocates didn't want to reinforce the very notion they were trying to combat. "Having someone from Washington explain that there's not really a conspiracy here doesn't really put the fire out," notes Cohen.

Grass-roots rage had made Common Core a potent issue. Many Republican officials who had backed the standards were now flip-flopping. Presidential aspirants performed some of the most remarkable acrobatics. In Oklahoma in 2013, Common Core supporters enlisted Mike Huckabee, the former governor of next-door Arkansas, to fight a growing repeal movement. Huckabee wrote a two-page letter urging lawmakers to stay the course: "I've heard the argument these standards 'threaten local control' of what's being taught in Oklahoma classrooms. Speaking from one conservative to another, let me assure you this simply is not true." Huckabee called the Common Core standards "near and dear to my heart … something to embrace."

Less than two years later, after announcing plans to seek the Republican presidential nomination, Huckabee had a different view. "We must kill Common Core and restore common sense," he declared on his campaign website. Huckabee was hardly alone in reversing his position.

Conservative politicians cast Common Core as a looming threat to liberty. Even in Texas, which never adopted the standards, the state legislature—in the name of defending local control of education—passed a law in June 2013 making it illegal for any school district to use the Common Core standards. In his successful run for governor, Greg Abbott vowed to "crush" Common Core.

At the Business Roundtable, Tillerson importuned his fellow CEOs at every meeting to "be visible in their support" and "pick up the phone and call key state leaders to voice their support for staying the course," according to Linn, who became the group's education specialist around that time. The Exxon CEO urged them to wield their lobbying and economic clout, especially in states where they operated major facilities with lots of jobs. Tillerson was so persistent that he annoyed some of his CEO peers.

Some companies—such as Intel and Cisco CSCO -0.04% —promoted the standards. But the response from others fell short. The Exxon Mobil CEO simply couldn't move his peers amid the political heat.

Tillerson didn't hesitate to flex his own muscle. In May 2013, after Pennsylvania delayed implementation, he fired off a letter reminding the governor and others that his company had "significant operations" there. Common Core, he advised them, was necessary to give Exxon Mobil "the confidence that the education standards we require for employment will be met by your state's graduates." An education blogger quickly branded this a "Mafia-style letter," and suggested Pennsylvania's governor "may soon wake to a horse's head laying in his bed, which will smell vaguely of gasoline."

Five months later, Tillerson was even less subtle, warning lawmakers contemplating repeal that Exxon Mobil might not hire anyone from states that don't have Common Core. "If I can't find the workforce in the state that I'm in, I will go to the next state and find that workforce," he told NBC's Tom Brokaw in an interview on stage at an education conference. "And I'm going to look in states that are using the Common Core State Standards because I have a high degree of confidence in the kids that graduate under that system."

The persistent advocacy from Exxon's gruff CEO, a staunch Republican and blunt critic of federal regulation, drew some particularly improbable attacks. Tom Borelli, then a leader of Tea Party group FreedomWorks, called Tillerson's support "another example of the Big Business establishment joining ranks with big government to expand centralized control of our lives."

By early 2014, the tide seemed to be turning against Common Core. Indiana became the first state to drop out entirely. South Carolina soon followed. Over the next year state lawmakers would introduce more than 100 bills to limit or halt implementation and 40 to drop Common Core altogether, according to a tally by the National Conference of State Legislators.

Oklahoma would become the site of the most dramatic reversal for Common Core to date. Mary Fallin had won election as the state's first woman governor as a strong supporter of Common Core. She vigorously defended the standards in a January 2014 speech to the National Governors Association.

That spring the Oklahoma legislature overwhelmingly passed a bill that didn't just scrap Common Core. It dictated that the state eventually implement new benchmarks—subject to a 10-point comparison to make sure they didn't even resemble Common Core.

Fallin had not indicated whether she would sign the legislation, and activists descended en masse to persuade her. One group of opponents besieged the Capitol wearing green T-shirts reading common core is not ok. National organizations swamped Fallin's office with thousands of calls urging her to sign the repeal. School administrators and teachers, meanwhile, warned of educational chaos; they had already prepared classroom plans for the fall aligned to Common Core. Business groups urged a veto. Tillerson, in Oklahoma

City to deliver a speech at an energy conference, urged Oklahoma not to retreat from its “prior commitment” to “high and meaningful standards.”

In the end, Fallin sided with Common Core’s opponents. On June 5, 2014, she signed the new legislation, citing the “widespread concerns” that Common Core “gives up local control of Oklahoma’s public schools”—the very concerns she’d previously dismissed.

It was about this time that Tillerson’s company adopted a new policy for its corporate political action committee. ExxonMobil PAC would make no more donations to elected officials actively opposed to Common Core, even those who typically back the company’s principal business interests.

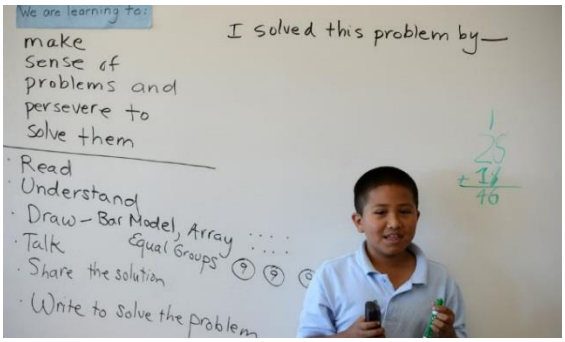
Among the first to be affected: Oklahoma Gov. Fallin. Her campaign committee had received a combined $6,000 in annual donations from ExxonMobil PAC in 2011, 2012, and 2013. In 2014, as she campaigned for reelection, ExxonMobil PAC gave her nothing.

Even before the defeat in Oklahoma, Common Core’s supporters had begun to recognize that they had to step up their defense—and do so in a more localized way. The real fight was occurring in individual states considering repeal. The proponents unrolled a “ground game,” helping launch state advocacy groups with full-time staff and websites, featuring testimonials from local teachers and business leaders supporting Common Core.

Recognizing the need for conservative political and PR savvy, Common Core backers turned to a new nonprofit they’d established, called the Collaborative for Student Success, to “ensure fact-based discussion.” (The group’s funding includes $27.9 million from the Gates Foundation, as well as grants from the ExxonMobil Foundation.) To run it, leaders of eight big foundations hired Karen Nussle, a former Newt Gingrich aide who had become a Washington PR and marketing operative. Nussle assumed the role of “conservative whisperer.” She established a rapid-response operation, to spin news about Common Core and respond fiercely to opponents’ charges.

She also retained William Bennett, one of the fathers of Common Core, as an advisor. In ads and media interviews aimed at calming the right, Bennett bashed the Obama administration for meddling (“That messed up everything,” he tells Fortune), but defended the standards as “still excellent” and a “conservative idea.” America, he told the Manchester Union Leader, needed to adopt “real standards” across the country “so we can stop being the dumb asses of the industrial world.” Nussle says Bennett, now an author and radio talk-show host, makes an ideal advocate because he is “unassailable as a hard-core conservative.” (Still, assail him they have. The conservative blog RedState.com put it this way: “Bill Bennett paid to pimp for Common Core.”)

A few business leaders stepped up their efforts too. State Farm helped fund Biz4Readiness, a smartphone app developed by the Committee for Economic Development, as a sort of electronic cheat sheet for CEOs to use in promoting the standards. It included statistics, talking points, videos, and rebuttals to “common myths” about Common Core. In 2014 the U.S. Chamber of Commerce and Business Roundtable bankrolled a two-month round of ads. They aired on Fox News rather than during a golf tournament.



A fourth-grader in San Pablo, Calif. explains his solution to a math problem for his classmates in 2013. Math teachers are changing the way they deliver lessons to students to adapt to new Common Core curriculum standards. Photo: Kristopher Skinner—TNS/Zumapress.com

The anger against Common Core remained fierce, with politicians facing intense pressure. And yet most state education officials and many teachers continued to view the substance of the standards as extremely valuable. In the face of these opposing positions, an almost-too-easy third way began to emerge.

Instead of dropping out, 27 states simply renamed their education standards. In most cases they tweaked some of the provisions while retaining the vast majority. In Florida, for example, the dreaded Common Core was dead! Long live … the Next Generation Sunshine State Standards!

This provided political cover for Republicans. For their part, supporters of Common Core concluded they had no need to fight such initiatives. (The advocates note that any serious standards will necessarily share many elements with Common Core. Says Rich McKeon, head of the education program at the Helmsley Charitable Trust, a philanthropy that has given millions to support the standards: “It’s hard to get rid of Common Core completely unless we don’t want kids to do a lot of math and writing and deep analytical thinking.”)

The moves lowered the temperature in the fight, and by 2015 the repeal forces seemed to be losing momentum. The battle of Arizona may have been the turning point. The state's Republican governor, Doug Ducey, took office in January 2015 after campaigning against Common Core. Tea Party–backed legislators promptly prepared a bill to dump the standards, which had been embraced under Ducey's GOP predecessor, Jan Brewer.

Opponents took aim at both the standards and their supporters, including former Intel CEO Barrett. Once again the accusations were wild. A Republican, Barrett had retired to Arizona and served as an education adviser to Brewer. A group called Arizonans Against Common Core declared that Barrett "has UN ties!" and advised that "Common Core Science Standards"—actually, there are none, since Common Core doesn't deal with science—"Teach Global Warming!" The repeal bill passed the Arizona house and went to the senate, where Republicans had a big majority.

This time, though, the business community had seemingly learned how to tangle with the organized opposition. The Arizona chamber of commerce—and Barrett—fought back hard. The chamber lobbied furiously, senator by senator, arguing that the state, which had been struggling for years to improve its poor-performing school system, needed the standards to attract jobs. They made the same case to Ducey, a former CEO of Cold Stone Creamery. Soon after, the governor publicly stated that getting rid of the Common Core standards wasn't "necessary" after all. A week later, the Arizona senate voted 16–13 to preserve the standards.

The adversaries of Common Core have no intention of capitulating. In 16 states, it now faces various implementation "reviews." The clock ran out in 2015 without any more states dropping out. But in 2016, legislative assaults will undoubtedly resume.

That said, Common Core has become a reality. Like Obamacare, it's reviled in many quarters. Yet it's increasingly impractical to undo. Countless schools have established curriculums designed around the standards, retrained teachers, and bought new books and materials. Reversing course would require redoing all of that again. Today, 42 states remain officially committed to the Common Core (under whatever name), while South Carolina, Indiana, and Alaska have standards of their own that experts say closely resemble Common Core. After decades of controversy and conflict, a single set of thoughtful, higher standards is shaping the education of most American schoolchildren. (Exxon Mobil is confident enough of the standards' staying power that it has rescinded its policy of withholding campaign contributions to opponents of Common Core.)

It remains unclear how well this grand experiment will meet its ultimate goal: better preparing kids (and our country) for a challenging future. A key element of the Common Core effort—common standardized tests to allow honest assessments of progress—remains unfulfilled, swept back by a wave of parental concern about over-testing and teacher anxiety about being judged too harshly too soon. (Indeed, even the Gates Foundation—a staunch advocate of testing accountability—has urged a two-year moratorium on using new Common Core exams for teacher evaluations or student promotions, citing the need to give everyone time to adjust.)

Of 43 states initially enrolled in one of the two consortia established to develop new Common Core tests, only about half remain. Dropout states, which must use their own tests, have cited teacher and parent concerns, as well as unhappiness with the new exams and their cost. But for states unwilling to repeal the standards, abandoning the tests has also become a way to assert local control—and appease anger about the Common Core.

The first states started using the new tests this year, producing refreshingly honest—if predictably dismal—results on student proficiency. As education experts see it, it will take several years to assess how successfully the combination of standards and "aligned" tests can drive improvements in the classroom.

"We're better off than we were before Common Core," says veteran education scholar Chester Finn, a senior fellow with Stanford's Hoover Institution. "We've got better standards. There's less lying about the performance of kids and schools. There's some better curriculum in place. If you were hoping for a 100% gain, today we're probably looking at a 37% gain. But honestly it's still early days. The aircraft carrier of an education system turns really slowly."

GE's Retreat

General Electric was a leading supporter of Common Core—until it began facing political pressure over the initiative.

Initially, no company supported the Common Core standards more enthusiastically than General Electric. In February 2012 the company's charitable foundation announced an $18 million grant—"the largest corporate commitment to date for the Common Core," its press release noted—to help states transition to the new standards. That August, GE gave another $7 million to Achieve Inc. to aid Common Core's implementation.

GE, which has been active in education philanthropy for decades, didn't just write checks. Led by foundation president Bob Corcoran, a 34-year GE veteran, it evangelized for Common Core. For three years the foundation convened an annual Business and Education Summit focused heavily on Common Core. The 2012 event was dedicated to developing a "unified business effort" backing the standards.

The foundation simultaneously held a separate weeklong conference for educators at the same location, dedicating most of the agenda to Common Core. GE Foundation education director Kelli Wells opened the 2011 event, according to a video posted by an attendee, by declaring, "We're going to be focusing on the Common Core so much that you're going to be eating and drinking and dreaming about the Common Core."

The GE Foundation website asserted that "the future health of business depends on this historic initiative." It urged executives to promote Common Core with state officials, give speeches, "engage the media," and "keep pressure on school boards." GE recruited 73 executives—including the CEOs of Alcoa, Boeing BA 1.58% , and State Farm—to sign an open letter backing Common Core, which was published as a full-page ad in the New York Times in February 2013.

GE also urged companies to remain resolute in the face of the political storm that the reforms were sure to generate, noting, "The business community can be the spine of stability in a changing environment, helping others stay the course too."

As it turned out, GE didn't maintain its own "spine of stability." In 2013 the Tea Party, which was already accusing GE of "crony capitalism" for backing the Export-Import Bank, began attacking the company on Common Core. That April, protesters from FreedomWorks, a Tea Party group, picketed GE's annual shareholders meeting at the New Orleans Convention Center. They carried signs reading GE LEAVE EDUCATION ALONE! and GENERAL ELECTRIC STAY OUT OF MY CHILD'S EDUCATION.

In July 2013 the GE Foundation devoted a conference to Common Core for the last time. That October, Corcoran retired as head of the foundation, replaced by GE's chief diversity officer, Deborah Elam. With that, GE would stop making new grants and advocating Common Core. Says the Business Roundtable's Dane Linn: "They stopped funding anything Common Core related."

Wells, who remains director of GE's U.S. education philanthropy, insists the foundation "never took a stance on Common Core, in the sense of the yes-or-no piece of it." She says GE merely gave money to help educators required to teach to the standards. "It wasn't something where the foundation was saying it's right or wrong," says Wells. "Both sides had valid points and positions." While denying any "retreat," Wells acknowledged that the growing controversy made GE uncomfortable. "That was not something we wanted to be involved with."

Corcoran says he left GE of his own accord. Unaware of the capitulation until Fortune informed him of it, he says the GE Foundation was "absolutely totally committed" to backing Common Core during his tenure, calling it "some of the best work the foundation has ever done."

Adds Corcoran: "If GE has moved away from that investment, it saddens me. If Common Core dies because it's been abandoned too early—moving on to new investments while others tear it apart—you won't see an effort to increase the quality of education systemically in this country for 20 more years."

A version of this article appears in the January 1, 2016 issue of Fortune.

FEBRUARY 2016

Evaluating the Content and Quality of Next Generation Assessments

NANCY DOOREY AND MORGAN POLIKOFF

Foreword by Amber M. Northern and Michael J. Petrilli

THOMAS B. FORDHAM INSTITUTE

ADVANCING EDUCATIONAL EXCELLENCE

The Thomas B. Fordham Institute is the nation's leader in advancing educational excellence for every child through quality research, analysis, and commentary, as well as on-the-ground action and advocacy in Ohio. It is affiliated with the Thomas B. Fordham Foundation, and this publication is a joint project of the Foundation and the Institute. For further information, please visit our website at www.edexcellence.net or write to the Institute at 1016 16th St. NW, 8th Floor, Washington, D.C. 20036. The Institute is neither connected with nor sponsored by Fordham University.

Contents

Foreword 4
Executive Summary 11

Introduction 25

Overview 25
The Four Assessments 26
Study Design and Key Questions 26
Organization of the Report 27
Approach to Alignment and Quality 27
Test Program Differences 28

Section I: Assessments of English Language Arts/Literacy and Mathematics 31

Overview of the Methodology 31
- Study Phases 31
- Selection of Test Forms 33
- Selection of Review Panels and Assignment to Forms 34
- The Study Criteria 36
- Methodology Modifications 38

Findings 40
- ELA/Literacy Content Criteria 41
- ELA/Literacy Depth Criteria 47
- Mathematics Content Criteria 54
- Mathematics Depth Criteria 57
- Criterion Level Ratings 63
- Summary Ratings 64
- Program Strengths and Areas for Improvement 66

Section II: Recommendations 70

For State Policymakers 70
For Test Developers 72

Section III: Suggestions for Methodological Improvement 74

English Language Arts/Literacy 74
Mathematics 77
Across Subjects 79
Summary 79

Appendices 80

Appendix A: Depth of Knowledge (DOK) of the Four Assessment Programs as Compared to the CCSS and Other Policy-Relevant Assessments 80
Appendix B: Key Terminology 82
Appendix C: The Methodology as Written 85
Appendix D: Author Biographies 93
Appendix E: Review Panelist Biographies 94
Appendix F: Full ELA/Literacy and Math Ratings and Summary Statements (Grades 5 and 8) 100
Appendix G: Testing Program Responses to Study and Descriptions of Test Changes for 2015–2016 113

Foreword

By Amber M. Northern and Michael J. Petrilli

We at the Thomas B. Fordham Institute have been evaluating the quality of state academic standards for nearly twenty years. **Our very first study**, published in the summer of 1997, was an appraisal of state English standards by Sandra Stotsky. Over the last two decades, we've regularly reviewed and reported on the quality of state K–12 standards for **mathematics**, **science**, **U.S. history**, **world history**, **English language arts**, and **geography**, as well as the **Common Core**, **International Baccalaureate**, **Advanced Placement** and other influential standards and frameworks (such as those used by **PISA, TIMSS, and NAEP**). In fact, evaluating academic standards is probably what we're best known for.

For most of those two decades, we've also dreamed of evaluating the tests linked to those standards—mindful, of course, that in most places the tests are the real standards. They're what schools (and sometimes teachers and students) are held accountable to and they tend to drive actual curricula and instruction. (That's probably the reason we and other analysts have never been able to demonstrate a close relationship between the quality of standards per se and changes in student achievement.) We wanted to know how well aligned the assessments were to the standards, whether they were of high quality, and what type of cognitive demands they placed on students.

But with fifty-one different sets of tests, such an evaluation was out of reach—particularly since any bona fide evaluation of assessments must get under the hood (and behind the curtain) to look at a sizable chunk of actual test items. Getting dozens of states—and their test vendors—to allow us to take a peek was nigh impossible.

So when the opportunity came along to conduct a groundbreaking evaluation of Common Core-aligned tests, we were captivated. We were daunted too, both by the enormity of the task and by the knowledge that our unabashed advocacy of the standards would likely cause any number of doubters and critics to sneer at such an evaluation coming from us, regardless of its quality or impartiality.

So let's address that first. It's true that we continue to believe that children in most states are better off with the Common Core standards than without them. If you don't care for the standards (or even the concept of "common" standards), or perhaps you come from a state that never adopted these standards or has since repudiated them, you should probably ignore this study. Our purpose here is not to re-litigate the Common Core debate. Rather, we want to know, for states that are sticking with the common standards, whether the "next generation assessments" that have been developed to accompany the standards deliver what they promised by way of strong content, quality, and rigor.

It is also true that the study was funded by a number of foundations that care about assessment quality and the Common Core (seven, in fact, including the Fordham Institute's own foundation). If you think that big private foundations are ruining public education, this study is also not for you.

Now is an especially opportune time to look closely at assessments, since the national testing landscape is in a state of flux. According to the Education Commission of the States, as of October 2015, six states and the District of Columbia planned to administer the Partnership for Assessment of Readiness for College and Careers (PARCC)

test in 2015–16 and fifteen states will deploy the Smarter Balanced Assessment Consortium (Smarter Balanced) test.[1] At least twenty-five others will administer state-specific assessments in math and English language arts. Some (Florida, Ohio, and Utah) will use tests developed by the American Institutes for Research (AIR); others (Indiana, Kentucky, and Virginia) are using Pearson-developed products; still others are choosing "blended" versions of consortium and state-developed items (Michigan and Massachusetts). A handful are undecided and currently in the midst of evaluating test vendors through their RFP process (Maine, Louisiana, and South Carolina). About half the states also require an additional assessment for college admissions, such as the ACT or SAT, which is generally administered in grade 11 (and sometimes statewide). And let's not forget that the new SAT will be unveiled in **March 2016**.

Hence there's no way any single study could come close to evaluating all of the products in use and under development in today's busy and fluid testing marketplace. But what we were able to do was to provide an in-depth appraisal of the content and quality of three "next generation" assessments—ACT Aspire, PARCC, and Smarter Balanced—and one best-in-class state test, the Massachusetts Comprehensive Assessment System (MCAS, 2014). In total, over thirteen million children (about 40 percent of the country's students in grades 3–11) took one of these four tests in spring 2015. Of course it would be good to encompass even more. Nevertheless, our study ranks as a major accomplishment—as well as possibly the most complex and ambitious single project ever undertaken by Fordham.

After we agreed to myriad terms and conditions, we and our team of nearly forty reviewers (more about them below) were granted secure access to operational items and test forms for grades 5 and 8 (the elementary and middle school capstone grades that are this study's focus).[2]

This was an achievement in its own right. It's no small thing to receive access to examine operational test forms. This is especially true in a divisive political climate where anti-testing advocates are looking for *any* reason to throw the baby out with the bathwater and where market pressure gives test developers ample reason to be wary of leaks, spies, and competitors. Each of the four testing programs is to be commended for allowing this external scrutiny of their "live" tests—tests that cost them much by way of blood, sweat, tears, and cash to develop and bring to market. They could have easily said "thanks, but no thanks." But they didn't.

Part of the reason they said yes was the care we took in recruiting smart, respected individuals to help with this project. Our two lead investigators, Nancy Doorey and Morgan Polikoff, together bring a wealth of experience in educational assessment and policy, test alignment, academic standards, and accountability. Nancy has authored reports for several national organizations on advances in educational assessment and she co-piloted the Center for K–12 Assessment and Performance Management at ETS. Morgan is assistant professor of education at the University of Southern California and a well-regarded analyst of the implementation of college and career readiness standards and the influence of curricular materials and tests on that implementation. He is an associate editor of the *American Educational Research Journal*, serves on the editorial board for *Educational Administration Quarterly*, and is the top finisher in the RHSU 2015 Edu-Scholar rankings for junior faculty.[3]

Nancy and Morgan were joined by two well-respected content experts who facilitated and reviewed the work of the ELA/Literacy and math review panels. Charles Perfetti, Distinguished University Professor of Psychology at University of Pittsburgh, served as the ELA/Literacy content lead, and Roger Howe, Professor of Mathematics at Yale, served as the math content lead.

1. J. Woods, "State Summative Assessments: 2015–2016 School Year" (Denver, CO: Education Commission of the States, 2015), http://www.ecs.org/ec-content/uploads/12141.pdf. According to ECS, fifteen states are members of the Smarter Balanced Assessment Consortia, and all but one plan to administer the full assessment in grades 3–8 math and English language arts.

2. The study targets "summative," not "formative," assessments, though most of these same test developers also make the latter available.

3. R. Hess, "2016 RHSU Edu-Scholar Public Influence: Top Tens," *Education Week* (blog), January 7, 2016, http://blogs.edweek.org/edweek/rick_hess_straight_up/2016/01/2016_rhsu_edu-scholar_public_influence_top_tens.html.

Given the importance and sensitivity of the task at hand, we spent months recruiting and vetting the individuals who would eventually comprise the panels led by Dr. Perfetti and Dr. Howe. We began by soliciting recommendations from each participating testing program and other sources, including content and assessment experts, individuals with experience in prior alignment studies, and several national and state organizations. Finalists were asked to submit CVs and detailed responses to a questionnaire regarding their familiarity with the Common Core, their prior experience in conducting alignment evaluations, and any potential conflicts of interest. Individuals currently or previously employed by participating testing organizations and writers of the Common Core were not considered. Given that most card-carrying experts in content and assessment have earned their experience by working on prior alignment or assessment-development studies, and that it's nearly impossible to find experts with zero conflicts, we prioritized balance and fairness. In the end, we recruited at least one reviewer recommended by each testing program to serve on each panel; this strategy helped to ensure fairness by equally balancing reviewer familiarity with the various assessments. (Their bios can be found in Appendix E.)

Which brings us to the matter at hand: How did our meticulously assembled panels go about evaluating the tests—and what did they find? You can read plenty on both questions in the Executive Summary and report itself, which includes ample detail about the study design, testing programs, criteria, and selection of test forms, and review procedures, among other topics.

But the short version is this: we deployed a brand new methodology developed by the Center for Assessment to evaluate the four tests—a methodology that was itself based on the Council of Chief State School Officers' 2014 **"Criteria for Procuring and Evaluating High-Quality Assessments."** Those criteria, say their authors, are "intended to be a useful resource for any state procuring and/or evaluating assessments aligned to their college and career readiness standards." This includes, of course, tests meant to accompany the Common Core standards.

About Those Criteria...

The CCSSO Criteria address the "content" and "depth" of state tests in both English language arts and mathematics. For ELA, "content" spans topics such as whether students are required to use evidence from texts; for math, they are concerned with whether the assessments focus strongly on the content most needed for success in later mathematics. The "depth" criteria for both subjects include whether the tests required a range of "cognitively demanding," high-quality items that make use of various item types (e.g., multiple choice, constructed response, etc.), among other things.

The Center for Assessment took these criteria and transformed them into a number of measurable elements that reviewers addressed. In the end, the newly minted methodology wasn't perfect. Our rock-star reviewers improved upon it and wanted others following in their footsteps to benefit from their learned experience. So we made adjustments along the way (see Section I, *Methodology Modifications* for more).

The panels essentially evaluated the extent of the match between the assessment and a key element of the CCSSO document. They assigned one of four ratings to each ELA and math-specific criterion, such that tests received one of four "match" ratings: Excellent, Good, Limited/Uneven, or Weak Match. To generate these marks, each panel reviewed the ratings from the grade 5 and grade 8 test forms, considered the results from the analysis of the program's documentation (which preceded the item review), and came to consensus on the rating.

What did they ultimately find? The summary findings appear below.

TABLE F-1

Overall Content and Depth Ratings for ELA/Literacy and Mathematics

	ACT Aspire	MCAS	PARCC	Smarter Balanced
ELA/Literacy CONTENT	L	L	E	E
ELA/Literacy DEPTH	G	G	E	G
Mathematics CONTENT	L	L	G	G
Mathematics DEPTH	G	E	G	G

LEGEND E Excellent Match G Good Match L Limited/Uneven Match W Weak Match

As shown, the PARCC and Smarter Balanced assessments earned an Excellent or Good Match to the subject-area CCSSO Criteria for both ELA/Literacy and mathematics. This was the case with both Content and Depth.

ACT Aspire and MCAS (along with the others) also did well regarding the quality of their items and the depth of knowledge assessed (both of which are part of the Depth rating). But the panelists also found that they did not adequately assess—or in some cases did not really assess at all—some of the priority content in both ELA/Literacy and mathematics at one or both grade levels in the study (Content).

What do we make of these bottom-line results? Simply put, developing a test—like all major decisions and projects in life—is full of trade-offs. PARCC and Smarter Balanced are a better match to the CCSSO criteria, which is not surprising, given that they were both developed with the Common Core in mind. ACT Aspire, on the other hand, was not developed for that explicit purpose. In a paper on their website, ACT officials Sara Clough and Scott Montgomery explain that ACT Aspire was

> *under development prior to the release of the Common Core State Standards [and] not designed to directly measure progress toward those standards. However, since ACT data, empirical research, and subject matter expertise about what constitutes college and career readiness was lent to the Common Core development effort, significant overlap exists between the Common Core State Standards and the college and career readiness constructs that ACT Aspire and the ACT measure.*[4]

Our reviewers also found some "overlap" in MCAS given that the state had added new Common Core items to its 2014 test. Yet the Bay State's intention was not a full redesign, particularly since it was then in the midst of deciding between MCAS and PARCC as its test of choice (the state ultimately decided on a hybrid).[5] To the extent that states want their tests to reflect the grade-level content in the new standards, they should choose accordingly.

The CCSSO Criteria do not consider testing time, cost, or comparability. But those are nonetheless key considerations for states as they make assessment decisions. Although PARCC and Smarter Balanced are a better match to the Criteria, they also take longer to administer and are more expensive. The estimated testing time for

4. S. Clough and S. Montgomery, "How ACT Assessments Align with State College and Career Readiness Standards" (Iowa City, IA: ACT, 2015), http://www.discoveractaspire.org/pdf/ACT_Alignment-White-Paper.pdf.

5. J. Fox, "Education Board Votes to Adopt Hybrid MCAS-PARCC Test," *Boston Globe*, November 17, 2015, https://www.bostonglobe.com/metro/2015/11/17/state-education-board-vote-whether-replace-mcas/aex1nGyBYZW2sucEW2082L/story.html. To the extent that states want their tests to reflect the grade-level content in the new standards, they should choose accordingly.

students in grades 5 and 8, on average, to complete both the ELA/Literacy and mathematics assessments for all four programs is as follows:

- ACT Aspire: three to three and one-quarter hours for all four tests (English, reading, writing, and mathematics)
- MCAS 2014: three and a half hours
- PARCC: seven to seven and a half hours[6]
- Smarter Balanced: five and a half hours

The longer testing times for PARCC and Smarter Balanced are due primarily to their inclusion of extended performance tasks. Both programs use these tasks to assess high-priority skills within the CCSS, such as the development of written compositions in which a claim is supported with evidence drawn from sources; research skills; and solving complex multi-step problems in mathematics. In addition to requiring more time than multiple-choice items, these tasks are also typically costlier to develop and score.[7]

Another trade-off pertains to inter-state comparability. Some states want the autonomy and uniqueness that come with having their own state test developed by their own educators. Other states prioritize the ability to compare their students with those in other states via a multi-state test. We think the extra time and money,[8] plus the comparability advantage, are trade-offs worth making, but we can't pretend that they're not tough decisions in a time of tight budgets and widespread anxiety about testing burden.

Of course we're mindful—as anyone in this field would be—of the recent backlash to testing and the so-called "opt-out movement." We understand that some local and state officials are wary of adopting longer tests. We also suspect that most of the concerns that parents have isn't with the length of one test in May, but with the pressure that educators feel to teach to the test and narrow the curriculum.

If we're right and that's the real problem, the answer is stronger tests, which encourage better, broader, richer instruction, and which make traditional "test prep" ineffective. Tests that allow students of all abilities, including both at-risk and high-achieving youngsters, to demonstrate what they know and can do. More rigorous tests that challenge students more than they've been challenged in the past. But, again, those tests tend to take a bit longer (say, five hours rather than two and a half hours) and cost a bit more. Our point is not to advocate for any particular tests but to root for those that have qualities that enhance, rather than constrict, a child's education and give her the best opportunity to show what she's learned.

A discussion of such qualities, and the types of trade-offs involved in obtaining them, are precisely the kinds of conversations that merit honest debate in states and districts.

******

We at Fordham don't plan to stay in the test-evaluation business. The challenge of doing this well is simply too overwhelming for a small think tank like ours. But we sincerely hope that others will pick up the baton, learn from

6. The 2015–16 PARCC revisions will reduce this time by an estimated one and a half hours.

7. That said, Matthew Chingos, in a 2012 study on state assessment spending, found that "collaborating with other states achieves cost savings simply by spreading fixed costs over more students…" (page 22). See M. Chingos, "Strength in Numbers: State Spending on K–12 Assessment Systems" (Washington, D.C.: Brookings Institution, November 29, 2012), http://www.brookings.edu/research/reports/2012/11/29-cost-of-ed-assessment-chingos.

8. Note that the per-pupil costs for PARCC, Smarter Balanced, and ACT Aspire are in the same ballpark, ranging from roughly $22 to $25 depending on the tested subjects. The MCAS, typically viewed as a higher-quality state test, costs $42 per student. The costs associated with many of the prior state tests were considerably lower than these figures so changing tests represented an increase for them. See M. Chingos, "Strength in Numbers." Cost estimates for PARCC and Smarter Balanced can be found here: http://www.parcconline.org/news-and-video/press-releases/248-states-select-contractor-to-help-develop-and-implement-parcc-tests; http://www.smarterbalanced.org/faq/7-what-does-it-cost-for-a-state-to-participate-in-smarter-balanced-assessment-consortium/. Per MCAS, "The approximate cost of the legacy MCAS assessment is $42 per student for ELA and mathematics per estimates presented to the Massachusetts State Board of Elementary and Secondary Education in fall 2015" (personal email communication with Michol Stapel, January 22, 2016). Per ACT Aspire, "The estimated price for 2016 is $25 per student and includes English, Mathematics, Reading, Writing, and Science subject tests" (personal email communication with Elizabeth Sullivan, January 21, 2016).

our experience, and provide independent evaluations of the assessments in use in the states that have moved away from PARCC, Smarter Balanced, and ACT Aspire.

Not only will such reviews provide critical information for state and local policymakers, as well as educators, curriculum developers and others, they might also deter the U.S. Department of Education from pursuing a dubious plan to make states put their new assessments through a federal evaluation system. In October 2015, the Department issued procedures for the "peer review" process that had been on hold for the last three years. The guidelines specify that states must produce evidence that they "used sound procedures in design and development to state tests aligned to academic standards, and for test administration and security." Count us among those who think renewed federal vetting of state tests invites unwanted meddling from Uncle Sam (and could spark another round of backlash akin to what happened to the Common Core itself a few years back.) Besides, twelve years during which the Department already had such guidance in place did little to improve the quality of state tests—hence the recent moves to improve them.

Parting Thoughts

We are living in a time of political upheaval, divisiveness, and vitriol. The public's faith in government and other large institutions is at an all-time low. So we're glad to be the bearers of good news for a change. All four tests we evaluated boasted items of high technical quality. Further, the next generation assessments that were developed with the Common Core in mind have largely delivered on their promises. Yes, they have improvements to make (you'll see that our reviewers weren't shy in spelling those out). But they tend to reflect the content deemed essential in the Common Core standards and demand much from students cognitively. They are, in fact, the kind of tests that many teachers have asked state officials to build for years.

Now they have them.

Acknowledgments

This research was made possible through the generous support of the Louis Calder Foundation, the High-Quality Assessment Project (including the Bill & Melinda Gates Foundation, the Lumina Foundation, the Charles and Lynn Schusterman Family Foundation, the William and Flora Hewlett Foundation, and the Helmsley Trust), and our sister organization, the Thomas B. Fordham Foundation.

We owe a debt of gratitude to Nancy Doorey, project manager and report coauthor, and Morgan Polikoff, alignment expert and report coauthor, for their invaluable contributions to this project. This study had its share of difficulties, and through their tireless efforts Nancy and Morgan proved themselves highly equipped to handle all of them. We also extend our thanks to Dr. Roger Howe and Dr. Charles Perfetti, who served as math and ELA/Literacy content leads for the study and assisted with everything from creating initial reviewer training materials to overseeing the review process and synthesizing final study findings. Thanks also to Melisa Howey and Lynne Olmos for their special assistance and to the rest of our esteemed panelists for the thoughtfulness and care with which they conducted their reviews.

We also extend sincere thanks to each of the testing organizations who participated in the study (ACT Aspire, MCAS, PARCC, and Smarter Balanced) and to the many members of their staff who conducted initial reviewer trainings for their respective programs, responded to the panelists' questions, and reviewed the final report drafts for accuracy. In particular, we thank Elizabeth Sullivan, Carrie Conaway, Judy Hickman, and Francine Markowitz for facilitating this work.

We also appreciate the contributions of our colleagues at the Human Resources Research Organization (HumRRO), who led a similar evaluation at the high school level (reported separately) and with whom we conducted several joint reviews; the National Center for the Improvement of Educational Assessment (NCIEA) for producing the study's methodology; Judy Wurtzel and Joanne Weiss for facilitating the study on behalf of funders; and Student Achievement Partners (SAP) for helping to design and deliver, along with Jami-Jon Pearson, Morgan Polikoff, and HumRRO staff, portions of the reviewer training.

Fordham Research Manager Victoria Sears skillfully helped manage all aspects of the project, led recruitment of panelists, provided input on drafts, and shepherded the project across the finish line. Chester E. Finn, Jr. provided valuable feedback and edits to drafts, Alyssa Schwenk handled funder and media relations, and Shep Ranbom assisted with managing the report's dissemination. We also thank Fordham interns Megan Lail, Damien Schuster, and Stephan Shehy for their assistance throughout the project, and Jonathan Lutton, who ushered the report through production. Finally, we thank Shannon Last, who served as our copy editor; Edward Alton, who designed the report's layout; and Thinkstock.com, from which our cover design originated.

Executive Summary

Approximately one-third of American freshmen at two-year and four-year colleges require remedial coursework and over 40 percent of employers rate new hires with a high school diploma as "deficient" in their overall preparation for entry-level jobs.[9, 10] Yet, over the past decade, as these students marched through America's public education system, officials repeatedly told them, and their parents, that they were on track for success. They passed their courses, got good grades, and aced state annual tests. To put it plainly, it was all a lie. Imagine being told year after year that you're doing just fine—only to find out when you apply for college or a job that you're simply not as prepared as you need to be.

Thankfully, states have taken courageous steps to address this preparedness gap. Over the past five years, every state has upgraded its K–12 academic standards to align with the demands of college and career readiness (CCR), either by adopting the Common Core State Standards (CCSS) or working with their own higher education and career training providers to strengthen or develop standards. New assessments intended to align to these more-rigorous standards made their debut in the past year or two, and, as was widely expected (and, indeed, inevitable), student proficiency rates are lower than on previous tests—often significantly lower. State and local officials must decide whether to forge ahead with the new tests and higher expectations or back down in order to cast more schools and students in a positive (if, perhaps, illusory) light.

Of course, test scores that more accurately predict students' readiness for entry-level coursework or training are not enough. The content of state assessments, too, is an important predictor of the impact of those tests on what is taught and learned. For instance, low-quality assessments poorly aligned with the standards will undermine the content messages of the standards; given the tests' role in accountability under the newly reauthorized Elementary and Secondary Education Act, it is only logical that such tests might contribute to poor-quality instruction.

In short, good tests matter. Of critical importance to this conversation, therefore, is whether the new tests are indeed good and worth fighting for. That's the central question this study seeks to answer.

The Tests

In the pages that follow, we evaluate the quality of four standardized assessments—three new, multi-state assessments and a well-regarded existing state assessment—to determine whether they meet new criteria developed by the Council of Chief State School Officers (CCSSO) for test quality. These new criteria, as explained in the following pages, ask that evaluators take a deep look at whether the assessments target and reliably measure the essential skills and knowledge needed at each grade level to achieve college and career readiness by the end of high school.

9. National Center for Education Statistics (NCES), Digest of Education Statistics, *Percentage of First-Year Undergraduate Students Who Took Remedial Education Courses, by Selected Characteristics: 2003–04 and 2007–08,* Table 241 (Washington, D.C.: NCES, 2010), https://nces.ed.gov/programs/digest/d10/tables/dt10_241.asp.

10. Conference Board et al., "Are They Really Ready To Work? Employers' Perspectives on the Basic Knowledge and Applied Skills of New Entrants to the 21st Century U.S. Workforce" (New York, NY: Conference Board, 2006), http://www.p21.org/storage/documents/FINAL_REPORT_PDF09-29-06.pdf.

We evaluate English language arts/literacy and mathematics assessments for grades 5 and 8 for this quartet of testing programs:

- ACT Aspire
- The Partnership for Assessment of Readiness for College and Careers (PARCC)
- The Smarter Balanced Assessment Consortium (Smarter Balanced)
- The Massachusetts Comprehensive Assessment System (MCAS, 2014)

The Study Design

The analysis that follows was designed to answer three questions:

1. Do the assessments place strong emphasis on the most important content for college and career readiness (CCR), as called for by the Common Core State Standards and other CCR standards? (**Content**)
2. Do they require all students to demonstrate the range of thinking skills, including higher-order skills, called for by those standards? (**Depth**)
3. What are the overall strengths and weaknesses of each assessment relative to the examined criteria for ELA/Literacy and mathematics? (**Overall Strengths and Weaknesses**)

To answer these questions, we use a new methodology based on the CCSSO's 2014 "Criteria for Procuring and Evaluating High-Quality Assessments."[11] Developed by experts at the National Center for the Improvement of Educational Assessment (NCIEA), this methodology evaluates the degree to which test items and supporting program documentation (e.g., test blueprints and documents describing the item creation process) measure the critical competencies reflected in college and career readiness standards, thereby sending clear signals about the instructional priorities for each grade.[12]

The evaluation was conducted by review panels composed of practitioners, content experts, and specialists in assessment. Following reviewer training and a calibration exercise, the panels evaluated test items across various dimensions, with three to four experts reviewing each test form. Results were aggregated for each test form, discussed among the panel members, combined with results from a review of program documentation, and turned into group ratings and summary statements about each program.

The quality and credibility of an evaluation of this type rests largely on the expertise and judgment of the individuals serving on the review panels. To recruit highly qualified yet impartial reviewers, the study team requested recommendations from each of the four testing programs; from other respected content, assessment, and alignment experts; and from several national and state organizations. Reviewers were carefully vetted for their familiarity with the CCSS, their experience with developing or evaluating assessment items, and potential conflicts of interest. Individuals currently or previously employed by participating testing organizations and writers of the CCSS were not considered. (For more information, see Section I, *Selection of Review Panels*.) To ensure fairness and a balance of reviewer familiarity with each assessment, each of the panels included at least one reviewer recommended by each testing program.

Two university-affiliated content leads facilitated and reviewed the work of the ELA/Literacy and math review panels. Dr. Charles Perfetti, Distinguished University Professor of Psychology at University of Pittsburgh, served as the ELA/Literacy content lead, and Dr. Roger Howe, Professor of Mathematics at Yale University, served as the mathematics content lead. The names and biographical summaries of all panelists appear in Appendix E.

11. Council of Chief State School Officers (CCSSO), "Criteria for Procuring and Evaluating High-Quality Assessments" (Washington, D.C.: CCSSO, 2014).

12. The National Center for the Improvement of Educational Assessment, Inc. (NCIEA), "Guide to Evaluating Assessments Using the CCSSO Criteria for High Quality Assessments: Focus on Test Content" (Dover, NH: NCIEA, February 2016): http://www.nciea.org/publication_PDFs/Guide%20to%20Evaluating%20CCSSO%20Criteria%20Test%20Content%20020316.pdf.

This study evaluates English language arts and math assessments at grades 5 and 8, while a parallel study led by the Human Resources Research organization (HumRRO) evaluates the high school assessments from the same four testing programs (see Table ES-1). Because both organizations used the same methodology, it made sense to conduct two portions of the review jointly and across all grades: the documentation review and the accessibility review. Documentation results specific to grades 5 and 8 are addressed in this report. Please see HumRRO's report for the results from their evaluation of the high school assessments, as well as results from the joint accessibility review (all grades).[13]

TABLE ES-1

Overview of the Parallel Fordham and HumRRO Studies

	ELA/Literacy Review	Math Review	Documentation Review	Accessibility Review
Fordham Study	Grades 5 and 8	Grades 5 and 8	Joint Panel	Joint Panel
HumRRO Study	High School	High School	(grades 5 and 8 findings presented in this report; high school findings presented in HumRRO report)	(presented in HumRRO report)

The CCSSO Criteria for High-Quality Assessments

To evaluate assessments intended to measure student mastery of the Common Core State Standards, we needed a new methodology that would capture their key dimensions. Traditional alignment methodologies offer the advantage of having been studied extensively, but treat each of the grade-level standards with equal importance, creating an inadvertent incentive for tests—and instruction—to be "a mile wide and an inch deep."

The CCSSO's "Criteria for Procuring and Evaluating High-Quality Assessments" was the basis of the new methodology. Specifically designed to address tests of college and career readiness, these criteria focus the evaluation on the highest priority skills and knowledge at each grade in the CCSS, addressing foundational as well as complex skills. By using the CCSSO Criteria as the basis of the methodology, the evaluation rewards those tests that focus on the essential skills and give clear signals about the instructional priorities for each grade.

The CCSSO Criteria address six domains, but only two pertain to the research questions addressed in this study: those for the assessment of ELA/Literacy standards and the assessment of mathematics standards (see Table ES-2).

In addition, CCSSO defined ratings for test content and depth, each of which is based on a subset of ratings. The Content rating reflects the degree to which each test assesses the material most needed for college and career readiness, and the Depth rating reflects the degree to which each test assesses the depth and complexity of the college and career readiness standards.

13. This study also originally included an evaluation of test program transparency, or the extent to which programs provide sufficient information to the public regarding assessment design and expectations (CCSSO criterion A.6). Due to several challenges associated with this review, however, we ultimately decided to drop this criterion from our study. Review panelists were not able to review all relevant documentation for each program, due to the vast volume of materials provided and publicly available. In addition, many test programs continued to release additional information (such as sample items) since our review occurred, rendering this panel's findings somewhat outdated.

TABLE ES-2

CCSSO Criteria Evaluated in This Study

Assessment of ELA/Literacy Standards

- Test Content Criteria
  - B.3 Requiring students to read closely and use evidence from texts
  - B.5 Assessing writing from sources
  - B.6 Emphasizing vocabulary and language skills
  - B.7 Assessing research and inquiry
  - B.8 Assessing speaking and listening
- Test Depth Criteria
  - B.1 Using a balance of high-quality literary and informational texts
  - B.2 Focusing on the increasing complexity of texts across grades
  - B.4 Requiring a range of cognitive demand
  - B.9 Ensuring high-quality items and a variety of item types

Assessment of Mathematics Standards

- Test Content Criteria
  - C.1 Focusing strongly on the content most needed for success in later mathematics (i.e., the major work of the grade)
  - C.2 Assessing a balance of concepts, procedures, and applications
- Test Depth Criteria
  - C.3 Connecting mathematics practices to mathematical content
  - C.4 Requiring a range of cognitive demand
  - C.5 Ensuring high-quality items and a variety of item types

Findings

Results are organized around the key research questions above.

✓ RESULTS FOR QUESTIONS #1 AND #2

Do the assessments place strong emphasis on the most important content for college and career readiness (CCR) as called for by the Common Core State Standards and other CCR standards? **(Content)**

Do they require all students to demonstrate the range of thinking skills, including higher-order skills, called for by those standards? **(Depth)**

The panels assigned one of four ratings to each ELA/Literacy and math criterion: Excellent Match, Good Match, Limited/Uneven Match, or Weak Match. To generate these, each panel reviewed the ratings from the grade 5 and grade 8 test forms, considered the results of the documentation review, and came to consensus on the criterion rating.

Table ES-3 shows the ratings for test content and depth in ELA/Literacy and mathematics across the four programs.

The PARCC and Smarter Balanced assessments earned an Excellent or Good Match to the CCSSO Criteria for both ELA/Literacy and mathematics. While ACT Aspire and MCAS did well regarding the quality of items (see Section I, *Results*) and the Depth of Knowledge assessed (Depth), the panelists found that these two programs do not adequately assess—or may not assess at all—some of the priority content in both ELA/Literacy and mathematics at one or both grades in the study (Content).

TABLE ES-3

Overall Content and Depth Ratings for ELA/Literacy and Mathematics

	ACT Aspire	MCAS	PARCC	Smarter Balanced
ELA/Literacy CONTENT	L	L	E	E
ELA/Literacy DEPTH	G	G	E	G
Mathematics CONTENT	L	L	G	G
Mathematics DEPTH	G	E	G	G

LEGEND E Excellent Match G Good Match L Limited/Uneven Match W Weak Match

Criterion Level Results for ELA/Literacy and Mathematics

The Content and Depth ratings are based on the results of subsets of the CCSSO Criteria, as described above. NCIEA also recommended that certain criteria be "emphasized," meaning awarded greater weight in the final determinations (though precise weightings were not specified). The panels, however, sometimes chose not to adhere to the weighting based on their level of confidence in reviewing each criterion (see Section I, *Methodology Modifications*).

Tables ES-4A and 4B show the distribution of the ELA/Literacy and math criteria ratings. Immediately striking in ELA is that the two consortia assessments (PARCC and Smarter Balanced, which received development grants from the U.S. Department of Education) earned twice as many ratings of Good and Excellent Match as the other two programs, earning eight high ratings to the four of ACT Aspire and MCAS. PARCC earned the most Excellent Match ratings (six), while Smarter Balanced was the only assessment with no ratings of Weak Match (partly because it was also the only program to test listening on the summative assessment).

TABLE ES-4A

ELA/Literacy Ratings Tally by Program

ACT Aspire	E	G	G	G	L	L	L	W	W
MCAS	E	G	G	G	L	L	W	W	W
PARCC	E	E	E	E	E	E	G	G	W
Smarter Balanced	E	E	E	E	G	G	G	G	L

TABLE ES-4B

Mathematics Ratings Tally by Program[14]

ACT Aspire	E	E	L	W
MCAS	E	E	E	L
PARCC	E	G	G	G
Smarter Balanced	E	G	G	L

LEGEND E Excellent Match G Good Match L Limited/Uneven Match W Weak Match

14. Although all four programs require the assessment of conceptual understanding, procedural skill/fluency, and applications (criterion C.2), final ratings could not be determined with confidence due to variations in how reviewers understood and implemented this criterion.

The ratings for mathematics (Table ES-4B) were more similar between programs, with PARCC earning four Excellent or Good Match ratings, Smarter Balanced and MCAS three each, and ACT Aspire two. MCAS scored particularly well on the three Depth criteria in mathematics, while PARCC is the only assessment that earned all Good Match or better scores.

Tables ES-5A and ES-5B on the following pages provide the final criterion ratings for each program, organized by Content and Depth. They also provide the specific attributes required to fully meet each criterion as indicated in the methodology.[15] Those criteria followed by an asterisk were awarded greater emphasis during development of the Content and Depth ratings.

TABLE ES-5A

Criterion Ratings for ELA/Literacy

CONTENT		ACT Aspire	MCAS	PARCC	Smarter Balanced
B.3*	**Reading:** Items require close reading and use of direct textual evidence, and focus on central ideas and important particulars. **To Meet the Criterion:** 1) Nearly all reading items require close reading and analysis of text, rather than skimming, recall, or simple recognition of paraphrased text. 2) More than half of the reading score points are based on items that require direct use of textual evidence. 3) Nearly all items are aligned to the specifics of the standards. 4) More than half of the reading score points are based on items that require direct use of textual evidence.	L	G	E	E
B.5*	**Writing:** Test programs assess a variety of types and formats of writing and the use of writing prompts that require students to confront and use evidence from texts or other stimuli directly. **To Meet the Criterion**: 1) All three writing types (expository, narrative, and persuasive/argument) are approximately equally represented across all forms in the grade band (K–5 and 6–12), allowing blended types (those that combine types) to contribute to the distribution. 2) All writing prompts require writing to sources (are text-based).	L	W	E	E
B.6	**Vocabulary and Language Skills:** Test forms place adequate emphasis on language and vocabulary items on the assessment, assess vocabulary that reflect requirements for college and career readiness, and focus on common student errors in language questions. **To Meet the Criterion:** 1) The large majority of vocabulary items (i.e., three-quarters or more) focuses on Tier 2 words and requires use of context, and more than half assess words important to central ideas. 2) A large majority (i.e., three-quarters or more) of the items in the language skills component and/or scored with a writing rubric mirror real-world activities, focus on common errors, and emphasize the conventions most important for readiness. 3) Vocabulary is reported as a sub-score or at least 13 percent of score points are devoted to assessing vocabulary/language. 4) Same as #3 for language skills.	G	L	E	G

* Criterion awarded greater weight in determination of Content and Depth rating.

LEGEND E Excellent Match G Good Match L Limited/Uneven Match W Weak Match

15. Note: As first implementers of the methodology, the reviewers made a number of modifications they deemed important for improvement. See Section I, *Methodology Modifications*.

CONTENT		ACT Aspire	MCAS	PARCC	Smarter Balanced
B.7	**Research and Inquiry:** Test forms include research items/tasks requiring students to analyze, synthesize, organize, and use information from multiple sources. **To Meet the Criterion:** The large majority (i.e., three-quarters or more) of the research items require analysis, synthesis, and/or organization of information.	L	W	E	E
B.8	**Speaking and Listening:** (*Not yet required by the criteria, so not included in the Content rating. Listening requirements are listed here because one program assesses listening.)* Items assess students' listening skills using passages with adequate complexity and assess students' speaking skills through oral performance tasks. **To Meet the Criterion:** 1) The large majority (i.e., at least three-quarters) of listening items meet the requirements outlined in B.1 and B.2 and evaluate active listening skills.	W	W	W	L
DEPTH		**ACT Aspire**	**MCAS**	**PARCC**	**Smarter Balanced**
B.1*	**Text Quality and Types:** Test forms include a variety of text types (narrative and informational) that are of high quality, with an increasing focus on diverse informational texts across grades. **To Meet the Criterion:** 1) Approximately half of the texts at grades 3–8 and two-thirds at high school are informational, and the remainder literary. 2) Nearly all passages are high quality (previously published or of publishable quality). 3) Nearly all informational passages are expository in structure. 4) For grades 6–12, the informational texts are split nearly evenly between literary nonfiction, history/social science, and science/technical texts.	G	G	G	E
B.2	**Complexity of Texts:** *(based on documentation review only)* Assessments include texts that have appropriate levels of text complexity for the grade or grade band (grade bands identified in the CCSS are K–5 and 6–12). **To Meet the Criterion:** 1) The documentation clearly explains how quantitative data are used to determine grade band placement. 2) Texts are then placed at the grade level recommended by qualitative review. 3) Text complexity rating process results in nearly all passages being placed at a grade band and grade level justified by complexity data.	G	G	G	G
B.4	**Matching the Complexity of the Standards:** Each test form contains an appropriate range of cognitive demand that adequately represents the cognitive demand of the standards. **To Meet the Criterion:** 1) The distribution of cognitive demand of the assessment matches the distribution of cognitive demand of the standards as a whole and matches the higher cognitive demand (DOK 3+) of the standards. *(Note: This is not a rating of test difficulty. Assessments that do not match the DOK distribution of the standards, even if there are too many high DOK items, may receive a rating less than Excellent Match. See Appendix A for more information.)*	W	L	E	G
B.9	**High-Quality Items and Variety of Item Types:** Test items are of high quality, lacking technical or editorial flaws and each test form contains multiple item types including at least one type in which students construct, rather than select, a response. **To Meet the Criterion:** 1) All or nearly all operational items reviewed reflect technical quality and editorial accuracy. 2) At least two item formats are used, including one that requires students to generate, rather than select, a response.	E	E	E	G

* Criterion awarded greater weight in determination of Content or Depth rating.

LEGEND E Excellent Match G Good Match L Limited/Uneven Match W Weak Match

Cells for which the ratings are not used in determining Content and Depth ratings (See Section I, *Weighting of Criteria for Content and Depth Ratings*.)

TABLE ES-5B

Criterion Ratings for Mathematics

CONTENT		ACT Aspire	MCAS	PARCC	Smarter Balanced
C.1*	**Focus:** Each test form contains a strong focus on the content most crucial for success in later mathematics (i.e., the major work of the grade). **To Meet the Criterion:** The vast majority (i.e., at least three-quarters in elementary grades, at least two-thirds in middle school grades, and at least half in high school) of score points in each assessment focus on the content that is most important for students to master in that grade band in order to reach college and career readiness (the major work of the grade).	W	L	G	G
C.2	**Concepts, Procedures, and Applications:** Each test form contains items that assess conceptual understanding, procedural skill/fluency, and application in approximately equal proportions. **To Meet the Criterion:** The distribution of score points reflects a balance of mathematical concepts, procedures/fluency, and applications.	*Due to variations in how reviewers understood and implemented this criterion, final ratings could not be determined with confidence.*			
DEPTH		**ACT Aspire**	**MCAS**	**PARCC**	**Smarter Balanced**
C.3	**Connecting Practice to Content:** Assessments test students' use of mathematical practices through test items that connect these practices with grade-level content standards. **To Meet the Criterion:** All or nearly all items that assess mathematical practices also align to one or more content standards.	E	E	E	E
C.4*	**Matching the Complexity of the Standards:** Each test form contains an appropriate range of cognitive demand that adequately represents the cognitive demand of the standards. **To Meet the Criterion:** 1) The distribution of cognitive demand of the assessment matches the distribution of cognitive demand of the standards as a whole and matches the higher cognitive demand (DOK 3+) of the standards. *(Note: This is not a rating of test difficulty. Assessments that do not match the DOK distribution of the standards, even if there are too many high DOK items, may receive a rating less than Excellent Match. See Appendix A for more information.)*	L	E	G	G
C.5*	**High-Quality Items and Variety of Item Types:** Test items are of high quality, lacking technical or editorial flaws, and each test form contains multiple item types, including at least one type in which students construct, rather than select, a response. **To Meet the Criterion:** 1) All or nearly all operational items reviewed reflect technical quality and editorial accuracy. 2) At least two item formats are used, including one that requires students to generate, rather than select, a response.	E	E	G	L

* Criterion awarded greater weight in determination of Content or Depth rating.

LEGEND E Excellent Match G Good Match L Limited/Uneven Match W Weak Match

In the ELA/Literacy assessments, all four programs receive high ratings for the quality of items and variety of item types. In addition, all pay close attention to the use of high-quality informational and literary texts and increasing the complexity of tests across grades, which are significant advances over many previous state ELA assessments. Significant differences exist across the testing programs, however, in the degree to which their writing tasks require students to use evidence from sources and the extent to which research skills are assessed. In these areas, PARCC and Smarter Balanced perform well, receiving higher ratings than either ACT Aspire, which receives a rating of Limited/Uneven Match on these criteria, or MCAS, which receives a rating of Weak Match. PARCC and Smarter Balanced assessments also contain a distribution of cognitive demand that better reflects that of the standards, when compared to ACT Aspire and MCAS.

In mathematics, PARCC and Smarter Balanced receive a rating of Good Match for the degree to which their tests focus on the most important content of the grade. ACT Aspire test forms receive a rating of Weak Match on this prioritized criterion, due to their test design choice, in which off-grade standards are assessed in order to monitor mastery across grades. MCAS receives a rating of Limited/Uneven because its grade 5 forms do not contain

Supplemental Analysis: Assessment of Higher-Order Thinking Skills

CCSSO criteria B.4 and C.4 capture the degree to which the range of cognitive demand on the test forms match that of the CCSS. We used Webb's Depth of Knowledge (DOK) taxonomy to assess cognitive demand, as it is by far the most widely used approach to categorizing cognitive demand Webb's DOK is composed of four levels. Level 1 is the lowest level (recall), Level 2 requires use of a skill or concept, and Levels 3 and 4 are higher-order thinking skills. We compared the DOK of the assessments to those of the Common Core State Standards, which were coded by content experts (see Section I, *Selection of Review Panels and Assignment to Forms*). We also compared the tests' DOK distributions to those of fourteen highly regarded previous state assessments, as well as the distribution reflected in several national and international assessments—including Advanced Placement (AP), the National Assessment of Education Progress (NAEP), and the Program for International Student Assessment (PISA).[16, 17]

We found that the CCSS call for greater emphasis on higher-order skills than fourteen highly regarded previous state assessments in ELA/Literacy at both grades 5 and 8 as well as in math at grade 8 (they are similar at grade 5). In addition, the grade 8 CCSS in both ELA/Literacy and math call for greater emphasis on higher-order thinking skills than either NAEP or PISA, both of which are considered to be high-quality, challenging assessments.

Overwhelmingly, the assessments included in our study were found to be more challenging—placing greater emphasis on higher-order skills—than prior state assessments, especially in mathematics (where prior assessments rarely included items at DOK 3 or 4 at all). In some cases, the increase was dramatic: PARCC's DOK in grade 8 exceeds even that of AP and PISA in both subjects. See Appendix A for more.

However, the panels found significant variability in the degree to which the four assessments match the distribution of DOK in the CCSS. In some cases, the panels found significant variability between the grade 5 and grade 8 assessments for a given program. PARCC tests generally have the highest DOK in ELA/Literacy, while ACT Aspire had the highest in mathematics. See Section I, Tables 14 and 22 for the DOK distribution of each program.

16. L. Yuan and V. Le, *Estimating the Percentage of Students who were Tested on Cognitively Demanding Items through the State Achievement Tests* (Santa Monica, CA: RAND Corporation, 2012).

17. *Ibid*.

sufficient focus on the critical content for the grade. With respect to item quality, ACT Aspire and MCAS receive the highest rating of Excellent Match, whereas PARCC receives a rating of Good Match and Smarter Balanced a rating of Limited/Uneven Match.[18]

✓ RESULTS FOR QUESTION #3

What are the overall strengths and weaknesses of each assessment relative to the examined criteria for ELA/Literacy and mathematics? **(Overall Strengths and Weaknesses)**

Each of the review panels developed summary statements for each assessment program, detailing their strengths and areas of improvement in ELA/Literacy and mathematics. In addition, they created summary statements for each test's Content and Depth ratings based on the prioritization of criteria recommended in the study methodology (see Appendix F). They also generated final statements summarizing the observed strengths and areas of improvement for each program.

ACT Aspire

English Language Arts:

In ELA/Literacy, ACT Aspire receives a Limited/Uneven to Good Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. The combined set of ELA/Literacy tests (reading, writing, and English) requires close reading and adequately evaluates language skills. More emphasis on assessment of writing to sources, vocabulary, and research and inquiry, as well as increasing the cognitive demands of test items, will move the assessment closer to fully meeting the criteria. Over time, the program would also benefit by developing the capacity to assess speaking and listening skills.

Content: ACT Aspire receives a Limited/Uneven match to the CCSSO Criteria for Content in ELA/Literacy. The assessment program includes an emphasis on close reading and language skills. However, the reading items fall short on requiring students to cite specific textual information in support of a conclusion, generalization, or inference and in requiring analysis of what has been read. In order to meet the criteria, assessing writing to sources, vocabulary, as well as research and inquiry need to be strengthened.

Depth: ACT Aspire receives a rating of Good Match for Depth in ELA/Literacy. The program's assessments are built on high-quality test items and texts that are suitably complex. To fully meet the CCSSO Criteria, more cognitively demanding test items are needed at both grade levels, as is additional literary narrative text—as opposed to literary informational texts.[19]

Mathematics:

In mathematics, ACT Aspire receives a Limited/Uneven to Good Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. Some of the mismatch with the criteria is likely due to intentional program design, which requires that items be included from previous and later grades.

18. The nature and timing of this review required Smarter Balanced to make the test items and forms available to reviewers through an alternate test interface that was more limited than the actual student interface used for the summative assessments, particularly with regard to how items appeared on the screen and how erroneous responses were handled. Though reviewers were not able to determine the extent to which these interface limitations impacted their findings, the study team worked with Smarter Balanced to ascertain which item issues were caused by interface differences and which were not. All item-relevant statements in the report reflect data not prone to interface differences.

19. ACT Aspire does not classify literary nonfiction texts that are primarily narrative in structure as "informational." See Appendix G for more information about ACT Aspire's interpretation of CCSSO criterion B.1.

The items are generally high quality and test forms at grades 5 and 8 have a range of cognitive demand, but in each case the distribution contains significantly greater emphasis at DOK 3 than reflected in the standards. Thus, students who score well on the assessments will have demonstrated a strong understanding of the standards' more complex skills. However, the grade 8 test may not fully assess standards at the lowest level of cognitive demand. The tests would better meet the CCSSO Criteria with an increase in the number of items focused on the major work of the grade and the addition of more items at grade 8 that assess standards at DOK 1.

Content: ACT Aspire receives a Limited/Uneven Match to the CCSSO Criteria for Content in Mathematics. The program does not focus exclusively on the major work of the grade, but rather, by design, assesses material from previous and later grades. This results in a weaker match to the criteria. The tests could better meet the criteria at both grades 5 and 8 by increasing the number of items that assess the major work of the grade.

Depth: ACT Aspire receives a good match to the CCSSO Criteria for Depth in Mathematics. The items are well crafted and clear, with only rare instances of minor editorial issues. The ACT Aspire tests include proportionately more items at high levels of cognitive demand (DOK 3) than the standards reflect and proportionately fewer at the lowest level (DOK 1). This finding is both a strength, in terms of promoting strong skills, and a weakness, in terms of ensuring adequate assessment of the full range of cognitive demand within the standards. While technically meeting the criterion for use of multiple item types, the range is nonetheless limited, with the majority comprising multiple-choice items. The program would better meet the criteria for Depth by including a wider variety of item types and relying less on traditional multiple-choice items.

MCAS

English Language Arts:

In ELA/Literacy, MCAS receives a Limited/Uneven to Good Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. The test requires students to closely read high-quality texts and a variety of high-quality item types. However, MCAS does not adequately assess several critical skills—including reading informational texts, writing to sources, language skills, and research and inquiry; further, too few items assess higher-order skills. Addressing these limitations would enhance the ability of the test to signal whether students are demonstrating the skills called for in the standards. Over time, the program would also benefit by developing the capacity to assess speaking and listening skills.

Content: MCAS receives a Limited/Uneven Match to the CCSSO Criteria for Content in ELA/Literacy. The assessment requires students to read closely well-chosen texts and presents test questions of high technical quality. However, the program would be strengthened by assessing writing annually, assessing the three types of writing called for across each grade band, requiring writing to sources, and placing greater emphasis on assessing research and language skills.

Depth: MCAS receives a rating of Good Match for Depth in ELA/Literacy. The assessments do an excellent job in presenting a range of complex reading texts. To fully meet the demands of the CCSSO Criteria, however, the test needs more items at higher levels of cognitive demand, a greater variety of items to test writing to sources and research, and more informational texts—particularly those of an expository nature.

Mathematics:

In mathematics, MCAS receives a Limited/Uneven Match to the CCSSO Criteria for Content and an Excellent Match for Depth relative to assessing whether students are on track to meet college and career readiness standards. The MCAS mathematics test items are of high technical and editorial quality. Additionally, the content is distributed well across the breadth of the grade level standards, and test forms closely reflect the range of cognitive demand of the standards. Yet the grade 5 tests have an insufficient degree of focus on the major work of the grade.

While mathematical practices are required to solve items, MCAS does not specify the assessed practices(s) within each item or their connections to content standards. The tests would better meet the criteria through increased focus on major work at grade 5 and identification of the mathematical practices that are assessed—and their connections to content.

Content: MCAS receives a Limited/Uneven Match to the CCSSO Criteria for Content in Mathematics. While the grade 8 assessment focuses strongly on the major work of the grade, the grade 5 assessment does not, as it samples more broadly from the full range of standards for the grade. The tests could better meet the Criteria through increased focus on the major work of the grade on the grade 5 test.

Depth: MCAS receives an Excellent Match to the CCSSO Criteria for Depth in Mathematics. The assessment uses high-quality items and a variety of item types. The range of cognitive demand reflects that of the standards of the grade. While the program does not code test items to math practices, mathematical practices are nonetheless incorporated within items. The program might consider coding items to the mathematical practices and making explicit the connections between specific practices and content standards.

PARCC

English Language Arts:

In ELA/Literacy, PARCC receives an Excellent Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. The tests include suitably complex texts, require a range of cognitive demand, and demonstrate variety in item types. The assessments require close reading, assess writing to sources, research, and inquiry, and emphasize vocabulary and language skills. The program would benefit from the use of more research tasks requiring students to use multiple sources and, over time, developing the capacity to assess speaking and listening skills.

Content: PARCC receives an Excellent Match to the CCSSO Criteria for Content in ELA/Literacy. The program demonstrates excellence in the assessment of close reading, vocabulary, writing to sources, and language, providing a high-quality measure of ELA/Literacy content as reflected in college and career readiness standards. The tests could be strengthened by the addition of research tasks that require students to use two or more sources and, as technologies allow, a listening and speaking component.

Depth: PARCC receives a rating of Excellent Match for Depth in ELA/Literacy. The PARCC assessments meet or exceed the depth and complexity required by the Criteria through a variety of item types that are generally of high quality. A better balance between literary and informational texts would strengthen the assessments in addressing the Criteria.

Mathematics:

In mathematics, PARCC receives a Good Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. The assessment is reasonably well aligned to the major work of each grade. At grade 5, the test includes a distribution of cognitive demand that is similar to that of the standards. At grade 8, the test has greater percentages of higher-demand items (DOK 3 and 4) than reflected by the standards, such that a student who scores well on the grade 8 PARCC assessment will have demonstrated strong understanding of the standards' more complex skills. However, the grade 8 test may not fully assess standards at the lowest level (DOK 1) of cognitive demand.

The test would better meet the CCSSO Criteria through additional focus on the major work of the grade, the addition of more items at grade 8 that assess standards at DOK 1, and increased attention to accuracy of the items—primarily editorial, but in some instances mathematical.

Content: PARCC receives a Good Match to the CCSSO Criteria for Content in Mathematics. The test could better meet the criteria by increasing the focus on the major work at grade 5.

Depth: PARCC receives a Good Match to the CCSSO Criteria for Depth in Mathematics. The tests include items with a range of cognitive demand, but at grade 8, that distribution contains a higher percentage of items at the higher levels (DOK 2 and 3) and significantly fewer items at the lowest level (DOK 1). This finding is both a strength, in terms of promoting strong skills, and a weakness, in terms of ensuring adequate assessment of the full range of cognitive demand within the standards. The tests include a variety of item types that are largely of high quality. However, a range of problems (from minor to severe) surfaced relative to editorial accuracy and, to a lesser degree, technical quality. The program could better meet the Depth criteria by ensuring that all items meet high editorial and technical standards and by ensuring that the distribution of cognitive demand on the assessments receives sufficient information across the range.

Smarter Balanced

English Language Arts:

In ELA/Literacy, Smarter Balanced receives a Good to Excellent Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. The tests assess the most important ELA/Literacy skills of the CCSS, using technology in ways that both mirror real-world uses and provide quality measurement of targeted skills. The program is most successful in its assessment of writing and research and inquiry. It also assesses listening with high-quality items that require active listening, which is unique among the four programs. The program would benefit by improving its vocabulary items, increasing the cognitive demand in grade 5 items, and, over time, developing the capacity to assess speaking skills.

Content: Smarter Balanced receives an Excellent Match to the CCSSO Criteria for Content in ELA/Literacy. The program demonstrates excellence in the areas of close reading, writing to sources, research, and language. The listening component represents an important step toward adequately measuring speaking and listening skills—a goal specifically reflected in the standards. Overall, Smarter Balanced is a high-quality measure of the content required in ELA/Literacy, as reflected in college and career readiness standards. A greater emphasis on Tier 2 vocabulary would further strengthen these assessments relative to the criteria.

Depth: Smarter Balanced receives a rating of Good Match for Depth in ELA/Literacy. The assessments use a variety of item types to assess student reading and writing to source. The program could better meet the depth criteria by increasing the cognitive demands of the grade 5 assessment and ensuring that all items meet high editorial and technical quality standards.

Mathematics:

In mathematics, Smarter Balanced has a Good Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. The test provides adequate focus on the major work of the grade, although it could be strengthened at grade 5.

The tests would better meet the CCSSO Criteria through increased focus on the major work at grade 5 and an increase in the number of items on the grade 8 tests that assess standards at the lowest level of cognitive demand. In addition, removal of serious mathematical and/or editorial flaws, found in approximately one item per form, should be a priority.[20]

Content: Smarter Balanced receives a Good Match to the CCSSO Criteria for Content in Mathematics. The tests could better meet the criteria by increasing the focus on the major work for grade 5.

Depth: Smarter Balanced receives a Good Match to the CCSSO Criteria for Depth in Mathematics. The exam includes a range of cognitive demand that fairly represents the standards at each grade level. The tests have a strong variety of item types including those that make effective use of technology. However, a range of problems (from minor to severe) surfaced relative to editorial accuracy and, to a lesser degree, technical

20. See footnote 18 for more on Smarter Balanced test interface.

quality. A wide variety of item types appear on each form, and important skills are assessed with multiple items, as is sound practice. The program could better meet the Depth criteria by ensuring that all items meet high editorial and technical standards and that a given student is not presented with two or more virtually identical problems.

******

For too many years, state assessments have generally focused on low-level skills and have given parents and the public false signals about students' readiness for postsecondary education and the workforce. They often weren't very helpful to educators or policymakers either. States' adoption of college and career readiness standards has been a bold step in the right direction. Using high-quality assessments of these standards will require courage: these tests are tougher, sometimes cost more, and require more testing time than the previous generation of state tests. Will states be willing to make the tradeoffs?

Introduction

Overview

Approximately one-third of American freshmen at two-year and four-year colleges require remedial coursework, and over 40 percent of employers rate new hires with a high school diploma as "deficient" in their overall preparation for entry-level jobs.[21, 22] Yet over the past decade, as these students marched through America's public education system, officials repeatedly told them, and their parents, that they were on track for success. They passed their courses, got good grades, and aced state annual tests. To put it plainly, it was all a lie. Imagine being told year after year that you're doing just fine—only to find out when you apply for college or a job that you're simply not as prepared as you need to be.

Thankfully, states have taken courageous steps to address this preparedness gap. Over the past five years, every state has upgraded its K–12 academic standards to align with college and career readiness, either by adopting the Common Core State Standards or by working with its own higher education and career training providers to strengthen or develop standards. But whether or not these improved standards will be faithfully implemented in the classroom depends a great deal on whether enough states adopt rigorous, high-quality assessments, and whether those assessments truly measure the complex skills and knowledge called for by the standards. If the tests are weak measures that, like many of their predecessors, focus too much on low-level and easy-to-measure skills, they make it less likely that the standards will achieve their ultimate goal of ensuring that high school graduates are indeed prepared for successful transition into the job market or higher education.

The selection of state assessments has recently been a topic of great debate in policy and political circles. As new assessments intended to align to the new college and career readiness standards made their debut in the past year or two, student proficiency rates have (predictably and inevitably) been lower—sometimes significantly so—than on previous tests, forcing state and local officials nationwide to make critical decisions. Should they forge ahead with the new tests and higher expectations or back down in order to cast more schools and students in a positive (if, perhaps, illusory) light?

Of course, test scores that more accurately predict students' readiness for entry-level coursework or training are not enough. The content of state assessments, too, is an important predictor of the impact of those tests on what is taught and learned. For instance, low-quality assessments with poorly aligned tasks undermine the content messages of the standards; given the tests' role in accountability, it is only logical that such tests might contribute to poor-quality instruction.

In short, good tests matter. Of critical importance to this conversation, however, is whether the new tests are indeed good and worth fighting for. That's the central question this study seeks to answer.

This report, then, provides much-needed information to policymakers, practitioners, and researchers about the quality of four current assessments and their potential to support effective implementation of college and career readiness standards. It takes an unprecedented, in-depth look at three new multi-state assessments and an

21. NCES, *Percentage of First-Year Undergraduate Students who Took Remedial Education Courses*.

22. Conference Board et al., "Are They Really Ready To Work?"

existing best-in-class state assessment. All four assert that they measure college and career readiness standards generally and the CCSS specifically. The study results are particularly relevant to states that have yet to make final decisions on which assessments they'll use in future years to measure student learning in English language arts/literacy and mathematics.

The Four Assessments

Three of the assessments evaluated in this study are currently in use across multiple states: the Partnership for the Assessment of Readiness for College and Careers (PARCC), the Smarter Balanced Assessment System (Smarter Balanced), and ACT Aspire. The first two are membership-based organizations that are governed by participating states;[23] the latter is developed by the maker of the widely known ACT college admissions test.

The fourth assessment, the 2014 version of the Massachusetts Comprehensive Assessment System (MCAS), is a highly regarded state assessment program from a state that has consistently outscored the rest of the nation on a host of achievement metrics.[24, 25] In 2011, Massachusetts adopted new curriculum frameworks in mathematics and ELA/Literacy, which incorporated the CCSS with Massachusetts-specific standards, and began to transition the MCAS assessments accordingly. By the spring of 2014, the transition was complete: MCAS included items that covered the CCSS and additional Massachusetts-specific standards.[26] Within this study, the 2014 MCAS serves as a comparison point or "best-case" example for the solo state option.

Study Design and Key Questions

The study utilizes a new methodology developed by the National Center for the Improvement of Educational Assessment (NCIEA) to determine how well new tests measure the requisite content, knowledge, and critical skills at key grade levels and, in doing so, whether they sufficiently tap higher-order thinking skills.

This study evaluates the summative (end-of-year) assessments administered in grades 5 and 8, the capstone grades for the elementary and middle school levels.[27] A parallel study led by the Human Resources Research Organization (HumRRO) evaluates the capstone assessment in grades 9–12 (typically administered at grade 10 or 11, depending on the program). Because both organizations used the same methodology, it made sense to conduct two portions of the review jointly—the documentation review and the accessibility review. The composition of these joint panels is described below. (HumRRO has published a separate report with the results of the evaluation of the high school assessments, which will also include findings from our joint accessibility review.)[28]

23. Congress and the U.S. Department of Education allocated $350 million in federal grants to groups of states seeking to develop new assessments aligned to college and career readiness standards. Two consortia of states were awarded Race to the Top assessment grants: the Partnership for Assessment of Readiness for College and Careers (PARCC) and the Smarter Balanced Assessment Consortium. By 2012, forty-five states were participating either as governing or participating (advisory) members in either or both of the consortia. That number is now 21 states, however, as of January 2016.

24. V. Bandeira de Mello et al., "Mapping State Proficiency Standards onto NAEP Scales: 2005–2007" (Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences, 2009).

25. National Assessment Governing Board, "Nation's Report Card," 2015, http://www.nationsreportcard.gov/reading_math_2015/.

26. MCAS items assessing standards unique to Massachusetts—meaning not part of the CCSS—were removed from the test forms prior to use in this study. In 2014, school districts in Massachusetts were allowed to choose either MCAS or PARCC as their summative assessment.

27. Some of these testing programs provide aligned formative and/or benchmark/interim assessments, which are not used for consequential purposes. Those assessments were not part of this study.

28. This study also originally included an evaluation of test program transparency, or the extent to which programs provide sufficient information to the public regarding assessment design and expectations (CCSSO criterion A.6). Due to several challenges associated with this review, however, we ultimately decided to drop this criterion from our study. Review panelists were not able to review all relevant documentation for each program, due to the vast volume of materials provided and publicly available. In addition, many test programs continued to release additional information (such as sample items) since our review occurred, rendering this panel's findings somewhat outdated.

TABLE 1

Parallel Makeup of the Fordham and HumRRO Studies

	ELA/Literacy Review	Math Review	Documentation Review	Accessibility Review
Fordham Study	Grades 5 and 8	Grades 5 and 8	Joint Panel	Joint Panel
HumRRO Study	High School	High School	(grades 5 and 8 findings presented in this report; high school findings presented in HumRRO report)	(presented in HumRRO report)

Together, these studies provide the public with the first in-depth look at this new generation of highly touted assessments.

Our study was designed to address this trio of questions:

1. Do the assessments place strong emphasis on the most important content for college and career readiness (CCR), as called for by the Common Core State Standards and other CCR standards? (**Content**)
2. Do they require all students to demonstrate the range of thinking skills, including higher-order skills, called for by those standards? (**Depth**)
3. What are the overall strengths and weaknesses of each assessment relative to the examined criteria for ELA/Literacy and mathematics? (**Overall Strengths and Weaknesses**)

Questions 1 and 2, regarding the Content and Depth of the assessments, are the major emphases of this evaluation, and Question 3 highlights the major findings of the panels.

Organization of the Report

The report is organized as follows. First, we take a high-level look at the ways in which this study's approach to the evaluation of test quality differs significantly from previous approaches and describe major design differences within and among the four participating test programs.

Next, we explain the methodology developed by the NCIEA and how it was operationalized for this study (Section I). The findings from our mathematics and English language arts/literacy review follow this (see Section I, *Results*), then the section concludes with a summary of the various strengths and weaknesses of each testing program relative to the CCSSO Criteria. Section II includes our recommendations for policymakers and test developers, and we conclude in Section III by offering suggestions for future improvements to the methodology. The multiple appendices include, among other topics, a discussion of Depth of Knowledge (DOK), key terminology, the biographies of the authors and panelists, and responses to the study from the testing programs.

Approach to Alignment and Quality

To evaluate assessments intended to measure student mastery of the Common Core State Standards (CCSS), we needed a methodology that would capture their key dimensions. Traditional alignment methodologies, such as Webb's alignment tool[29] and the Surveys of Enacted Curriculum,[30] offer the advantage of having been studied extensively, but they measure only a fairly narrow range of test-standards alignment issues. For instance,

29. N. L. Webb, *Alignment of Science and Mathematics Standards and Assessments in Four States*, Research Monograph No. 18 (Madison, WI: University of Wisconsin–Madison, National Institute for Science Education, 1999).

30. A. C. Porter, "Measuring the Content of Instruction: Uses in Research and Practice," *Educational Researcher* 31, no. 7 (2002): 3–14.

neither of these methodologies could be used to report on issues such as a math test's coverage of the CCSS for Mathematical Practice, or whether an ELA/Literacy test contains passages with appropriate text complexity for the grade level.

Because of the limitations of existing approaches, a new methodology was created based on the CCSSO's "Criteria for Procuring and Evaluating High-Quality Assessments."[31] Specifically designed to address tests of college and career readiness, these criteria focus the evaluation on the highest-priority skills and knowledge at each grade in the CCSS, addressing foundational as well as complex skills.

The methods were developed by experts at the NCIEA and refined in several cycles of revision. They represent a very different approach to gauging test quality and alignment than earlier generations of such studies. Here we discuss two of these fundamental differences, both of which are the result of critical design attributes of the CCSS and similar college and career readiness standards.

First, most prior alignment approaches assume a one-to-one match between standards and test items. The CCSS and kindred standards of college and career readiness, however, describe numerous complex competencies that cannot be assessed with individual test items, such as the ability to craft clearly written arguments supported by evidence drawn from multiple sources or the ability to apply multiple mathematical skills to solve a complex problem. Such competencies are essential for postsecondary education, training, and citizenship, and states need to know whether their assessments are measuring them. This new approach to alignment and quality asks evaluators to determine the degree to which the tests measure these more complex competencies.

The second key difference of this approach is that it focuses the evaluation on the highest-priority skills and knowledge at each grade. Prior approaches treated each of the grade-level standards with equal importance, creating an inadvertent incentive for tests—and instruction—to be "a mile wide and an inch deep." The CCSS clearly define the prioritized skills and competencies at each grade level and call upon both instruction and assessment to focus strongly, although not exclusively, on these priorities. By focusing the review on the critical competencies, this methodology rewards those tests that give clear signals about the instructional priorities for each grade.

Test Program Differences

Although each of the four tests in this study was developed to assess college and career readiness standards, they differ in significant ways. These differences provide important context for interpreting the results of our study.

1. **Test Design:** Content standards are not written to guide the creation of assessments; rather, they are written to define the specific content and skills students are to master at each grade level. Thus, in creating assessments from standards, programs must first interpret and bundle standards into assessment targets that define what is to be measured. Even though each testing program was tasked with assessing student learning against college and career readiness standards, and the CCSS in particular, the level of emphasis on specific skills and knowledge—such as explaining mathematical reasoning—varies. Simply put, some programs, by design, follow the contours of the CCSS more closely than others, and the impact of these test-design choices is reflected in the study results. As another example, ACT Aspire's ELA/Literacy assessment is broken into three components—reading, writing, and English—while the other programs do not have this split.

31. Council of Chief State School Officers (CCSSO), "Criteria for Procuring and Evaluating High-Quality Assessments" (Washington, D.C.: CCSSO, 2014).

2 **Test Delivery:** The 2014 MCAS is a paper-based assessment, whereas the other three are designed and primarily delivered as computer-based tests (although each currently offers a paper version). Administering the test via computer offers a wider array of options for the presentation of test items as well as how student responses are entered. For example, audio, video, and animation can be used in the presentation of computerized test items, and student responses may involve manipulation of data to identify patterns or live editing of texts with grammatical errors. Paper-based versus computer delivery also impacts the degree to which a test program can assess certain standards, such as listening and modeling skills.[32]

3 **Test Forms and Computer Adaptivity:** The set of items presented to a student in a test form is predetermined for the PARCC, MCAS, and ACT Aspire assessments. Smarter Balanced assessments, however, are computer adaptive, meaning that items or small groups of items are selected as the student proceeds through the test, based at least in part on all of the student's prior responses. Their summative tests have an adaptive and a performance task (PT) component. The PT is a fixed set of items that are not chosen adaptively. The adaptive portions of the test adjust after each item is administered and scored, except in cases in which multiple questions are based on the same reading or listening passage(s) or scenario. Such groups of items are administered as a unit.[33] The majority of students see only on-grade-level items, but those performing near the ends of the performance spectrum may be given off-grade-level items near the end of the test to better assess their performance and increase the precision of their score.

This performance-based adaptation allows the test to produce scores with smaller margins of error for students near the ends of the performance spectrum and/or to shorten the overall length of the test for most students.[34] It also, however, creates the potential for greater variation in content across student forms, which could impact the degree to which individual forms meet the CCSSO Criteria. This study followed the processes recommended by the NCIEA for the creation of simulated Smarter Balanced test forms and, in order to evaluate the degree to which other forms would vary, also included the results of a simulation study of 1,000 forms per grade and content area.

4 **Testing Time:** Test developers are forced to strike a balance between accurately measuring student learning and the time and cost required to do so. Our four testing programs show different solutions to this balancing act. The estimated testing time for students in grades 5 and 8, on average, to complete both the ELA/Literacy assessments and the mathematics assessments for the respective programs are as follows:

- ACT Aspire: three to three-and-a-quarter hours for all four tests (English, reading, writing, and mathematics)
- MCAS: three-and-a-half hours
- PARCC: seven to seven-and-a-half hours (the 2015–16 revisions will reduce this by an estimated one-and-a-half hours)[35]
- Smarter Balanced: five-and-a-half hours

The longer testing times for PARCC and Smarter Balanced are primarily due to the inclusion of the extended performance tasks. Both programs use these tasks to assess high-priority skills within the CCSS, such as the development of written compositions in which a claim is supported with evidence drawn from

32. The CCSS define modeling as the process of choosing and using appropriate mathematics and statistics to analyze empirical situations, to understand them better, and to improve decisions.

33. The Smarter Balanced adaptive engine, which selects the items to be presented to the student, is programmed to first ensure that the requirements in the test blueprint for content and cognitive demand are met. Within these constraints, it also adapts to improve score precision.

34. In CAT, the algorithm is typically programmed to continue asking a student questions until a certain level of precision is achieved OR until other conditions are met, such as a time limit or the assigned item pool is exhausted. See, for example, H. Wainer et al., *Computerized Adaptive Testing: A Primer* (2000).

35. See Appendix G for a description of the changes made to the 2015–16 versions of PARCC and the other three assessments, as relevant.

sources; research skills; and solving complex multi-step problems in mathematics. In addition to requiring more time than selected-response items, these tasks are typically more costly to develop and score.

5 **Ownership of the Assessments and Setting Proficiency Cut Scores**: PARCC and Smarter Balanced are governed by their respective groups of member states, which collectively set, as required by the initial grants, the proficiency cut scores to be used by all members for federal reporting of student performance. Each member state, however, may determine the cut scores to be used within its own state accountability systems—cut scores that may impact grade-to-grade promotions, the awarding of high school diplomas, and/or educator evaluations.[36]

In contrast, the ACT Aspire assessments are the property of ACT Aspire. Individual states contract with ACT Aspire for their use and may set their own proficiency cut scores to be used for both federal and state accountability purposes. For grade 10, ACT Aspire is able to inform states' cut-score decisions with historical data regarding the minimum scores associated with a high likelihood of success in credit-bearing first-year college courses[37] (i.e., the ACT Aspire College Readiness Benchmarks).

The MCAS is a custom assessment developed and owned by the state of Massachusetts. State policymakers are charged with setting the proficiency cut scores used for both federal and state accountability purposes.

Next we turn our attention to the methodology.

36. Both Smarter Balanced and PARCC plan to analyze the relationship between student performance on the high school assessments and subsequent performance in entry-level post-secondary courses to inform future state decision making.

37. "Success" is defined as a 50 percent chance of obtaining a B or higher, or about a 75 percent chance of obtaining a C or higher, in corresponding credit-bearing first-year college courses.

Section I: Assessments of English Language Arts/Literacy and Mathematics

Overview of the Methodology

In 2014, the Council of Chief State School Officers (CCSSO) released a set of criteria approved by the chiefs for evaluating the new generation of assessments designed to measure college and career readiness standards, the "Criteria for Procuring and Evaluating High-Quality Assessments."[38] CCSSO's stated goal was to provide guidance and support to states seeking to ensure that new assessments aligned to college and career readiness standards are not only valid, reliable, and fair, but also "match the depth, breadth, and rigor of the standards; accurately measure student progress toward college and career readiness; and provide valid data to inform teaching and learning."

Under the leadership of Brian Gong and Scott Marion, the NCIEA utilized these criteria as the basis for a new methodology for evaluating assessments based on college and career readiness standards, including the CCSS. Our study and the parallel high school study conducted by HumRRO are the first to implement this new methodology.

Below we provide an overview of its key elements. We first describe the various phases of the evaluation—from the review of individual test items to the creation of final program-level summary statements. Next, we present details on the selection of reviewers and test forms as well as the assignment of reviewers to forms. Third, we review the CCSSO Criteria and describe how they were used to address our key questions. Finally, we describe several modifications made to the NCIEA-written methodology, due to issues that arose during implementation.

Study Phases

The study comprises two major phases. The first is a review of the test items (the "item review") and documentation (the "generalizability review"). The second is the development of program-level (overall) ratings for each CCSSO criterion.

The heart of the evaluation is the review of actual test items and forms. The results of the item reviews are "rolled up" to the form levels, which are then rolled up to the grade level, then to the overall-testing-program level—the latter of which comprise the ratings.

Phase 1: Item and Documentation Review

The item review requires individual evaluation of actual student test items against each of the subject area sub-criteria outlined in the CCSSO Criteria. These sub-criteria are described alongside the criterion-level results in the *Results* section.

38. Council of Chief State School Officers (CCSSO), "Criteria for Procuring and Evaluating High-Quality Assessments" (Washington, D.C.: CCSSO, 2014).

The programs were first asked to provide test item metadata. Metadata refer to descriptive data about test items and passages, such as their alignment to standards, their level of text complexity (i.e., qualitative and quantitative measures of the complexity of text passages), and item type (e.g., multiple choice, constructed-response, and technology-enhanced). These metadata were pre-populated into customized electronic coding sheets and used throughout the review process. Reviewers used these coding sheets to move through each test form one item at a time, rating each item on each of the applicable dimensions. Some of these ratings relied on the metadata provided by the programs, and others were strictly based on expert judgment.[39]

The methodology also requires that panels evaluate each program's documentation to determine whether the results from reviewing one or two test forms per grade are likely generalizable to the test program as a whole (see *Selection of Forms* for more). In other words, would item-review ratings likely remain the same, improve, or decline if all possible test forms built from the same blueprints and other test specifications had been reviewed?[40]

For the documentation review, testing programs were provided with the criteria to be evaluated in the study and asked to provide documentation that reviewers could use to assess them. Each program provided a large volume of materials they deemed to support the criteria, such as test blueprints, item specifications, and cognitive-demand definitions. These materials detailed their approach to building and validating their assessments to demonstrate that they adequately assessed college and career readiness.

Phase 2: Development of Program Ratings

In accordance with the methodology, reviewers next aggregated the ratings of individual panelists across test forms, grade levels, and programs.

To begin this process, panelists first met to discuss their individual scores and comments for each test form reviewed.[41] The panels then decided on a final "group match score" for each form, which was simply their agreed-upon score after talking through rater differences and settling on a score that represented the collective wisdom of the panel.[42]

Next, review panels developed final sub-criterion ratings and comments for each grade overall, taking into account the final scores for the two forms at that grade level. Reviewers then inspected each program's scores across the grade 5 and grade 8 forms, considered the results of the documentation review, and came to consensus on the degree to which each program "matched" the respective CCSSO criterion. They issued final ratings on the following scale: Excellent, Good, Limited/Uneven, or Weak Match to the criterion.[43] They also developed summary statements with a rationale for the ratings. Finally, review panels developed ratings for the Content and Depth of each assessment, based on the prioritization of criteria recommended in the study methodology (see Section I, Table 3). They also generated final statements summarizing the observed strengths and areas of improvement for each program.

In short, each consecutive step of the evaluation built upon the prior step, which helped to foster shared understanding among reviewers and strengthen the internal consistency of the results. Figure 1 illustrates the entire process for arriving at the final Content and Depth scores, as well as the final summary statements for each assessment program.

39. Field-tested items were not included in the formal review.

40. This portion of the study was conducted jointly between HumRRO and Fordham.

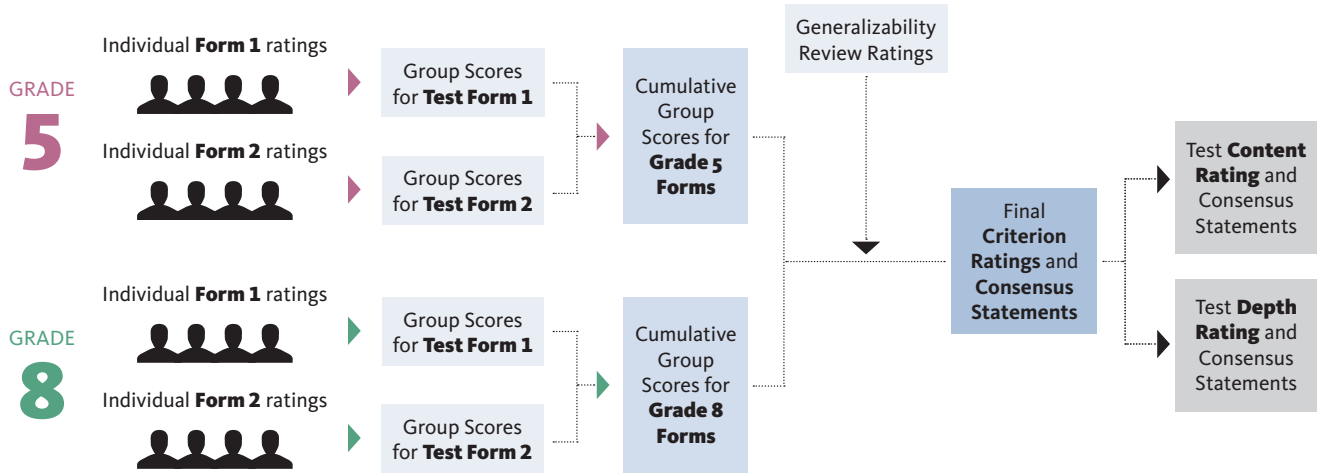
41. Per the methodology, two forms were reviewed for each program per grade and subject area (except for MCAS, which had only one form per grade).

42. In instances where they were unable to reach consensus, minority opinions are reflected in the final statements at the criterion and program level.

43. If the program documentation supported reviewers' form-level findings, suggesting that the results would hold regardless of the number of forms reviewed (i.e., forms developed in different years that are based on the same test blueprints and specifications), the final criterion rating was based solely on the aggregation of the grade 5 and 8 form-level ratings, as explained in the methodology. If, however, the documentation did not demonstrate that the rating would remain the same if additional forms had been reviewed, the panel determined whether or not to adjust the final criterion rating and, if so, stated the rationale.

FIGURE 1

The Process for Generating Final Program Ratings for ELA/Literacy and Mathematics



Selection of Test Forms

The NCIEA's methodology has as its primary goal that evaluators obtain an accurate measure of each test program's content and quality within reasonable time and cost parameters. Thus, for each assessed program, the methodology recommends reviewing two forms, which "should be sufficient basis for an evaluation when coupled with a review of generalizability documentation." This recommendation allows implementers to test for consistency of test quality across forms and helps ensure that results are not idiosyncratic to a particular form. Further, in the case of ELA/Literacy, since each test form can accommodate only a handful of text passages and writing prompts, evaluating two forms gives reviewers a broader look at quality. We followed this two-form recommendation per content area and grade level.[44]

Programs were asked to select two operational forms of their choosing for each subject at each grade level.[45] They could not be "special forms" (e.g., used only for students with particular special needs).

As indicated, Smarter Balanced is a computer-adaptive test (CAT), meaning each student taking the test can receive a different set of questions based on their response to prior questions (see *Test Program Differences* in the Introduction). Thus there exist hundreds or more possible forms to consider. Methods for analyzing the alignment of computer adaptive assessments are in their infancy.[46] One approach would be to simply analyze a representative sample of the item pool, ignoring how items were placed on student test forms. Yet doing so would violate a primary study objective, which is to determine whether complete test forms administered to students meet the CCSSO Criteria.

Instead, the NCIEA methodology recommends analyzing two test forms for computer-adaptive tests, one for a student at the 40th percentile of achievement and the other at the 60th percentile, in order to represent a range of possible forms that "typical students" are likely to see. To augment the two forms, the NCIEA also recommends that the program provide simulation data that specifically summarize the characteristics of 1,000

44. The one exception was for MCAS, which has only one operational test form per subject and grade available for review.

45. An operational form refers to a form that was actually used by students.

46. S. L. Wise et al., "Evaluating Content Alignment in Computerized Adaptive Testing," *Educational Measurement: Issues and Practice* (2015; published online ahead of print).

simulated forms per grade and content area. Smarter Balanced provided such information for this study, and the NCIEA summarized those results for our panelists and answered their related questions. The summary explained whether the simulation data provided confidence that the results observed on two forms were likely to apply across the full range of Smarter Balanced forms.

We applied the methodology as written for computer adaptive tests and are confident in the accuracy of our ratings for the forms that panelists reviewed. Still, if we had chosen a different approach to evaluating CAT, such as analyzing larger numbers of forms and at additional student percentiles, our results for Smarter Balanced may have been different.

Selection of Review Panels and Assignment to Forms

Selection and Recruitment of Panelists

The quality and credibility of a complex evaluation of this type rests largely on the expertise and objectivity of the individuals serving on its review panels, such that it is critical to recruit highly qualified yet impartial reviewers with diverse experience and backgrounds. For each panel, we sought to recruit for each panel a mix of practitioners, content experts, and assessment experts.

We began by soliciting reviewer recommendations from each participating testing program and other sources, including content and assessment experts, individuals with experience in prior alignment studies, and several national and state organizations. Finalists were asked to submit CV as well as detailed responses to a questionnaire about their familiarity with the Common Core State Standards, prior experience in conducting alignment evaluations, and potential conflicts of interest. Follow-up phone calls were conducted as necessary. Individuals currently or previously employed by participating testing organizations and writers of the CCSS were not considered. Given that most content and assessment experts have become such by working on prior alignment or assessment-development studies, and that it appeared impossible to find individuals with zero conflicts who are also experts, we prioritized balance and fairness. We recruited at least one reviewer recommended by each testing program to serve on each panel; this strategy helped to ensure fairness by equally balancing reviewer familiarity with the various assessments.[47]

In addition, two university-affiliated content leads facilitated the work of the ELA/Literacy and math review panels. Dr. Charles Perfetti, Distinguished University Professor of Psychology at University of Pittsburgh, served as the ELA/Literacy content lead and Dr. Roger Howe, Professor of Mathematics at Yale University, served as the mathematics content lead. The names and biographical summaries of all panelists appear in Appendix E.

Reviewers received several days of training, both online and in-person, prior to conducting the analysis. See *How Were Reviewers Trained?* for more.

47. One reviewer in ELA/Literacy grade 5 was forced to drop out of the study just before the item review due to medical issues; we were unable to replace her, so some ELA/Literacy grade 5 forms were reviewed by three (and not four) individuals.

How Were Reviewers Trained?

Training for our item review was delivered via several online and in-person sessions. Online training began with a two-hour webinar that described the broader review process and briefly introduced the CCSSO Criteria. The second element of the online training was a series of subject-specific, one-hour presentations by the test vendors describing important features of their assessments, such as their approach to alignment and their form specifications. Panelists also individually reviewed overviews of the methodology and the CCSS for their assigned subject and grade.[48]

In-person training was conducted in Washington, D.C. in the summer of 2015. The three-day training included a series of PowerPoint training modules ranging from one to three hours, each detailing a relevant criterion. Modules laid out the intention of the criterion, the procedures for rating each test item and reading passage, and included a guided-practice activity using sample items. (They also addressed the meaning of key terms used in the methodology.) These modules were developed by Student Achievement Partners (SAP) with input and editing from the study team. The training was delivered by SAP, HumRRO, Dr. Morgan Polikoff, and several consultants, including alignment and assessment experts. Reviewers also received training on how to access the online systems of the various assessments and to correctly complete the scoring worksheets.

Following this training, reviewers independently completed a calibration exercise (led by Dr. Polikoff) using released test items from the 2014 New York Common Core-aligned exam. Individual ratings were collated and examined for disagreement, which was defined as less than 80 percent agreement across raters on any item or passage rating. They were subsequently discussed in-depth at the start of day three, in order to improve consensus, increase inter-rater reliability, and attempt to clarify misunderstandings of the criteria. At this point, the rating of actual test forms began.

The subset of reviewers who participated in the documentation review were also provided training and support by assessment consultant Dr. Jami-Jon Pearson.

Assignment to Forms

There were thirty-two total reviewers across subjects and grade levels. Of these, sixteen were practitioners (practicing teachers in the content area), eight were content experts (from higher education or content area consultants), and eight were assessment experts. Thus, for each grade level and subject there were eight total reviewers, four of whom were practitioners, two who were content experts, and two who were assessment experts (though some had expertise in multiple areas).

For the item review, reviewers were stratified by expertise (practitioners, content experts, and assessment experts) and randomly assigned to test forms. See Table 2 below, which illustrates this jigsaw arrangement.

For each content area and grade level, a total of eight experts reviewed six total forms from ACT Aspire (two forms), PARCC (two forms), and Smarter Balanced (two forms), and eight experts reviewed the form from MCAS. This approach ensured that there were four reviewers for each of the seven test forms at each grade level—see in Table 2 that each column has four X's indicating the four reviewers. These four reviewers included two practicing teachers in the content area, one content expert (either from higher education or a consultant), and one assessment expert. This approach also ensured that there were six unique reviewers across the two test forms for each program at each grade level—except for MCAS, which had only one test form.[49]

48. Separate trainings were provided for the documentation analysis, which were conducted online.

49. The difference in the number of experts is due to the number of forms reviewed for each program. Two forms per grade level (grades 5 and 8) and content area were reviewed for ACT Aspire, PARCC, and Smarter Balanced, and only one form per grade and content area for MCAS.

After the item reviews were completed, a subset of reviewers convened to develop final program ratings. This group comprised seven reviewers in each subject drawn from the sixteen who participated in the item review. As was the case with the item review, each subject-area panel included practitioners, assessment experts, and content experts, plus reviewers recommended by each of the four programs.[50] Their deliberations occurred over three days via phone and online meetings.

TABLE 2

Reviewer Assignment Jigsaw for Grade 5 Tests in a Given Subject*

Reviewers	ACT Aspire		PARCC		Smarter Balanced		MCAS
	Form 1	Form 2	Form 1	Form 2	Form 1	Form 2	Form 1
Practitioner 1	✕			✕	✕	✕	
Practitioner 2	✕	✕				✕	✕
Practitioner 3		✕	✕		✕		
Practitioner 4			✕	✕			✕
Content Expert 1	✕			✕		✕	✕
Content Expert 2		✕	✕		✕		
Assessment Expert 1	✕	✕			✕	✕	
Assessment Expert 2			✕	✕			✕

*Note: The same configuration was implemented for Grade 8 in both subjects.

The Study Criteria

The study methodology evaluates assessment programs against a subset of the CCSSO "Criteria for Procuring and Evaluating High-Quality Assessments."[51] The Criteria "focus on the critical characteristics that should be met by high-quality assessments aligned to college and career readiness standards" and are based in part on the seminal Standards for Educational and Psychological Testing.[52] The methodology was created primarily by the NCIEA and refined iteratively over several months in the winter and spring of 2014–15.

The CCSSO Criteria are the backbone of the methodology. These are briefly presented in Tables 3 and 4 (see Appendices A and C for more).

50. Furthermore, all but one of the twenty-eight total test forms (math plus ELA/Literacy) had at least one reviewer on the Phase 2 panel.

51. Council of Chief State School Officers (CCSSO), "Criteria for Procuring and Evaluating High-Quality Assessments" (Washington, D.C.: CCSSO, 2014).

52. American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) Joint Committee on Standards for Educational and Psychological Testing, "Standards for Educational and Psychological Testing" (Washington, D.C.: AERA, 1999).

TABLE 3

The CCSSO Criteria for the Assessment of ELA/Literacy

	Content	Depth
B.1* focuses on text type (narrative vs. informational) and text quality. It calls for a) a variety of text types, b) an increasing focus on diverse informational texts across grades, and c) the use of high-quality text passages.		✕
B.2* focuses on text complexity (qualitative and quantitative). It requires that assessments include texts that have appropriate levels of text complexity for the grade or grade band (grade bands identified in the CCSS are K–5 and 6–12).		✕
B.3* focuses on key dimensions of reading items. It includes test items that require close reading and direct textual evidence and that focus on central ideas and important particulars.	✕	
B.4 focuses on the cognitive demand of the assessment (i.e., the type of student performance that is required to complete each task on the assessment). It requires an appropriate range of cognitive demand that adequately represents the cognitive demand of the standards.		✕
B.5* focuses on key dimensions of writing items. It includes assessing a variety of types and formats of writing and the use of writing prompts that require students to confront and use evidence from texts or other stimuli directly.	✕	
B.6 focuses on key dimensions of language and vocabulary items. It includes placing adequate emphasis on language and vocabulary items on the assessment, assessing vocabulary that reflect requirements for college and career readiness, and focusing on common student errors in language questions.	✕	
B.7 focuses on key dimensions of research items. It expects that items will require students to analyze, synthesize, organize, and use information from multiple sources.	✕	
B.8 focuses on key dimensions of listening and speaking items. While acknowledging that the technology does not yet exist for a complete assessment of these skills, this criterion assesses students' listening skills using passages with adequate complexity and assesses students' speaking skills through oral performance tasks.	✕	
B.9 focuses on high-quality items and a variety of item types. It calls for multiple item types, including at least one type where students construct, rather than select, a response as well as high-quality items lacking technical or editorial flaws.		✕

Notes:

Nine ELA/Literacy criteria are evaluated. These criteria have between two and eight sub-criteria each, which are the actual targets of the review.

*The methodology suggests that criteria B.3 and B.5 receive greater emphasis in developing the Content rating, and that criteria B.1 and B.2 receive greater emphasis in developing a Depth rating.[53] These emphases are indicated by an asterisk.

53. Though the methodology called for B.2 to be emphasized in determining the ELA/Literacy Depth rating, the reviewers ultimately chose not to do so for reasons described in Section III.

TABLE 4

The CCSSO Criteria for the Assessment of Mathematics

	Content	Depth
C.1* focuses on mathematics item content. It requires a heavy focus on the content most crucial for success in later mathematics (the major work of the grade).[55]	✕	
C.2 focuses on the variety of skills assessed. It requires measurement of conceptual understanding, procedural skill, and application in approximately equal proportions.	✕	
C.3* focuses on mathematical practices and their connection to content. The criterion calls for students' use of mathematical practices through test items that connect these practices with grade-level content standards.		✕
C.4 focuses on the cognitive demand of the assessment. As in ELA/Literacy, this criterion requires an appropriate range of cognitive demand that adequately represents the cognitive demand of the standards.		✕
C.5 focuses on high-quality items and a variety of item types. As in ELA/Literacy, this criterion requires multiple item types, including at least one constructed-response type, as well as high-quality items lacking technical or editorial flaws.		✕

Notes:

Five mathematics criteria are evaluated. As in ELA/Literacy, each of these criteria has at least two sub-criteria each of which are the targets of the review.

*The methodology suggests that criterion C.1 receives greater emphasis in developing a Content rating and that criterion C.3 receives greater emphasis in developing a Depth rating.[54] These emphases are indicated by an asterisk.

Methodology Modifications

Fordham and HumRRO were the first organizations to implement the new methodology developed by NCIEA. Thus, while the goal of this initial implementation was to conduct a review as faithful as possible to the original methodology, several logistical or other issues arose that limited our ability to do so. Some pertained to both subjects and others were unique to programs or the content area. We describe seven modifications below.

1 Challenges Associated with Particular Testing Programs

Because the methodology was designed to be neutral with respect to any particular test, it will not perfectly fit each test's design or specifications.

One issue that affected PARCC and Smarter Balanced pertains to the assignment of test items to individual Common Core State Standards. These programs use evidence-centered design (ECD),[56] which is an approach that bases test-design decisions on the key inferences intended to be made about student performance—also called "claims." Neither of these programs creates items through a one-to-one mapping of items to standards; instead they map items to the specific evidence statements or targets for their claims. Because our methodology requires metadata that assume a one-to-one or one-to-many mapping of items to specific standards, both programs had to produce metadata they would not normally produce, and that do not fully correspond to a test-construction philosophy based on claims. As a result, both programs' tests conflict in some way with criteria in the methodology. For instance, ELA/Literacy reviewers were asked to evaluate the alignment ratings for individual reading items, but mapping individual items to specific standards is not how either PARCC or Smarter Balanced designed their alignment.

54. The methodology called for C.3 to be emphasized in the determination of the Depth rating but the panel found this criterion to be poorly operationalized and used their professional judgment to reduce its weight. See Section III, *Suggestions for Methodological Improvement*.

55. A list of the major work standards is available at http://achievethecore.org/content/upload/Focus%20in%20Math_091013_FINAL.pdf.

56. ACT Aspire also uses evidence-centered design, but this issue did not affect them because they had item alignment metadata as well.

A unique issue arose with ACT Aspire regarding CCSSO criterion B1. This criterion places narrative informational texts into the "informational text" category along with expository texts. ACT Aspire includes literary fiction and literary nonfiction texts in its Literary Narrative category, but does not classify literary nonfiction texts that are primarily narrative in structure as informational. Review panelists followed the methodology's guidance when evaluating this criterion.[57]

There was also an issue unique to the MCAS program. For MCAS mathematics, the metadata provided did not include information on mathematical practices. Thus, according to the letter of the C.3 criterion, MCAS did not meet it. However, in reviewing the items, panelists were confident that MCAS items did indeed assess mathematical practices—the program just did not code them in their documentation. Thus panelists took this into account in reaching their final rating for MCAS on this criterion.

2 Item Alignment and Quality

Sub-criteria under B.9 and C.5 call for reviewers to evaluate item quality (including the evaluation of item alignment to standards). In particular, reviewers are asked to verify each of the metadata alignment codes provided by the testing program. Per the methodology, reviewers must agree with all of the alignment codes for 90 percent of test items, in order to receive the top rating for item quality. For multiple reasons,[58] this procedure would have resulted in vast swaths of items (especially in ELA/Literacy) being rated as poorly aligned, leaving test forms rated weakly on this sub-criterion, even though reviewers believed that items were high quality.

The only approach that resulted in conclusions that had face validity to reviewers was to remove item alignment from B.9 and C.5. Thus, these criteria are based only on item quality and not on judgments of the degree to which reviewers agreed with alignment metadata. Alignment to the standards is still considered in the methodology in the criteria under Content in both subjects, but reviewers did not evaluate alignment to the metadata in their review.

3 Cognitive Demand

The procedures for evaluating criteria B.4 and C.4 (which we evaluated using Webb's depth of knowledge, or DOK) were intended to ensure that a) the cognitive demand distribution on the assessments adequately matched the depth of the cognitive demand in the standards and b) that the higher-level cognitive demands of the standards, in particular, were not under-assessed. During implementation, reviewers discovered that the distribution of DOK on some assessments was actually higher on average than in the standards. These exams sometimes received low ratings because the DOK distribution did not match the standards, but some reviewers believed that assessments with higher average cognitive demand than the standards should not necessarily be penalized. In calculating final ratings, reviewers used their professional judgment to adjust these cases higher when needed.

4 Test Complexity Metadata

The B.2 sub-criteria focus on text complexity (quantitative and qualitative measures of the difficulty of a reading passage). In order to evaluate these criteria, reviewers need access to the metadata. However, we were unable to include text complexity data in our analysis for several reasons: 1) vendors often used different methods for evaluating text complexity, 2) qualitative text complexity data varied dramatically across programs, and 3) they were often too voluminous to be displayed in the coding worksheet in any readable format. However, as the documentation review did consider text complexity, these results are used in final scoring.

57. For more on ACT Aspire's interpretation of B1, see Appendix G.

58. For example, some programs indicated that each item aligned to multiple standards; hence, even if reviewers agreed with most of these alignments for a given item, the item would nonetheless be deemed misaligned since the requirement was complete agreement.

5 Major Work of the Grade in Mathematics

The C.1 criterion in mathematics uses the language of focusing "exclusively" on the major work of the grade, which would penalize items that mostly focus on major work but also include some non-major work content. We instructed reviewers to rate an item favorably if it focused "primarily" on major work standards.

6 Application, Conceptual Understanding, and Procedural Skill

Criterion C.2 assesses the extent to which the assessment is adequately balanced across items that assess application, conceptual understanding, and procedural fluency. There were a number of difficulties with the implementation of this criterion; these issues are described in more detail in Section III. The end result is that we chose not to report a criterion-level score, instead providing qualitative feedback only.

7 Weighting of Criteria for Content and Depth Ratings

The methodology recommends that certain criteria be emphasized more heavily as reviewers were rolling up scores from the individual-criterion level to the Content and Depth ratings. In two cases, our reviewers modified these recommendations during implementation. For ELA/Literacy Depth, the recommendation was that B.1 and B.2 be most emphasized in determining the Depth rating; however, because of the issues mentioned above, reviewers did not emphasize the latter in determining the Depth rating. For Mathematics Depth, the methodology recommended that C.3 receive the predominant emphasis. However, as discussed in the Results section, reviewers did not believe the rating for C.3 should drive the Depth rating because of the way it was operationalized. Instead, C.4 and C.5 were equally emphasized.

Overall, we believe the methodology represents a rigorous and detailed approach to evaluating the quality of new assessments against the CCSS and other college and career readiness standards—and were pleased with how our panelists implemented it for this study. As first implementers, we also have a number of recommendations for improving the design and implementation of the methodology in future studies (Section III).

Findings

Findings are organized around three key questions:

1. Do the assessments place strong emphasis on the most important content for college and career readiness (CCR), as called for by the Common Core State Standards and other CCR standards? (**Content**)
2. Do they require all students to demonstrate the range of thinking skills, including higher-order skills, called for by those standards? (**Depth**)
3. What are the overall strengths and weaknesses of each assessment relative to the examined criteria for ELA/Literacy and mathematics? (**Overall Strengths and Weaknesses**)

Results for English language arts appear first, followed by math. For each criterion, we first delineate the requirements for assessments to earn a top score on the evaluation. In cases where it is relevant, we present the tentative scoring guidance proposed by NCIEA (e.g., percentages of items required to score a "2: Meets the criterion").[59] These are tentative guidelines that may be revised for future implementations of the methodology.[60]

We also provide additional detail in the form of illustrative quotes and fine-grained analyses to help contextualize the findings. Illustrative quotes are gleaned from reviewers' comments and responses to open-ended questions, which they were asked to complete after their reviews.[61] Although submitted by individual reviewers, we include

59. Depending on relevance, tentative scoring guidance appears in the report body or the Appendix.

60. The tentative cut-offs were provided to support reviewers in interpreting the more general language of the criteria—but because they were "tentative," reviewers could (and did) use their professional judgment as they interpreted and applied the criteria.

61. Reviewers were asked to provide feedback on: a) the greatest strength and weakness of the methodology; b) ways in which the methodology could be improved; c) the strengths and weaknesses of each program they evaluated; and d) suggestions for future improvements to each testing program.

only those quotes that represent consensus opinion. In some cases, we provide more fine-grained results than those that appear in the criterion results. (For instance, we present the average proportion of reading items requiring close reading.) Unless otherwise specified, all descriptive statistics are calculated by averaging form-level reviewer ratings across test forms and grades.[62]

English Language Arts/Literacy Content Criteria

The four test programs varied significantly in the degree to which they emphasize the most important content of the CCSS for the grade level. Summary ratings and panel statements are presented in Table 5.

TABLE 5

Overall ELA/Literacy Content Ratings

L

ACT Aspire

LIMITED/UNEVEN MATCH

The assessment program includes an emphasis on close reading and language skills.

However, the reading items fall short on requiring students to cite specific textual information in support of a conclusion, generalization, or inference and in requiring analysis of what has been read. In order to meet the criteria, ACT Aspire should strengthen assessing writing to sources, vocabulary, and research and inquiry.

L

MCAS

LIMITED/UNEVEN MATCH

The assessment requires students to read closely well-chosen texts and presents test questions of high technical quality.

However, the program would be strengthened by assessing writing annually, assessing the three types of writing called for across each grade band, requiring writing to sources, and placing greater emphasis on assessing research and language skills.

E

PARCC

EXCELLENT MATCH

The program demonstrates excellence in the assessment of close reading, vocabulary, writing to sources, and language, providing a high-quality measure of ELA/Literacy content, as reflected in college and career readiness standards.

The tests could be strengthened by the addition of research tasks that require students to use two or more sources and, as technologies allow, a listening and speaking component.

E

Smarter Balanced

EXCELLENT MATCH

The program demonstrates excellence in the areas of close reading, writing to sources, research, and language. The listening component represents an important step toward adequately measuring speaking and listening skills—a goal specifically reflected in the standards. Overall, Smarter Balanced is a high quality measure of the content required in ELA/Literacy and literacy, as reflected in college and career readiness standards.

A greater emphasis on Tier 2 vocabulary would further strengthen these assessments relative to the criteria.

Note: These ratings are based on five criteria (B.3, B.5, B.6, B.7, and B.8); the first two receive greater emphasis.

62. Proportions presented in the text and used in determining form-level ratings were typically based on percentages of score points associated with items. Calculating proportions based on item counts instead may have resulted in somewhat different ratings. This may be especially true for computer-adaptive tests such as Smarter Balanced, where the number of score points does not necessarily correspond to the actual weight in determining the final score. This note applies to all criteria for which score points were used.

Criterion B.3
Do the tests require students to read closely and use evidence from texts to obtain and defend responses?

This criterion reflects the high priority in the CCSS that students be able to read and understand increasingly complex texts, both literary and informational. The following were required to fully meet this criterion:

1. *Nearly all reading items require close reading and analysis of text, rather than skimming, recall, or simple recognition of paraphrased text.*
2. *Nearly all reading items focus on central ideas and important particulars.*
3. *Nearly all items are aligned to the specifics of the standards.*
4. *More than half of the reading score points are based on items that require direct use of textual evidence.*

As shown in Table 6, both PARCC and Smarter Balanced received the highest rating (Excellent Match) on this criterion. On average across forms and grades, 87 percent of PARCC reading items and 69 percent of Smarter Balanced reading items were scored as requiring direct textual evidence. The 2014 MCAS missed the Excellent rating and fell to Good Match largely due to an insufficient number of items that require direct citing of evidence from texts—just 29 percent. As one MCAS ELA/Literacy grade 8 reviewer explained, "While … students likely had to use the information in the texts in order to answer the questions, the items, as a whole, did not require students to provide direct evidence from the text."

TABLE 6

Reading Summary (Criterion B.3)

L

ACT Aspire

LIMITED/UNEVEN MATCH

Although most reading items require close reading of some kind, too many can be answered without analysis of what was read. Items that purport to require specific evidence from text often require only recall of information from text. To meet this criterion, the test items should require students to cite specific text information in support of some conclusion, generalization, or inference drawn from the text.

G

MCAS

GOOD MATCH

Most reading items require close reading and focus on central ideas and important particulars. Some questions, however, do not require the students to provide direct textual evidence to support their responses. In addition, too many items do not align closely to the specifics of the standards.

E

PARCC

EXCELLENT MATCH

Nearly all reading items require close reading, the understanding of central ideas and the use of direct textual evidence.

E

Smarter Balanced

EXCELLENT MATCH

Nearly all reading items align to the reading standards requiring close reading, the understanding of central ideas, and use of direct textual evidence in support of a conclusion, generalization, or inference.

The ACT Aspire test was rated as Limited/Uneven Match, falling short both in the percentage of items requiring direct textual evidence (51 percent) and in the percentage of items requiring analysis of text (75 percent, as compared to 89 percent or more for the other three programs). Multiple ELA/Literacy reviewers commented on ACT Aspire's lack of questions requiring students to use direct textual evidence or reference multiple texts.

Criterion B.5
Do the tests require students to write narrative, expository, and persuasive/argumentation essays (across each grade band, if not in each grade) in which they use evidence from sources to support their claims?

The following were required to fully meet this criterion:

1. *All three writing types are approximately equally represented across all forms in the grade band (K–5; 6–12), allowing blended types (i.e., writing types that blend two or more of narrative, expository, and persuasive/argumentation) to contribute to the distribution.*
2. *All writing prompts require writing to sources (meaning they are text-based).*

The grade bands identified in the CCSS are K–5 and 6–12. Because this study evaluated one grade from each band (grades 5 and 8), panelists also reviewed documentation for each program to check their writing requirements across each grade band. As shown in Table 7, PARCC and Smarter Balanced each received a rating of Excellent Match, as each writing type was assessed at least once across each grade band and all extended writing tasks required students to use sources. As one ELA/Literacy grade 8 reviewer commented for PARCC, "writing tasks were highly analytical and required close reading and thoughtful analysis."

TABLE 7

Writing Summary (Criterion B.5)

L

ACT Aspire

LIMITED/UNEVEN MATCH

Although the program documentation shows that a balance of all three writing types is required across each grade band, the writing prompts do not require writing to sources. As a result, the program insufficiently assesses the types of writing required by college and career readiness standards.

W

MCAS

WEAK MATCH

Writing is assessed at only one grade level per band, and there is insufficient opportunity to assess writing of multiple types. In addition, the writing assessments do not require students to use sources. As a result, the program inadequately assesses the types of writing required by college and career readiness standards.

E

PARCC

EXCELLENT MATCH

The assessment meets the writing criterion, which requires writing to sources. Program documentation shows that a balance of all three writing types is required across each grade band.

E

Smarter Balanced

EXCELLENT MATCH

The writing items are of high quality, and the writing prompts all require the use of textual evidence. Program documentation shows that a balance of all three writing types is required across each grade band.

ACT Aspire's writing test, available annually at grades 3–8 and early high school, does include the three types of writing called for by the criterion, but does not require students to write using evidence from sources, resulting in a rating of Limited/Uneven Match. Multiple ELA/Literacy reviewers commented on how infrequently students were asked to write to and compare multiple passages.

Given that MCAS does not evaluate writing at grades 5 and 8, it received a score of Limited/Uneven Match.

MCAS does, however, evaluate writing at grades 4 and 7, so we asked our panelists to review those writing items and evaluate them based on the associated CCSS grade-level standards. This exercise was conducted only to give the program information about its treatment of writing; it was not a part of the final scoring. Ultimately, both the writing prompts on the fourth- and seventh-grade assessments did not require writing to sources, which likely would have also resulted in a Weak Match had those prompts been a part of the actual review. As one ELA/Literacy grade 8 reviewer remarked, "the biggest flaw in showing college and career readiness is that writing is not assessed at every grade; therefore, there is very little evidence [relative] to students' abilities in the various writing modes."

Criterion B.6
Do the tests require students to demonstrate proficiency in the use of language, including academic vocabulary and language conventions, through tasks that mirror real-world activities?

This criterion calls for assessments to include vocabulary questions that focus on words important to the central ideas of a text and that require students to use context to determine meaning. In addition, the tests should focus on what are referred to as Tier 2 words. According to the standards, these are "general academic" words that are far more likely to appear in written text than in speech. They "appear in all sorts of texts: informational texts (words such as *relative*, *vary*, *formulate*, *specificity*, and *accumulate*), technical texts (*calibrate*, *itemize*, *periphery*), and literary texts (*misfortune*, *dignified*, *faltered*, *unabashedly*)."[63]

This same criterion also addresses language conventions. As opposed to asking a student to select the correctly punctuated sentence from a list of four or five options, this criterion calls for items, whether multiple-choice or technology-enhanced, that mirror real-world activities (i.e., place errors in context and require students to edit or revise for clarity or correctness).

The following were required to fully meet this criterion:

1. *The large majority of vocabulary items (i.e., three-quarters or more) focus on Tier 2 words and require the use of context, and more than half assess words important to central ideas.*
2. *A large majority (i.e., three-quarters or more) of the items in the language skills component and/or scored with a writing rubric (i.e., points in writing tasks that are allocated toward a language sub-score), mirror real-world activities, focus on common errors, and emphasize the conventions most important for readiness.*
3. *Vocabulary is reported as a sub-score or at least 13 percent of score points are devoted to assessing vocabulary/language.*
4. *Language skills are reported as a sub-score or at least 13 percent of score points are devoted to assessing language skills (language skills items plus score points).*

As shown in Table 8, PARCC excelled on this criterion, receiving high marks for both the vocabulary (85 percent of vocabulary items were on Tier 2 words and phrases) and the language items, leading to a score of Excellent Match. As one grade 8 reviewer commented, "[PARCC] items were generally great, especially the technology-enhanced vocabulary items." ACT Aspire and Smarter Balanced each received a rating of Good Match, and in each case the deficiency was in the number of vocabulary items that assess Tier 2 words (47 percent and 74 percent of

63. Common Core State Standards, "ELA/Literacy, Appendix A," http://www.corestandards.org/assets/Appendix_A.pdf, 33–36.

vocabulary items, respectively), whereas these assessments met the criteria for language and for the percent of score points devoted to language and vocabulary. MCAS's vocabulary items were generally adequate at grade 8 but insufficient at grade 5 (71 percent of vocabulary items included Tier 2 words and phrases, falling just below the 75 percent target), and the few language conventions items at grade 5 failed to require students to edit or revise for clarity or correctness, resulting in a rating of Limited/Uneven Match.

TABLE 8

Vocabulary and Language Skills Summary (Criterion B.6)

G

ACT Aspire

GOOD MATCH

Language items meet the criterion for being tested within writing activities, though more items are needed that are embedded in real world tasks such as editing. The vocabulary items do not meet the criterion because there are too few of them and not enough assess Tier 2 words.

L

MCAS

LIMITED/UNEVEN MATCH

Vocabulary items are sufficient and generally aligned to the criterion; however, the grade 5 items need more words at the Tier 2 level. Furthermore, a lack of program documentation means that the quality of vocabulary assessments cannot be substantiated across forms. MCAS does not meet the criterion for assessing language skills, which call for them to be assessed within writing assessments that mirror real-world activities including editing and revision.

E

PARCC

EXCELLENT MATCH

The test contains an adequate number of high-quality items for both language use and Tier 2 vocabulary and awards sufficient score points, according to the program's documentation, to both of these areas.

G

Smarter Balanced

GOOD MATCH

Language skill items are contained in a sub-score and meet the criterion for being assessed within writing and mirroring real-world activities such as editing and revision. The number of items that test vocabulary is a bit low; further, items coded as vocabulary too often did not test Tier 2 vocabulary words.

Criterion B.7
Do the tests require students to demonstrate research skills, including the ability to analyze, synthesize organize, and use information from sources?

This criterion presents a new challenge to test developers in that it requires development of tasks that contain multiple sources and assess whether students have accurately gleaned, analyzed, and synthesized evidence from those sources in their response. The criterion also requires that the tasks mirror real-world research activities, which is difficult when such activities typically involve online searches conducted over days, weeks, or months.

The following was required to fully meet this criterion:

1 *Three-quarters or more of the research items on each test form require analysis, synthesis, and/or organization of information.*

PARCC and Smarter Balanced included research tasks within their performance tasks. (The latter are extended tasks with multiple questions that ask students to analyze and use evidence from at least two sources, which may include audio, video, and text.) Both programs earned a rating of Excellent Match (Table 9).

ACT Aspire included a single item per test form that required analysis and organization of information, but the panels found this to be inadequate to earn more than a rating of Limited/Uneven Match. The 2014 MCAS did not contain any items that assessed research skills, resulting in a rating of Weak Match.

Another CCSSO criterion is to be evaluated "over time, as assessment advances allow." Criterion B.8 reads, "Do the tests measure students' speaking and listening communication skills?" Only one program, Smarter Balanced, has incorporated listening items, which the panel commended; none of the programs assess speaking skills at this time. Because the criterion calls for speaking and listening to be assessed over time, ratings for this criterion were not included in the overall ELA/Literacy content ratings.

TABLE 9

Research and Inquiry Summary (Criterion B.7)

L

ACT Aspire

LIMITED/UNEVEN MATCH

Although the one item at each grade level involving research and inquiry did indeed require analysis and organization of information, this single item is insufficient to provide a quality measure of research and inquiry.

W

MCAS

WEAK MATCH

The assessment has no test questions devoted to research.

E

PARCC

EXCELLENT MATCH

The research items require analysis, synthesis, and/or organization, as well as the use of multiple sources, therefore meeting the criterion for Excellent.

E

Smarter Balanced

EXCELLENT MATCH

The research items require analysis, synthesis, and/or organization, as well as the use of multiple sources, therefore meeting the criterion for Excellent.

English Language Arts/Literacy Depth Criteria

All four programs fared very well on Depth, which required students to demonstrate the higher-order thinking skills called for by the CCSS, via high-quality items that reflect a variety of item types (Table 10).

PARCC received the highest rating of Excellent Match and the remaining three programs received ratings of Good Match.

TABLE 10

Overall ELA/Literacy Depth Ratings

G

ACT Aspire

GOOD MATCH

The program's assessments are built on high-quality test items and texts that are suitably complex.

To fully meet the CCSSO Criteria, more cognitively demanding test items are needed at both grade levels, as well as additional literary narrative text, as opposed to literary informational texts.

G

MCAS

GOOD MATCH

The assessments do an excellent job in presenting a range of complex reading texts.

To fully meet the demands of the CCSSO Criteria, however, the test needs more items at higher levels of cognitive demand, a greater variety of items to test writing to sources and research, and more informational texts, particularly those of an expository nature.

E

PARCC

EXCELLENT MATCH

The PARCC assessments meet or exceed the depth and complexity required by the criteria through a variety of item types that are generally high quality.

A better balance between literary and informational texts would further strengthen the assessments in addressing the criteria.

G

Smarter Balanced

GOOD MATCH

The assessments use a variety of item types to assess student reading and writing to sources.

The program could better meet the depth criteria by increasing the cognitive demands of the grade 5 assessment and ensuring that all items meet high editorial and technical quality standards.

Note: These ratings are based on four criteria (B.1, B.2, B.4, B.9); the first two receive greater emphasis.

Criterion B.1
Do the tests require a balance of high-quality literary and informational texts?

The CCSS place great emphasis on students being able to read literary texts (e.g., classic novels), informational texts (e.g., newspapers, scientific and historical texts), as well as technical documents, in order to build their content knowledge.

The following were required to fully meet this criterion:

1. *Approximately half of the texts at grades 3–8 and two-thirds at high school are informational, and the remainder literary.*
2. *Nearly all passages are high quality (previously published or of publishable quality).*
3. *Nearly all informational passages are expository in structure.*
4. *For grades 6–12, the informational texts are split nearly evenly for literary nonfiction, history/social science, and science/technical.*

The texts used across all four programs were found to be of high quality (previously published or of publishable quality). Smarter Balanced tests contained both the requisite balance and text quality, earning a rating of Excellent Match. However, ACT Aspire, MCAS, and PARCC were all found to have a slight imbalance across literary and informational texts, resulting in a rating of Good Match (Table 11).[64]

TABLE 11

Text Quality and Types Summary (Criterion B.1)

G

ACT Aspire

GOOD MATCH

The texts are of high quality, and the proportion of informational texts meets the criterion.

The assessment would better align to the criterion, however, with additional literary narrative text, as opposed to literary informational text.

G

MCAS

GOOD MATCH

The quality of the texts is very high.

Regarding the balance of text types, some forms had too few informational texts.

G

PARCC

GOOD MATCH

Although the passages are consistently of high quality, the tests would have better reflected the criterion with additional literary nonfiction passages.

E

Smarter Balanced

EXCELLENT MATCH

Overall text quality is high, and among informational texts there is a high proportion of expository text types.

64. ACT Aspire does not classify literary nonfiction texts that are primarily narrative in structure as "informational." See Appendix G for more on ACT Aspire's interpretation of CCSSO criteron B.1.

Criterion B.2
Do the tests require appropriate levels of text complexity, increasing the level each year so that students are ready for the demands of college and career by the end of high school?

As explained in the methodology, the intent of this criterion was to gather quantitative and qualitative data from the testing programs regarding text complexity to determine whether reading passages had been assigned to grade bands and grade levels appropriately. Unfortunately, these data could not be used (see Section I, *Methodology Modifications*), so the evaluation of this criterion was based solely on the requirements recorded within each program's test documentation and specifications.

The following were required to fully meet this criterion:

1. *Documentation clearly explains how quantitative data are used to determine grade band placement.*
2. *Texts are placed at the grade level recommended by the qualitative review.*

In every case, the documentation met the criterion (Table 12). However, because the documentation is not a guarantee of what will appear on actual test forms, the panel decided on a maximum rating of Good Match. All four programs received this rating.

TABLE 12

Complexity of Texts Summary (Criterion B.2)

G

ACT Aspire

GOOD MATCH

It is based solely on the review of program documentation, which is determined to have met the criterion. The test blueprints and other documents clearly and explicitly require texts to increase in complexity grade-by-grade and for texts to be placed in grade bands and grade levels based on appropriate quantitative and qualitative data.

G

MCAS

GOOD MATCH

It is based solely on the review of program documentation, which is determined to have met the criterion. The test blueprints and other documents clearly and explicitly require texts to increase in complexity grade-by-grade and for texts to be placed in grade bands and grade levels based on appropriate quantitative and qualitative data.

G

PARCC

GOOD MATCH

It is based solely on the review of program documentation, which is determined to have met the criterion. The test blueprints and other documents clearly and explicitly require texts to increase in complexity grade-by-grade and for texts to be placed in grade bands and grade levels based on appropriate quantitative and qualitative data.

G

Smarter Balanced

GOOD MATCH

It is based solely on the review of program documentation, which is determined to have met the criterion. The test blueprints and other documents clearly and explicitly require texts to increase in complexity grade-by-grade and for texts to be placed in grade bands and grade levels based on appropriate quantitative and qualitative data.

Criterion B.4
Are all students required to demonstrate a range of high order, analytical thinking skills in reading and writing based on the depth and complexity of the standards?

Research over the last decade has shown that existing state tests were not adequately testing—and therefore not requiring the teaching of—higher-order skills such as analysis, synthesis, the development of a logical argument, and use of concepts to solve non-routine problems.[65] The CCSSO Criteria recognize these shortcomings of prior tests and recommend that "all students demonstrate a range of higher-order, analytical thinking skills in reading and writing based on the depth and complexity of college and career ready standards...." Webb's Depth of Knowledge (DOK) taxonomy[66] was used in this study to classify thinking skills (see Appendix A). In this taxonomy, level 1 is the lowest level (recall), level 2 requires use of a skill or concept, and levels 3 and 4 are considered the higher-order thinking skills.

To receive the highest rating on this criterion, the distribution of cognitive demand on test forms had to match the distribution of cognitive demand of the standards as a whole and match the higher cognitive demand (DOK 3+) of the standards. Note that criterion B.4 is not a rating of test difficulty. Assessments that do not match the distribution of complexity of the standards, including if they have too many high DOK items, may receive a rating of less than Excellent Match.

TABLE 13

Matching the Complexity of the Standards (Criterion B.4)

W

ACT Aspire

WEAK MATCH

To better reflect the depth and complexity of the standards, both grade-level tests should require more items with higher cognitive demands, although this problem is greater at grade 8.

L

MCAS

LIMITED/UNEVEN MATCH

More items that measure the higher levels of cognitive demand are needed to sufficiently assess the depth and complexity of the standards.

E

PARCC

EXCELLENT MATCH

The test is challenging overall; indeed, the cognitive demand of the grade 8 test exceeds that of the CCSS.

G

Smarter Balanced

GOOD MATCH

The cognitive demand of items cover a sufficient range and, in grade 8, the percentage of more demanding items (DOK 3 and 4) correspond well to the demand of the standards. However, the grade 5 test needs more items at higher levels of cognitive demand to reflect fully the depth and complexity of the standards.

65. Yuan and Le, 2012.

66. We chose the Webb DOK approach because it is widely used and familiar; however, results might have been different if another approach to cognitive complexity had been used. Future iterations of the methodology could use newer or more innovative approaches to cognitive complexity if desired.

To determine whether the tests reflected the depth and complexity of the standards, we first needed to measure those aspects of the CCSS. As described in Appendix A, the study team contracted with content experts in advance of the study to determine the distribution of the cognitive demand of the grade 5 and grade 8 CCSS standards.

The results show a great deal of variation across programs in their emphasis on higher-level skills and match to the standards. PARCC scored an Excellent Match, as the DOK of PARCC at both grade levels exceeded that of the standards.[67] Smarter Balanced scored Good Match for meeting the DOK of the standards at grade 8 but underemphasizing DOK 3–4 at grade 5. MCAS scored Limited/Uneven Match for underemphasizing DOK 3–4 by roughly 10–20 percent at each grade. ACT Aspire scored Weak Match for underemphasizing DOK 3+ by roughly 30 percent at grade 8.

The results of the DOK analysis for the CCSS and the four programs are shown in Table 14. Interestingly, at each grade, 46 percent of standards content require the use of higher-order skills (level 3 or 4).

More detail on this analysis, including the distributions of DOK on other national and international tests, appears in Appendix A.

TABLE 14

The Distribution of Cognitive Demand in ELA/Literacy: The CCSS vs. Tests

		Grade 5			Grade 8		
		LEVEL 1	LEVEL 2	LEVELS 3 & 4	LEVEL 1	LEVEL 2	LEVELS 3 & 4
CCSS		18%	37%	46%	10%	44%	46%
ACT Aspire	W	34%	38%	28%	46%	36%	18%
MCAS	L	10%	63%	27%	5%	59%	37%
PARCC	E	5%	45%	50%	2%	29%	69%
Smarter Balanced	G	19%	59%	22%	15%	41%	44%

LEGEND E Excellent Match G Good Match L Limited/Uneven Match W Weak Match

Note: Percentages in the table represent percentages of score points at each DOK level. Results for a particular grade and program were generated by averaging across all raters and forms for that grade and program (e.g., averaging the four raters of ACT Aspire's form 1 and the four raters of ACT Aspire's form 2 at grade 5).

67. The review focused primarily on match to the DOK of the standards, but reviewers also used professional judgment in determining final ratings.

Criterion B.9
Are a variety of item types used, including at least one that requires students to generate, rather than select, a response? Are the test items of high quality?

One aspect of item quality that the CCSSO Criteria emphasize is a diversity of item types. Specifically, the criteria recognize the widespread dissatisfaction with the ability of multiple-choice-only tests to fully measure complex student skills.

The CCSSO Criteria also address the editorial and technical quality of items. On all standardized assessments of student learning, it is imperative that items be of high editorial and technical quality and accuracy. The tests evaluated in this study are used for accountability purposes, so item quality must be above reproach. Each of these programs requires multiple rounds of item review and field-testing; even so, items with quality issues can sometimes make their way into the tests.

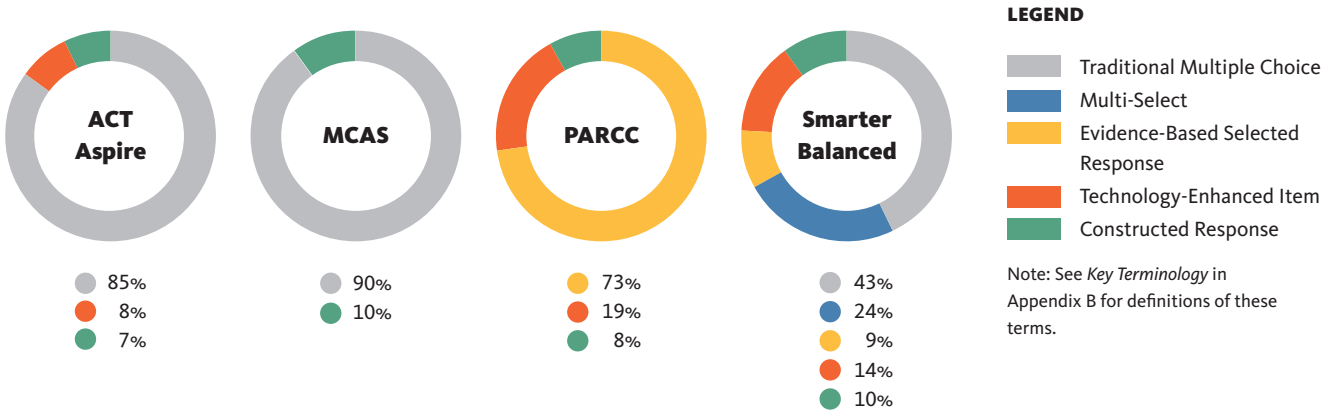
The following were required to fully meet this criterion:

1. *At least two item formats are used, including one that requires students to generate, rather than select, a response.*
2. *All or nearly all operational items reviewed reflect high editorial and technical quality and accuracy.*

All four programs use multiple item types, including at least one student-constructed response type, so they all met this portion of the criterion. The actual variety of item types, however, differed much more significantly, as shown in Table 15. Reviewers commended the innovative and appropriate use of technologies within Smarter Balanced and PARCC items, such as audio files in listening items and text editing. The ACT Aspire assessments, while also computer-based, had a much more limited set of item types, with heavy reliance on traditional multiple-choice items. The 2014 MCAS is a paper-and-pencil test that has a near-complete focus on traditional multiple-choice items, cited by multiple reviewers as a serious limitation.

TABLE 15

Distribution of Item Types in the ELA/Literacy Tests



Note: See *Key Terminology* in Appendix B for definitions of these terms.

The results for item quality are very strong (Table 16). Three programs—ACT Aspire, MCAS and PARCC—were rated as Excellent Match. These programs were each seen as having exceptionally high editorial accuracy and technical quality (less than 5 percent of items with quality issues). On Smarter Balanced items, some items (approximately one item per form) contained quality issues, ranging from editorial errors to readability or technical quality concerns, resulting in a score of Good Match.[68] For instance, reviewers sometimes felt that multiple answers could be considered correct even on items that were scored as having just one correct answer; identified items in which students had to define vocabulary from context as having insufficient context to determine the definition; and noted minor typographical issues, such as punctuation and spelling.[69]

TABLE 16

High-Quality Items and Variety of Item Types (Criterion B.9)

E

ACT Aspire

EXCELLENT MATCH

The test includes items that exhibit high technical quality and editorial accuracy. Multiple item formats are used, including student-constructed responses.

E

MCAS

EXCELLENT MATCH

Multiple item formats are used, including student-generated response items. The items exhibit high technical quality and editorial accuracy. The paper-and-pencil format precludes the use of technology-enhanced items, but the criterion for multiple item types is met.

E

PARCC

EXCELLENT MATCH

The tests use multiple item formats, including student-constructed responses.

G

Smarter Balanced

GOOD MATCH

The tests use multiple formats and technology-enhanced items including constructed responses. However, editorial or technical issues, including readability, were noted in a number of items.

68. The nature and timing of this review required Smarter Balanced to make the test items and forms available to reviewers through an alternate test interface that was more limited than the actual student interface used for the summative assessments, particularly with regard to how items appeared on the screen and how erroneous responses were handled. Though reviewers were not able to determine the extent to which these interface limitations impacted their findings, the study team worked with Smarter Balanced to ascertain which item issues were caused by interface differences and which were not. All item-relevant statements in the report reflect data not prone to interface differences.

69. Some Smarter Balanced "multi-select" items have more than one answer by design.

Mathematics Content Criteria

The mathematics review panel found that the four test programs varied somewhat in the degree to which they emphasize the most important content of the CCSS at each grade level (Table 17). PARCC and Smarter Balanced earned ratings of Good Match to the CCSSO Criteria. The ACT Aspire tests and the 2014 MCAS received ratings of Limited/Uneven Match.

TABLE 17

Overall Mathematics Content Ratings

L

ACT Aspire

LIMITED/UNEVEN MATCH

The program does not focus exclusively on the major work of the grade, but rather, by design, assesses material from previous and later grade(s). This results in a weaker match to the criteria.

The tests could better meet the criteria at both grades 5 and 8 by increasing the number of items that assess the major work.

L

MCAS

LIMITED/UNEVEN MATCH

While the grade 8 assessment focuses strongly on the major work of the grade, the grade 5 assessment does not, as it samples more broadly from the full range of standards for the grade.

The tests could better meet the criteria through increased focus on the major work of the grade on the grade 5 test.

G

PARCC

GOOD MATCH

The test could better meet the criteria by increasing the focus on the major work at grade 5.

G

Smarter Balanced

GOOD MATCH

The tests could better meet the criteria by increasing the focus on the major work in grade 5.

Note: Because of methodological challenges pertaining to criterion C.2 (see Section III), the Content ratings for mathematics are based primarily on the ratings for C.1.

Criterion C.1
Do the tests focus strongly on the content most needed for success in later mathematics?

Certain CCSS standards (termed "major work of the grade") have been deemed most important for ensuring that students are on track toward readiness for college or career. This criterion reflects that expectation. To score well on it, assessments must devote adequate attention toward these major work standards and also assess the full breadth of the major work (i.e., must assess, with at least one item, a large proportion of the major work "clusters").

The following was required to fully meet this criterion:

1. *The vast majority (i.e., at least three-quarters at elementary grades, at least two-thirds in middle school grades, and at least half in high school) of score points in each assessment focuses on the content that is most important for students to master in that grade in order to reach college and career readiness (also called the major work of the grade), and at least 90 percent of the major work clusters must be assessed by at least one item.*

As shown in Table 18, both PARCC and Smarter Balanced received a rating of Good Match on this criterion, falling short of the top rating because the grade 5 assessment was insufficiently focused (less than 75 percent of score points for each program) on the major work. The MCAS was rated Limited/Uneven Match, again due to insufficient focus in fifth grade. Only about 53 percent of MCAS score points focused on the major work, far below the 75 percent threshold, and also lower than PARCC (that just missed it at 72 percent) and Smarter Balanced (66 percent).

The ACT Aspire received the lowest score of Weak Match on this criterion because both grades' tests were far below the major work tentative cutoff (32 percent of score points on major work for grade 5 and 44 percent for grade 8, as compared to a cutoff of 75 percent). Though reviewers recognized that ACT Aspire's design intentionally included lower grade items, even those lower grade items were not concentrated on the major work of the respective grade. Reviewers for both grades 5 and 8 commented that ACT Aspire's test questions did not focus adequately on the major work of the grade.

TABLE 18

Focus Summary (Criterion C.1)

W

ACT Aspire

WEAK MATCH

ACT Aspire forms do not consistently place sufficient emphasis on the major work of the given grade, due in part to intentional test design, which requires inclusion of selected content from earlier and later grades. Still, many of the items coded to standards from lower grades do not address the major work of the relevant grade.

L

MCAS

LIMITED/UNEVEN MATCH

The grade 8 assessment is focused on the major work of the grade. The grade 5 assessment is significantly less focused on the major work of the grade than called for by the criterion, as it samples content across the full set of standards for the grade.

G

PARCC

GOOD MATCH

While the grade 8 tests focus strongly on the major work of the grade, the grade 5 tests fall short of the threshold required for the top rating.

G

Smarter Balanced

GOOD MATCH

While the grade 8 tests focus strongly on the major work of the grade, the grade 5 tests fall short of the threshold required for the top rating.

Criterion C.2
Do the tests assess a balance of concepts, skills, and applications?

The CCSS require students to demonstrate conceptual understanding and procedural skill/fluency, and apply this knowledge.

The following were required to fully meet this criterion:

1 *On each test form, at least 25 percent and no more than 50 percent of score points are allocated to each of the three categories: mathematical concepts, procedures/fluency, and applications.*

As described in Section III, there were a number of methodological challenges in implementing this criterion.[70] As a consequence, our reviewers agreed to provide only qualitative statements, rather than the ratings awarded to the other criteria. This criterion, therefore, was not used in the determination of the overall Content rating.

In general, the test forms from all four programs showed attention to conceptual understanding, procedural skill, and application. However, each program fell short of the goal of balance (which was operationalized as an even distribution) in one way or another. For ACT Aspire at both grades, reviewers noted that items directly assessing procedural skill were underrepresented. For MCAS at grade 5, reviewers found few items assessing conceptual understanding and an overabundance of application items. The grade 5 PARCC exam similarly had an overabundance of application items, some of which reviewers noted had shallow contexts. Finally, the Smarter Balanced exams at both grade levels had a slight wealth of application items, and reviewers also noticed that some forms were more heavily focused on applications than others.

70. To wit: All four programs require, in their program documentation, the assessment of conceptual understanding, procedural skill/ fluency, and application, although most do not clearly distinguish between procedural skill/fluency and conceptual understanding. Also, specific balance across these three types is not required. Due to variation across reviewers in how this criterion was understood and implemented, final ratings could not be determined with confidence.

Mathematics Depth Criteria

All four programs fared well on Depth in mathematics (Table 19), requiring all students to demonstrate the higher-order thinking skills called for by the CCSS. MCAS received a rating of Excellent Match, and the other three programs received a rating of Good Match.

TABLE 19

Overall Mathematics Depth Ratings

G

ACT Aspire

GOOD MATCH

The items are well crafted and clear, with only rare instances of minor editorial issues.

The ACT Aspire tests include proportionately more items at high levels of cognitive demand (DOK 3) than the standards reflect, and proportionately fewer at the lowest level (DOK 1). This finding is both a strength, in terms of promoting strong skills, and a weakness, in terms of ensuring adequate assessment of the full range of cognitive demand within the standards.

While technically meeting the criterion for use of multiple item types, the range is nonetheless limited, with the large majority comprising multiple-choice items.

The program would better meet the criteria for Depth by including a wider variety of item types and relying less on traditional multiple-choice items.

E

MCAS

EXCELLENT MATCH

The assessment uses high-quality items and a variety of item types. The range of cognitive demand reflects that of the standards of the grade. While the program does not code test items to math practices, mathematical practices are nonetheless incorporated within items.

The program might consider coding items to the mathematical practices and making explicit the connections between specific practices and specific content standards.

G

PARCC

GOOD MATCH

The tests include items with a range of cognitive demand, but at grade 8 that distribution contains a higher percentage of items at the higher levels (DOK 2 and 3) and significantly fewer items at the lowest level (DOK 1). This finding is both a strength, in terms of promoting strong skills, and a weakness, in terms of ensuring adequate assessment of the full range of cognitive demand within the standards.

The tests include a variety of item types that are largely of high quality. However, a range of problems (from minor to severe) surfaced relative to editorial accuracy and, to a lesser degree, technical quality.

The program could better meet the Depth criteria by ensuring that all items meet high editorial and technical standards and by ensuring that the distribution of cognitive demand on the assessments provides sufficient information across the range.

G

Smarter Balanced

GOOD MATCH

The exam includes a range of cognitive demand that fairly represents the standards at each grade level.

The tests have a strong variety of item types, including those that make effective use of technology. However, a range of problems (from minor to severe) surfaced relative to editorial accuracy and, to a lesser degree, technical quality. A wide variety of item types appear on each form, and important skills are assessed with multiple items, as is sound practice. Yet, individual forms sometimes contained two or three items measuring the same skill that were nearly identical, with only the numerical values changed in the item stem and a different set of answer choices. Such near-duplication may not impact the accuracy of the score, but a greater variety of question stems/scenarios is desirable.

The program could better meet the Depth criteria by ensuring that all items meet high editorial and technical standards and that a given student is not presented with two or more virtually identical problems.

Note: The mathematics Depth ratings are an aggregation of ratings from three criteria (C.3–C.5). The methodology recommended that C.3 receive greater emphasis but reviewers chose instead to consider equally C.4 and C.5.

Criterion C.3
Do the tests connect mathematical practices to content?

The CCSS have a focus on mathematics practices through their Standards for Mathematical Practice (SMPs). The SMPs describe a variety of types of mathematical expertise that students are expected to develop, such as reasoning, modeling, and attending to precision.

The following was required to fully meet this criterion:

1 *All or nearly all items that assess mathematical practices also align to one or more content standards.*

All four of the programs earned a score of Excellent Match on this criterion (Table 20). All ACT Aspire, PARCC, and Smarter Balanced items that had been coded in item metadata to SMPs also assessed at least one content standard. Since reviewers were not verifying the SMP coding provided by the programs—just checking for alignment to standards—these three programs met this criterion with 100 percent of items. MCAS did not code items to mathematical practices, but reviewers agreed that items assessed these practices nonetheless, and it also earned a score of Excellent Match.

TABLE 20

Connecting Practice to Content (Criterion C.3)

E

ACT Aspire

EXCELLENT MATCH

All items that are coded to mathematics practices are also coded to one or more content standard.

E

MCAS

EXCELLENT MATCH

Although no items are coded to mathematical practices, the practices were nonetheless assessed within items that also assessed content.

E

PARCC

EXCELLENT MATCH

All items that are coded to mathematics practices are also coded to one or more content standard.

E

Smarter Balanced

EXCELLENT MATCH

All items that are coded to mathematics practices are also coded to one or more content standard.

Criterion C.4
Are all students required to demonstrate a range of high-order, analytical thinking skills in mathematics based on the depth and complexity of the standards?

The approach to studying DOK in mathematics was analogous to ELA/Literacy. Webb's Depth of Knowledge (DOK) taxonomy was again used: level 1 is the lowest level (recall), level 2 requires use of a skill or concept, and levels 3 and 4 are considered the higher-order thinking skills. As in ELA/Literacy, the study team contracted with content experts in advance of the study to determine the distribution of the cognitive demand of the grade 5 and grade 8 standards. The experts coded 7 percent to 9 percent of CCSS content as being on DOK levels 3 or 4 in mathematics, depending on the grade (see Table 22).

To receive the highest rating on this criterion, the distribution of cognitive demand on test forms had to match the distribution of cognitive demand of the standards as a whole and match the higher cognitive demand (DOK 3+) of the standards. (As was the case in the ELA/Literacy review of cognitive demand, this is not a rating of test difficulty.) Assessments that do not match the distribution of complexity of the standards, including if they have too many high DOK items, may receive a rating of less than Excellent Match.

As shown in Table 21, the highest score for this criterion was awarded to MCAS. At both grades, the distribution of cognitive demand in the standards closely mirrored that of the CCSS, which was the goal of this criterion.

TABLE 21

Matching the Complexity of the Standards (Criterion C.4)

L

ACT Aspire

LIMITED/UNEVEN MATCH

At both grades 5 and 8, the test forms include significantly more items of high cognitive demand (DOK 3) than reflected in the standards, and proportionately fewer at the lowest level (DOK 1). While these items increase the challenge of the tests, standards that call for the lowest level of cognitive demand (DOK 1) may be under-assessed.

E

MCAS

EXCELLENT MATCH

At each grade level, the distribution of cognitive demand closely reflects that of the standards.

G

PARCC

GOOD MATCH

The distribution of cognitive demand of items reflects that of the standards very well at grade 5, while the grade 8 test includes proportionately more items at the higher levels of cognitive demand (DOK 2 and 3). As a result, grade 8 standards that call for the lowest level of cognitive demand may be under-assessed.

G

Smarter Balanced

GOOD MATCH

The distribution of cognitive demand of items reflects that of the standards very well at grade 5. At grade 8, the test includes proportionately fewer items at the lowest levels of cognitive demand (DOK 1) than in the standards, and proportionately more items at the mid-level of cognitive demand (DOK 2). As a result, grade 8 standards that call for the lowest level of cognitive demand may be under-assessed.

PARCC and Smarter Balanced both earned a score of Good Match for similar reasons—while their grade 5 assessments were fairly closely aligned to the DOK of the standards, their grade 8 assessments overemphasized DOK 2 and underemphasized DOK 1 relative to the standards (just 13 percent to 16 percent at DOK 1 versus 51 percent in the standards).[71] (See Table 22 for each program's distribution of cognitive demand, as compared to the CCSS.) ACT Aspire earned the lowest score of Limited/Uneven Match for this criterion. The ACT Aspire exam was seen as too heavily concentrated at the higher DOK levels relative to the standards (e.g., 35 percent at DOK 3 or 4 in grade 8, versus 9 percent CCSS) and thereby under-assessing the lower-level skills. Reviewers remarked that the test is "tipped heavily" toward higher DOK.

TABLE 22

The Distribution of Cognitive Demand in Mathematics: CCSS vs. Tests

		Grade 5			Grade 8		
		LEVEL 1	LEVEL 2	LEVELS 3 & 4	LEVEL 1	LEVEL 2	LEVELS 3 & 4
CCSS		43%	50%	7%	51%	40%	9%
ACT Aspire	L	23%	40%	37%	20%	45%	35%
MCAS	E	40%	58%	2%	40%	46%	14%
PARCC	G	34%	55%	11%	13%	62%	24%
Smarter Balanced	G	46%	36%	18%	16%	75%	9%

LEGEND E Excellent Match G Good Match L Limited/Uneven Match W Weak Match

Note: Percentages in the table represent percentages of score points at each DOK level. Results for a particular grade and program were created by averaging across all raters and forms for that grade and program (e.g., averaging the four raters of ACT Aspire's form 1 and the four raters of ACT Aspire's form 2 at grade 5).

Criterion C.5
Are a variety of item types used, including at least one that requires students to generate, rather than select, a response? Are the test items of high quality?

One aspect of item quality that the CCSSO Criteria emphasize is a diversity of item types. Specifically, the criteria recognize the widespread dissatisfaction with the ability of multiple-choice-only tests to fully measure complex student skills.

The CCSSO Criteria also address the editorial and technical quality of items. The tests evaluated in this study are used for accountability purposes, so item quality must be above reproach. Each of these programs requires multiple rounds of item review and field-testing, but issues with quality can nonetheless make their way in.

The following were required to fully meet this criterion:

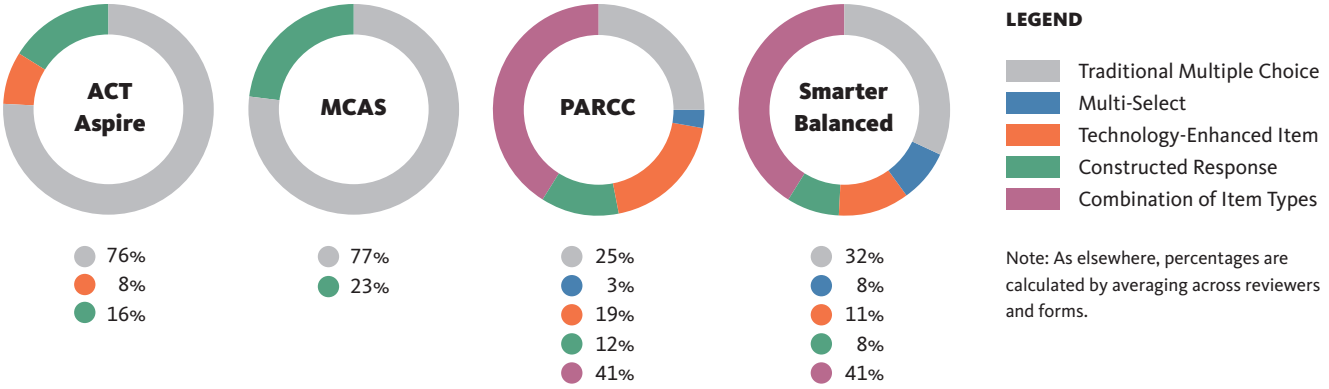
1. *At least two item formats are used, including one that requires students to generate, rather than select, a response.*
2. *All or nearly all operational items reviewed reflect both high technical quality and high editorial accuracy.*

71. PARCC uses a partial credit scoring model for multi-point items, such that students can sometimes earn partial credit for demonstrating DOK 1 or 2 skills on an item that might require DOK 3 for the maximum score. Since reviewers did not use scoring guides in their analysis of item DOK, PARCC's emphasis on higher-level skills might be overstated.

All four programs use multiple item types, including at least one student-constructed response type, so they all met this portion of the criterion. The actual variety of item types, however, differed much more significantly, as shown in Table 23. PARCC and Smarter Balanced typically had a wide range of item types, including multiple choice, constructed response, multi-select, and technology-enhanced. Both of these assessments had less than 50 percent traditional multiple-choice items. In contrast, both ACT Aspire and MCAS had 75 percent or more of items as traditional multiple-choice items.

TABLE 23

Distribution of Item Types in Mathematics Tests



Note: As elsewhere, percentages are calculated by averaging across reviewers and forms.

As shown in Table 24, the results for item quality are more varied. MCAS and ACT Aspire items were rated very highly in terms of quality, with only minor editorial issues. (Though reviewers did note that some ACT Aspire grade 8 items were susceptible to simplification via calculator use.) Thus, both of these assessments earned a score of Excellent Match on C.5.

Reviewers noted more issues related to item quality on the PARCC and Smarter Balanced forms than on the ACT Aspire or MCAS forms. Although this concern applies to a small percentage of items, the review panels expressed the need that a very high bar be set on the quality of items used on consequential tests. Indeed, most of the problems noted on PARCC and Smarter Balanced forms were editorial; some concerned the layout of the item on the screen,[72] but others were mathematical. Reviewers of both PARCC and Smarter Balanced forms noted that the technology enhancements did not always improve item quality.

Reviewers found quality issues more frequently on the PARCC forms than on the Smarter Balanced forms, albeit those instances were rare (one reviewer noted "very few 'broken' problems or major issues" on PARCC items). Still, the panel believed that these issues were significant enough to potentially impact the accuracy of the score. Thus, PARCC earned a score of Good Match on C.5.

Although less frequent, some quality concerns on Smarter Balanced items were viewed as more serious by the panel, at times interfering with the assessed construct. Some reviewers noted mathematical errors or imprecision in certain items (approximately one item per form).[73] The most common of these issues pertained to excessive precision in numbers used in application problems (e.g., more digits after the decimal point than would be appropriate given the problem context) and a lack of precise language in stem wording that made multiple

72. For example, there was one instance in which a mathematical formula appeared "broken" across lines on the screen.

73. See footnote 68 for more on Smarter Balanced's test interface.

answers plausible on items where there was only one right answer allowed. Several reviewers also noted that some Smarter Balanced forms had multiple items measuring the same skill that were nearly identical, with only the numbers in the item stem and the set of answer choices changed.[74] In light of these issues, Smarter Balanced earned a score of Limited/Uneven Match for C.5.

TABLE 24

High-Quality Items and Variety of Item Types (Criterion C.5)

E

ACT Aspire

EXCELLENT MATCH

The program uses multiple item types, including constructed response on the Extended Task items. These items, although they carry high point values, are limited in number; the rest of the items are predominantly multiple-choice.

The large majority of items are of high technical and editorial quality, with only very minor issues of editing, language, or accuracy. At the grade 8 level, some items appear to be susceptible to simplification by use of calculators, which are allowed on all items at grade 8, in contrast to the other programs that allow them on a restricted set of items.

E

MCAS

EXCELLENT MATCH

Both grade 5 and grade 8 forms include multiple item types, including constructed-response. The items are of high technical and editorial quality, with very minor issues of editing, language, and accuracy at grade 8.

G

PARCC

GOOD MATCH

The program includes a wide variety of item types, including several that require student-constructed responses. However, there are a number of items with quality issues, mostly minor editorial but sometimes mathematical.

L

Smarter Balanced

LIMITED/UNEVEN MATCH

The program includes a wide variety of item types, many of which make effective use of technology.

The program could be improved by ensuring that virtually identical items are not presented to individual students. Further, a good deal of variability across forms and grades is observed, with some forms fully meeting the item quality criterion and others only partially meeting it. Issues exist with the editorial quality and mathematical accuracy of individual items, most of which are minor but some of which could impact assessment of the targeted skill, resulting in a rating of Limited/Uneven.

74. Smarter Balanced staff indicate that the use of near-duplicate items is intentional in the assessment of mathematical fluency in Smarter Balanced assessments.

Criterion Level Ratings

Tables 25A and 25B show the final tally of the ELA/Literacy and Math criteria ratings. Immediately striking in ELA/Literacy is that the two consortia assessments earned twice as many ratings of Good and Excellent Match as the other two programs, earning eight high ratings to the four for ACT Aspire and MCAS. PARCC earned the most Excellent Match ratings (six), while Smarter Balanced was the only assessment with no ratings of Weak Match (partly because it was also the only program to test listening on the summative assessment).

The ratings for mathematics (Table 25B) were more similar between programs, with PARCC earning four Excellent or Good Match ratings, Smarter Balanced and MCAS three, and ACT Aspire two. MCAS scored particularly well on the three Depth criteria in mathematics, while PARCC is the only assessment that earned all Good Match or better scores.

TABLE 25A

ELA/Literacy Ratings Tally by Program

ACT Aspire	E	G	G	G	L	L	L	W	W
MCAS	E	G	G	G	L	L	W	W	W
PARCC	E	E	E	E	E	E	G	G	W
Smarter Balanced	E	E	E	E	G	G	G	G	L

TABLE 25B

Mathematics Ratings Tally by Program[75]

ACT Aspire	E	E	L	W
MCAS	E	E	E	L
PARCC	E	G	G	G
Smarter Balanced	E	G	G	L

LEGEND E Excellent Match G Good Match L Limited/Uneven Match W Weak Match

75. Although all four programs require the assessment of conceptual understanding, procedural skill/fluency, and applications (criterion C.2), final ratings could not be determined with confidence due to variations in how reviewers understood and implemented this criterion.

Summary Ratings

Tables 26A and 26B summarize the results across all tests, criteria, and subject areas. Appendix F includes these ratings plus the panel's summary statements, which are shown in separate tables throughout the *Results* section.

TABLE 26 A

ELA/Literacy Ratings Summary

Criteria	ACT Aspire	MCAS	PARCC	Smarter Balanced
I. CONTENT: Assesses the content most needed for College and Career Readiness	L	L	E	E
B.3 Reading:* Tests require students to read closely and use specific evidence from texts to obtain and defend correct responses.	L	G	E	E
B.5 Writing:* Tasks require students to engage in close reading and analysis of texts. Across each grade band, tests include a balance of expository, persuasive/argument, and narrative writing.	L	W	E	E
B.6 Vocabulary and language skills: Tests place sufficient emphasis on academic vocabulary and language conventions as used in real-world activities.	G	L	E	G
B.7 Research and inquiry: Assessments require students to demonstrate the ability to find, process, synthesize, and organize information from multiple sources.	L	W	E	E
B.8 Speaking and listening: Over time, and as assessment advances allow, the assessments measure speaking and listening communication skills.**	~~W~~	~~W~~	~~W~~	~~L~~
II. DEPTH: Assesses the depth that reflects the demands of College and Career Readiness	G	G	E	G
B.1 Text quality and types:* Tests include an aligned balance of high-quality literary and informational texts.	G	G	G	E
B.2 Complexity of texts:* Test passages are at appropriate levels of text complexity, increasing through the grades, and multiple forms of authentic, high-quality texts are used.***	G	G	G	G
B.4 Cognitive demand: The distribution of cognitive demand for each grade level is sufficient to assess the depth and complexity of the standards.	W	L	E	G
B.9 High-quality items and variety of item types: Items are of high technical and editorial quality and test forms include at least two item types with at least one that requires students to generate a response.	E	E	E	G

* The criteria recommended to be more heavily emphasized are underlined.

** The methodology indicates that criterion B.8 (speaking and listening) should be included "over time, and as assessment advances allow." Thus B.8 ratings are not included in the overall rating for Content.

*** The criterion B.2 rating is based solely on program documentation, as reviewers were not able to rate the extent to which quantitative measures are used to place each text in a grade band. Thus, reviewers did not consider the criterion B.2 rating as heavily when deciding the overall depth rating.

LEGEND E Excellent Match G Good Match L Limited/Uneven Match W Weak Match

Cells for which the ratings are not used in determining Content and Depth ratings (See Section I, *Weighting of Criteria for Content and Depth Ratings*.)

TABLE 26 B

Mathematics Ratings Summary

Criteria	ACT Aspire	MCAS	PARCC	Smarter Balanced
I. CONTENT: Assesses the content most needed for College and Career Readiness	L	L	G	G
C.1 Focus:* Tests focus strongly on the content most needed in each grade or course for success in later mathematics (i.e., major work).	W	L	G	G
C.2: Concepts, procedures, and applications: Assessments place balanced emphasis on the measurement of conceptual understanding, fluency and procedural skill, and the application of mathematics.**	–	–	–	–
II. DEPTH: Assesses the depth that reflects the demands of College and Career Readiness	G	E	G	G
C.3 Connecting practice to content:* Test questions meaningfully connect mathematical practices and processes with mathematical content.	E	E	E	E
C.4 Cognitive demand: The distribution of cognitive demand for each grade level is sufficient to assess the depth and complexity of the standards.	L	E	G	G
C.5 High-quality items and variety of item types: Items are of high technical and editorial quality and test forms include at least two item types, at least one that requires students to generate a response.	E	E	G	L

* The criteria recommended to be more heavily emphasized are underlined.

** Both programs require, in their program documentation, the assessment of conceptual understanding, procedural skill/fluency, and application, although most do not clearly distinguish between procedural skill/fluency and conceptual understanding. Also, specific balance across these three types is not required. Due to variation across reviewers in how this criterion was understood and implemented, final ratings could not be determined with confidence. Therefore, for criterion C.2, only qualitative observations are provided for grades 5 and 8. (See Section I, *Findings* for more information.)

LEGEND E Excellent Match G Good Match L Limited/Uneven Match W Weak Match

– Cells for which no quantitative rating could be determined

Program Strengths and Areas for Improvement

After completing the ratings, each subject area panel reviewed their findings and developed summary statements regarding the overall strengths and areas for improvement for each program. These comments pertain to how the programs performed against the CCSSO Criteria as operationalized. Feedback from the panels concerning future improvements to the methodology can be found in Section III.

ACT Aspire

English Language Arts:

In ELA/Literacy, ACT Aspire receives a Limited/Uneven to Good Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. The combined set of ELA/Literacy tests (reading, writing, and English) requires close reading and adequately evaluates language skills. More emphasis on assessment of writing to sources, vocabulary, and research and inquiry, as well as increasing the cognitive demands of test items, will move the assessment closer to fully meeting the criteria. Over time, the program would also benefit by developing the capacity to assess speaking and listening skills.

Content: ACT Aspire receives a Limited/Uneven Match to the CCSSO Criteria for Content in ELA/Literacy. The assessment program includes an emphasis on close reading and language skills. However, the reading items fall short on requiring students to cite specific textual information in support of a conclusion, generalization, or inference and in requiring analysis of what has been read. In order to meet the criteria, assessing writing to sources, vocabulary, and research and inquiry need to be strengthened.

Depth: ACT Aspire receives a rating of Good Match for Depth in ELA/Literacy. The program's assessments are built on high-quality test items and texts that are suitably complex. To fully meet the CCSSO Criteria, more cognitively demanding test items are needed at both grade levels, as is additional literary narrative text, as opposed to literary informational texts.[76]

Mathematics:

In mathematics, ACT Aspire receives a Limited/Uneven to Good Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. Some of the mismatch with the criteria is likely due to intentional program design, which requires that items be included from previous and later grade(s).

The items are generally high quality and test forms at grades 5 and 8 have a range of cognitive demand, but in each case the distribution contains significantly greater emphasis at DOK 3 than reflected in the standards. Thus, students who score well on the assessments will have demonstrated strong understanding of the standard's more complex skills. However, the grade 8 test may not fully assess standards at the lowest level of cognitive demand.[77] The tests would better meet the CCSSO Criteria with an increase in the number of items focused on the major work of the grade and the addition of more items at grade 8 that assess standards at DOK 1.

Content: ACT Aspire receives a Limited/Uneven Match to the CCSSO Criteria for Content in Mathematics. The program does not focus exclusively on the major work of the grade, but rather, by design, assesses material from previous and later grade(s). This results in a weaker match to the criteria. The tests could better meet the criteria at both grades 5 and 8 by increasing the number of items that assess the major work.

76. As discussed previously, ACT Aspire does not classify literary nonfiction texts that are primarily narrative in structure as "informational." See Appendix G for more information about ACT Aspire's interpretation of CCSSO criterion B.1.

77. As noted previously, reviewers did not account for PARCC's partial credit scoring model, which may lower the average DOK of the test.

Depth: ACT Aspire receives a Good Match to the CCSSO Criteria for Depth in Mathematics. The items are well crafted and clear, with only rare instances of minor editorial issues. The ACT Aspire tests include proportionately more items at high levels of cognitive demand (DOK 3) than the standards reflect, and proportionately fewer at the lowest level (DOK 1). This finding is both a strength, in terms of promoting strong skills, and a weakness in terms of ensuring adequate assessment of the full range of cognitive demand within the standards. While technically meeting the criterion for use of multiple item types, the range is nonetheless limited, with the large majority comprising multiple-choice items. The program would better meet the criteria for Depth by including a wider variety of item types and relying less on traditional multiple-choice items.

MCAS

English Language Arts:

In ELA/Literacy, MCAS receives a Limited to Good Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. The test requires students to closely read high-quality texts and a variety of high-quality item types. However, MCAS does not adequately assess several critical skills, including reading informational texts, writing to sources, language skills, and research and inquiry; further, too few items assess higher-order skills. Addressing these limitations would enhance the ability of the test to signal whether students are demonstrating the skills called for in the standards. Over time, the program would also benefit by developing the capacity to assess speaking and listening skills.

Content: MCAS receives a Limited/Uneven Match to the CCSSO Criteria for Content in ELA/Literacy. The assessment requires students to read closely well-chosen texts and presents test questions of high technical quality. However, the program would be strengthened by assessing writing annually, assessing the three types of writing (narrative, expository, and persuasive/argumentation) called for across each grade band, requiring writing to sources, and placing greater emphasis on assessing research and language skills.

Depth: MCAS receives a rating of Good Match for Depth in ELA/Literacy. The assessments do an excellent job in presenting a range of complex reading texts. To fully meet the demands of the CCSSO Criteria, however, the test needs more items at higher levels of cognitive demand, a greater variety of items to test writing to sources and research, and more informational texts, particularly those of an expository nature.

Mathematics:

In mathematics, MCAS receives a Limited/Uneven Match to the CCSSO Criteria for Content and an Excellent Match for Depth relative to assessing whether students are on track to meet college and career readiness standards. The MCAS mathematics test items are of high technical and editorial quality. Additionally, the content is distributed well across the breadth of the grade level standards, and test forms closely reflect the range of cognitive demand of the standards. Yet the grade 5 tests have an insufficient degree of focus on the major work of the grade.

While mathematical practices are required to solve items, MCAS does not specify the assessed practices(s) within each item or their connections to content standards. The tests would better meet the criteria through increased focus on the major work at grade 5 and identification of the mathematical practices that are assessed—and their connections to content.

Content: MCAS receives a Limited/Uneven Match to the CCSSO Criteria for Content in Mathematics. While the grade 8 assessment focuses strongly on the major work of the grade, the grade 5 assessment does not, as it samples more broadly from the full range of standards for the grade. The tests could better meet the criteria through increased focus on the major work of the grade on the grade 5 test.

Depth: MCAS receives an Excellent Match to the CCSSO Criteria for Depth in Mathematics. The assessment uses high-quality items and a variety of item types. The range of cognitive demand reflects that of the standards of the grade. While the program does not code test items to math practices, mathematical practices are nonetheless incorporated within items. The program might consider coding items to the mathematical practices and making explicit the connections between specific practices and specific content standards.

PARCC

English Language Arts:

In ELA/Literacy, PARCC receives an Excellent Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. The tests include suitably complex texts, require a range of cognitive demand, and demonstrate variety in item types. The assessments require close reading, assess writing to sources, research, and inquiry, and emphasize vocabulary and language skills. The program would benefit from the use of more research tasks requiring students to use multiple sources and, over time, developing the capacity to assess speaking and listening skills.

Content: PARCC receives an Excellent Match to the CCSSO Criteria for Content in ELA/Literacy. The program demonstrates excellence in the assessment of close reading, vocabulary, writing to sources, and language, providing a high-quality measure of ELA/Literacy content, as reflected in college and career readiness standards. The tests could be strengthened by the addition of research tasks that require students to use two or more sources and, over time, a listening and speaking component.

Depth: PARCC receives a rating of Excellent Match for Depth in ELA/Literacy. The PARCC assessments meet or exceed the depth and complexity required by the criteria through a variety of item types that are generally of high quality. A better balance between literary and informational texts would strengthen the assessments in addressing the criteria.

Mathematics:

In mathematics, PARCC receives a Good Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. The assessment is reasonably well aligned to the major work of each grade. At grade 5, the test includes a distribution of cognitive demand that is similar to that of the standards. At grade 8, the test has greater percentages of higher-demand items (DOK 3 and 4) than reflected by the standards, such that a student who scores well on the grade 8 PARCC assessment will have demonstrated strong understanding of the standard's more complex skills. However, the grade 8 test may not fully assess standards at the lowest level (DOK 1) of cognitive demand.

The test would better meet the CCSSO Criteria through additional focus on the major work of the grade at grade 5, the addition of more items at grade 8 that assess standards at DOK 1, and increased attention to accuracy of the items—primarily editorial, but in some instances mathematical.

Content: PARCC receives a Good Match to the CCSSO Criteria for Content in Mathematics. The test could better meet the criteria by increasing the focus on the major work at grade 5.

Depth: PARCC receives a Good Match to the CCSSO Criteria for Depth in Mathematics. The tests include items with a range of cognitive demand, but at grade 8, that distribution contains a higher percentage of items at the higher levels (DOK 2 and 3) and significantly fewer items at the lowest level (DOK 1). This finding is both a strength, in terms of promoting strong skills, and a weakness, in terms of ensuring adequate assessment of the full range of cognitive demand within the standards. The tests include a variety of item types that are largely of high quality. However, a range of problems (from minor to severe) surfaced relative to editorial accuracy and, to a lesser degree, technical quality. The program could better meet the Depth criteria by ensuring that all items meet high editorial and technical standards and by ensuring that the distribution of cognitive demand on the assessments provides sufficient information across the range.

Smarter Balanced

English Language Arts:

In ELA/Literacy, Smarter Balanced receives a Good to Excellent Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. The tests assess the most important ELA/Literacy skills of the CCSS, using technology in ways that both mirror real-world uses and provide quality measurement of targeted skills. The program is most successful in its assessment of writing and research and inquiry. It also assesses listening with high quality items that require active listening, which is unique among the four programs. The program would benefit by improving its vocabulary items, increasing the cognitive demand in grade 5 items, and, over time, developing the capacity to assess speaking skills.

Content: Smarter Balanced receives an Excellent Match to the CCSSO Criteria for Content in ELA/Literacy. The program demonstrates excellence in the areas of close reading, writing to sources, research, and language. The listening component represents an important step toward adequately measuring speaking and listening skills—a goal specifically reflected in the standards. Overall, Smarter Balanced is a high-quality measure of the content required in ELA/Literacy, as reflected in college and career readiness standards. A greater emphasis on Tier 2 vocabulary would further strengthen these assessments relative to the criteria.

Depth: Smarter Balanced receives a rating of Good Match for Depth in ELA/Literacy. The assessments use a variety of item types to assess student reading and writing to source. The program could better meet the depth criteria by increasing the cognitive demands of the grade 5 assessment and ensuring that all items meet high editorial and technical quality standards.

Mathematics:

In mathematics, Smarter Balanced is a Good Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. The test provides adequate focus on the major work of the grade, although it could be strengthened at grade 5.

The tests would better meet the CCSSO Criteria through increased focus on the major work at grade 5 and an increase in the number of items on the grade 8 tests that assess standards at the lowest level of cognitive demand. In addition, removal of serious mathematical and editorial flaws, found in approximately one item per form, should be a priority.

Content: Smarter Balanced receives a Good Match to the CCSSO Criteria for Content in Mathematics. The tests could better meet the criteria by increasing the focus on the major work in grade 5.

Depth: Smarter Balanced receives a Good Match to the CCSSO Criteria for Depth in Mathematics. The exam includes a range of cognitive demand that fairly represents the standards at each grade level. The tests have a strong variety of item types, including those that make effective use of technology. However, a range of problems (from minor to severe) surfaced relative to editorial accuracy and, to a lesser degree, technical quality. A wide variety of item types appear on each form, and important skills are assessed with multiple items, as is sound practice. The program could better meet the Depth criteria by ensuring that all items meet high editorial and technical standards.

Section II: Recommendations

The size and complexity of this report make it easy to miss its most important finding: these new, computer-based tests are major improvements over the previous generation of state tests. There is no better evidence than the fact that a best-in-class state assessment, the 2014 Massachusetts Comprehensive Assessment System, does not measure many of the important competencies that are part of today's college and career readiness standards. To be fair, MCAS is quite strong in the quality of its items and the degree to which its math test matches the depth of knowledge in the standards, but significant enhancements would be needed for it to meet the CCSSO Criteria for a high-quality assessment. In contrast, both PARCC and Smarter Balanced, while not perfect, measured a much wider array of skills and content called for in the Common Core State Standards.

But those two new assessments remain far from ubiquitous. Three years ago, forty-five states were members of either PARCC or Smarter Balanced for ELA/Literacy and mathematics. As of January 2016, nearly thirty states are planning to use either a customized state assessment or a vendor-developed option, such as ACT Aspire.[78] Based on the results of this evaluation, and with an eye to the future, we offer the following recommendations for state policymakers and test developers.

For State Policymakers

1 Make quality non-negotiable.

State assessments serve as a yardstick for gauging the quality and progress of public schools and the students in them. Whether used for high-stakes accountability purposes (such as teacher evaluation) or lower stakes (such as public reporting of school results), weak assessments leave state leaders (as well as educators and parents) with blinders on, misleading them or showing them only a portion of the truth. In the worst cases, weak assessments paint a rosier-than-justified picture of student performance, despite serious skill and knowledge deficits. This sets students up for unsuccessful transitions into college or the workplace, leaves states without the warning signals needed to protect the integrity of their future workforce, and deprives parents of an accurate picture of their children's educational progress. Quality assessments are also important for formative purposes for educators in that they reinforce the expectations of the standards, send teachers clear messages about what they should be teaching, and provide useful feedback on student progress.

The recently revived peer-review process for assessments at the U.S. Department of Education is explicitly focused on assessment quality along multiple dimensions; we view this as a promising development. That said, peer-review guidance was in place throughout the No Child Left Behind (NCLB) era, yet states still adopted assessments of questionable quality and alignment. We hope the peer-review process is taken seriously and that the Department helps to ensure that, whatever tests states adopt, that they are high quality and support effective standards implementation. The approach we used in this report provides relevant evidence about assessment quality that could feed into a peer-review evaluation.

78. National Council of State Legislators, "Hot Topics in Higher Education: Reforming Remedial Education," http://www.ncsl.org/research/education/improving-college-completion-reforming-remedial.aspx.

2 When developing or revising assessments, carefully prioritize the set of skills and knowledge at each grade that should serve as the focus of instruction, building public understanding and support as you do so.

State tests typically do not assess every grade level standard, just as none of the four tests in this evaluation assess every Common Core State Standard. The selection of the standards to be tested, then, is incredibly important, as it sends powerful signals throughout the educational system about the priorities for instructional time.

States that have joined a consortium have made these selections, delineated in their testing blueprints, and will likely be revised from time to time by the consortia. Yet the thirty or so states that are embarking solo on assessment development should pause to clarify the essential subset of skills at each grade to be assessed. The CCSSO's "Criteria for Procuring and Evaluating High Quality Assessments," the foundation of this evaluation, provides helpful guidance. This document lays out what educational leaders from a number of states, informed by research from higher education and employers, have determined to be the essential building blocks of college and career readiness. Most, if not all, of this "short list" of skills resonates with policymakers and the public as common-sense expectations. For example, the ability to read several texts about a topic and craft a well-written argument supported by evidence is an important skill for all high school graduates today. The state's employers and higher education institutions may want to augment CCSSO's short list with a few skills that are particularly important to them.

We also recommend active public involvement during this process because the assessment of college and career readiness is neither quick nor inexpensive—nor lacking controversy. Some of the skills most highly valued by colleges, such as the writing of clear, well-supported arguments or the development, testing, and refinement of mathematical models, cannot be assessed with short, low-cost, multiple-choice questions. Instead, they require a willingness to invest in extended-response items with human scoring or in technology-enhanced items. It is vital to build public support for both the cost of developing such an assessment and the time needed for students to complete it. Pressure to roll back testing decisions, such as recent opt-out initiatives, are due in part to the failure to build public understanding and buy-in for high-quality tests that measure high-priority skills and knowledge.

3 Ensure quality is maintained while addressing concerns about testing time and costs.

Even though high-quality tests tend to require more testing time and come with larger price tags, there are constructive steps that states can take to minimize both testing time and cost. A study by the Council of Great City Schools found great variability in the number of hours per year that students spend taking mandatory tests, ranging from ten hours to nearly thirty.[79] Many of the tests turned out to be redundant or not aligned to the state standards and assessments. The 2015–16 versions of the tests evaluated in this study will require no more than three hours per content area, which we find reasonable given the set of complex skills they are assessing. So a logical first step in minimizing testing time is to eliminate redundant and misaligned tests at the state and local levels. CCSSO has released a framework that states can use to identify unnecessary tests and to "ensure [that] every student is taking assessments that are necessary, of high quality and provide meaningful information."[80]

States can also ease testing costs while maintaining high-quality assessments. Three of the tests in this study serve as examples: states can share costs by participating in a consortium of states (PARCC or Smarter Balanced) or can purchase a test that has already been developed by a test provider (ACT Aspire). Both options, however, require that the state yield some control over a) the specification of the set of skills and

79. Council of Great City Schools, press release, "Student Assessments in Public Schools Not Strategic, Often Redundant," October 24, 2015, http://www.cgcs.org/cms/lib/DC00001581/Centricity/Domain/4/Testing%20Report.pdf.

80. CCSSO, "Comprehensive Statewide Assessment Systems: A Framework for the Role of the State Education Agency in Improving Quality and Reducing Burden," June 2015, http://www.ccsso.org/Resources/Publications/Comprehensive_Statewide_Assessment_Systems_A_Framework_for_the_Role_of_the_State_Education_Agency_in_Improving_Quality_and_Reducing_Burden.html.

knowledge assessed and b) the setting of cut scores. A third option is to secure high-quality test items from a credible source—such as one of the consortia or a bank of secure items shared across multiple states—while maintaining full state control over the specific items used and how the state will define proficiency.[81]

Policymakers may also encounter critics who demand the elimination of all state tests, some of whom assert that the goal of improving education could be better served by using those dollars at the classroom level. However, this is almost certainly false, considering how small testing budgets are compared to the overall costs of education. Matthew Chingos of the Urban Institute determined that the elimination of all federally and state-mandated testing would save an average of just $34 per student per year, an amount that would reduce pupil-teacher ratios by just 0.1 student.[82] Taking steps to minimize testing costs while maintaining quality is sound governance, but eliminating the yardstick altogether would be foolhardy.

4 Work with other state leaders to press the assessment industry and researchers for improvements in test item types and scoring engines to better measure key constructs in a cost-effective way.

Great advances have been made in testing over the past five years, as the majority of states have transitioned to computer-based delivery and the two multi-state consortia have used their unprecedented buying power to fund rapid research and development. Nevertheless, further advances are needed to assess accurately and efficiently all of the skills that states value and to improve the assessment of students with disabilities. The fact that forty states use the CCSS or a variant of them makes it possible for these states to use their collective influence to press for further improvements.[83] For example, the mathematical practices call for high school students to adeptly use various common technologies, such as spreadsheets and statistical software, and to "make sound decisions about when each of these tools might be helpful, recognizing both the insights to be gained and their limitations."[84] But advances in both technology-enhanced items and schools' bandwidth are needed to incorporate such technological tools and assess student's ability to use them strategically. Another example: the ELA/Literacy standards call for students to express orally well-supported ideas and probe the ideas of others, yet no current state test includes the assessment of speaking skills. By working through the CCSSO or other national organizations, all states with college and career readiness standards can prioritize areas for improvement and collectively strategize about how best to meet them.

For Test Developers

1 Ensure that every item meets the highest standards for editorial accuracy and technical quality.

Simple errors, the review panels noted, can reduce test validity and sometimes impact the accuracy of scores. When test scores are used for so many consequential purposes, these tests must maintain a very high bar for item quality. It's true that testing programs already have extensive review processes in place, but our reviewers still noted editorial and substantive issues that undermined item integrity. Even inconsequential errors can also spread rapidly through social media and news stories, exacerbating public understanding of the quality of state assessment items.

81. The Massachusetts State Board of Education voted on November 17, 2015 to award a new MCAS contract to include a next-generation assessment for English language arts and math using both PARCC items and items specific to Massachusetts. See Massachusetts Department of Elementary and Secondary Education, press release, "Board of Elementary and Secondary Education Approves Path to Next-Generation MCAS," November 17, 2015, http://www.doe.mass.edu/news/news.aspx?id=21314.

82. M. Chingos, "Testing Costs a Drop in the Bucket" (Washington, D.C.: Brookings Institution, February 2, 2015), http://www.brookings.edu/blogs/up-front/posts/2015/02/02-standardized-tests-chingos.

83. National Conference of State Legislatures, "Common Core Status Map," updated November 23, 2015, http://www.ccrslegislation.info/CCR-State-Policy-Resources/common-core-status-map.

84. Common Core State Standards, "Standards for Mathematical Practice," www.corestandards.org/Math/Practice.

2 Use technology-enhanced items (TEIs) strategically to improve test quality and enhance student effort.

Used well, TEIs can expand the depth and breadth of what can be measured and help maintain student engagement and effort during the assessment, which are important to obtaining meaningful results. Panelists were concerned, however, that in some cases TEIs (i.e., drag-and-drop items, equation editors, and line plotters) were used seemingly to no advantage—though information about their impact on student engagement was not part of the review. Because TEIs tend to be more expensive to develop than simpler item types, using them only when they enhance the quality of measurement will help to minimize costs.

3 Focus research and development on areas of targeted importance relative to measuring student performance on CCR standards.

Most of the skills and abilities the CCSS and other college and career readiness standards call for can currently be measured well. A few, however, require advances in the field of measurement and/or test development and scoring costs, if they are to be used at scale. These include:

- *Research skills:* The ability to formulate an inquiry into a topic, locate multiple sources of information, evaluate their quality, and cull evidence to develop and support a claim is a skill needed by all adults—whether it's put to use in researching jobs, choosing an insurance plan, or deciding how to vote. Current tests do not evaluate this integrated set of skills, but focus on some of the skill components such as the ability to organize information or to use evidence from sources to develop and support a claim. Test developers should have as their goal to measure robustly this integrated set of skills within reasonable cost and test-time constraints.
- *Use of technological math tools, such as spreadsheets and software designed for use in statistics, geometry, and modeling:* These tools are widely used in colleges and the workplace, which means that high school graduates need to be able to use them in solving complex problems. Incorporating such tools within state assessments is a new expectation, but one that warrants focused effort.
- *Speaking and listening skills:* Of the four assessments evaluated in this study, only Smarter Balanced appraised listening skills, and none gauged mastery of speaking skills. How to reliably measure these skills is a challenge for those in the assessment field, but advances here would obviously enhance the value of future tests.

******

We close with a note of optimism. For too many years, state assessments have generally focused on low-level skills and have given students and parents false signals about the readiness of their children for postsecondary education and the workforce. (They weren't very helpful to educators and policymakers, either.) Many students were forced to waste time and money on remedial coursework. Thus, the choices concerning state assessments are some of the most important decisions that state education leaders face. Students deserve carefully crafted tests that will measure the skills and knowledge they need to transition successfully into postsecondary education or the workplace. Educators need and deserve good tests that honor their hard work and give useful feedback, which enables them to improve their craft and boost their students' success. And policymakers, parents, and the public need tests that will tell them whether their students are developing the skills and knowledge that today's high school graduates need.

States' adoption of college and career readiness standards has been a bold step in the right direction. The fact that multiple states are already using ACT Aspire, PARCC, and Smarter Balanced, each of which is as good or better than the previous best-in-class Massachusetts state assessment, fuels our optimism. Yet, adopting and sticking with high-quality assessments requires courage. These tests are tougher, sometimes cost more, and can require more testing time than the previous generation of state tests.

Are we up to the task?

Section III: Suggestions for Methodological Improvement

The methodology used in this study is highly comprehensive and includes evaluating individual test items, passages, prompts, and documentation along many dimensions. Given the breadth and scope of what reviewers were asked to do, it's notable that they deemed the final results an accurate analysis of each assessment against the CCSSO Criteria.

For instance, one ELA/Literacy reviewer noted:

> *"This methodology's greatest strength is its comprehensiveness of approach. It takes all tests' supporting documents into consideration, relies on quality training of evaluators, and sound research approaches to arrive at useful and worthy metrics on how state/national assessments are doing."*

A mathematics reviewer also remarked:

> *"The development from scratch of a procedure that obtains such an in-depth review of a variety of tests is impressive. I am sure there are refinements to the process that can give a better evaluation of these tests, but the one we implemented gave, in my opinion, a fair and in depth evaluation of the tests."*

As compared to prior alignment methodologies (e.g., Webb's alignment tool and the Surveys of Enacted Curriculum), the task for reviewers implementing this methodology is great. While the methodology offers some advances over existing methodologies—most notably its close reflection of the principles of new college and career readiness standards—there are ways in which it could be improved to provide a more complete picture of the quality of new assessments.

As the first implementers of these new methods, it is not surprising that we identified a number of improvements that could be addressed in subsequent iterations. In this chapter, we lay out the challenges, and, where appropriate, suggestions for revising the methodology to address these limitations—many of which would make it even more comprehensive (as well as time- and labor-intensive). The section is organized by content area and criterion.

English Language Arts/Literacy

Methodological Concerns

B.1 This criterion includes both text quality and text type/balance. Several reviewers noted that these are different dimensions, and would have preferred that they not be combined. For future implementations of the methodology, we recommend that text quality and text type/balance be reported separately.

B.1.1 Balance of text types was determined simply by counting the number of texts of each type that students encountered on a test form. While this is a reasonable approach, reviewers noted that some passages were longer or included many more items than others, and that these might also be appropriate ways to weight text passages. We recommend considering these alternative ways of weighting text passages, either in the study design or in the roll-up from items to forms, to determine whether the results are sensitive to the choice of weighting scheme.

B.1.2 The purpose of this criterion is to evaluate the quality of texts—whether they are previously published or of publishable quality. In practice, almost all texts used on these assessments were previously published, so tests generally received full credit for this criterion. However, reviewers felt that text quality varied among texts, even though they were all previously published, but the methodology did not offer them the opportunity to render a judgment on text quality.

We recommend that future iterations of the methodology consider how to provide a measure of text quality that includes reviewer judgments.

B.1.3 The purpose of this criterion is to evaluate the structure of informational texts, with the goal being that "nearly all informational texts are expository in structure" (as opposed to narrative in structure). To operationalize this goal, the suggested cutoffs are 90 percent or above of informational text being expository for a score of 2, 75 percent to 89 percent for a score of 1, and 74 percent or below for a score of 0. Unfortunately, these suggested cutoffs don't work well with tests that have a small number of informational texts. For example, if an assessment has just two informational texts, it can only earn a score of 2 (fully met) or 0 (not met) on this criterion.

This is one of several of the criteria that are challenging to implement for a single test form. We recommend that certain criteria might be more appropriately evaluated across multiple forms, and this may be one. For instance, it is not necessary to go through an entire test form item by item to evaluate text passages on their structure; it may be more appropriate to evaluate this criterion across many forms. (Of course, this approach is not possible for tests, such as MCAS, that have just one or two forms.) Evaluating passages on multiple forms will result in a more precise determination of text passage quality and structure. Alternatively, the suggested cutoffs in the methodology might be made more flexible to account for the fact that there are typically small numbers of passages (and, as indicated, can sometimes swing the score from a 2 to a 0).

B.3.1 and B.3.4 These two criteria focus on "close reading" and "direct textual evidence." In general, reviewers understood the latter to be a subset of the former, but even after extensive training and discussion, there were often disagreements about how these two terms should be understood and implemented. The variability was often around whether direct textual evidence was truly required when the student only had to use the text to find an answer, as opposed to having to show the actual text evidence that supported an answer. It is not clear whether the misunderstanding was a problem with the methodology or with the training provided to reviewers in this study.

We recommend strengthening and clarifying definitions of these two terms and providing reviewers with more detailed training on them.

B.4.1 The methodology calls for the highest score when the distribution of DOK on a form matches that of the CCSS (a DOK index of .80 or above), and for lower scores to be assigned as the degree of match declines. However, this guideline does not differentiate between forms where the DOK misalignment is due to the test having too low DOK versus too high. There was broad agreement among reviewers that tests with a greater emphasis on DOK 3 and 4 items should be rated more positively than tests with a greater emphasis on DOK 1 and 2.

There are many possible solutions to this problem. The best approach would likely be to add guidance in the scoring criteria to account for this possibility. For instance, one way to determine whether the DOK misalignment occurred because the test is "too high" versus "too low" would be to create an aggregate DOK value for each test by treating the DOK index as ordered.[85] Revised guidance could specify that a lower DOK index threshold would be acceptable if the average DOK of the test exceeded that of the standards. Another approach would be to specify a priori DOK distributions that would be acceptable—for instance, one-third each at levels 1, 2, and 3 to 4.

B.6.3 and B.6.4 On these criteria, assessments get top scores when they report a sub-score for language or vocabulary, even when those assessments include very few items in those sub-scales. In contrast, some assessments that have many items on vocabulary or language but do not report a sub-score receive no or limited credit. While it is important to provide parents and teachers with direct feedback in the form of sub-scores about student knowledge of language and vocabulary, reviewers felt that assessments should not receive full credit if those sub-scales were based on an unreliably low number of items.

We recommend that the scoring guidance be revised so that assessments that have language or vocabulary sub-scales cannot receive the maximum score, unless they have an adequate number of test items in those areas.

B.7.1 This criterion evaluates test items assessing research. Reviewers felt the definition of "research item" was too broad and allowed credit for items that did not truly mirror research activities.[86]

We recommend that the definition of a research item be enhanced to ensure that the focus is on the use of two or more discrete sources (e.g., texts, audio recording, or videos) and of research skills applied in an authentic way. Also, the sufficiency of research items (e.g., the percentage of the test devoted to research items) should be addressed, because a test that has only a single research item comprising just a small proportion of the total score points could score a 2 (met the criterion) under the current methodology.

B.9.1 The current scoring guidance for this criterion requires only a single non-multiple choice test item to earn a score of 2. However, reviewers noted that assessments varied a great deal in the extent to which they employed a variety of closed and open-ended questions. Reviewers believed that assessments with a wider variety of item types measuring more authentic skills should receive more credit.

We recommend that the scoring guidance be revised to set suggested score point cutoffs for constructed-response items. For example, the scoring guidance might indicate that the assessment must have 25 percent of score points from constructed-response items in order to receive a score of 2.

Implementation Concerns

B.5.2 This criterion is focused on the proportion of writing prompts that require writing to a source. However, the denominator used to calculate this proportion is the number of items coded to a writing standard. Thus, depending on how vendors coded writing items, some assessments had many items that counted toward this criterion that were not actual writing items. This would have resulted in them failing the criterion if we did not manually correct the issue as we were implementing the study.

We recommend that future versions of the coding document use the correct denominator for this calculation, which is the number of writing prompts.

85. This approach was used with a somewhat different DOK scale in Polikoff & Struthers, 2013. As an example, suppose the standards were 25 percent at DOK 1 and 2 and 50 percent at DOK 3, while the assessment was 10 percent at DOK 1 and 2 and 80 percent at DOK 3. The DOK index for this comparison would be .70, well below the cutoff for a 2 (which is .80). But treating the scale as ordered the standards would have an average DOK of 2.25 (25%*1 + 25% *2 + 50% * 3), while the test would have an average DOK of 2.7.

86. That definition is as follows: These tasks require students to use multiple (minimum of two) informational texts about one topic (may be written, audio, or visual), analyze these multiple texts, and synthesize and/or organize information across texts. Research tasks may include multiple-choice and technology-enhanced items and must include at least one constructed-response item for which the student writes a response. See Appendix B, *Key Terminology*.

B.9.2 This criterion is focused on item quality. According to the methodology, two measures are used to determine the rating: the proportion of items with technical or editorial issues and the proportion of items for which reviewers agree with the vendors' alignment determinations (as provided in the metadata). The reviewers felt strongly that it was inappropriate to use alignment in evaluating this criterion, especially because it had been used in other criteria (e.g., B.3.1). Further, using alignment to evaluate B.9.2 is not fair to vendors who take different approaches, such as aligning items to claims rather than to individual content standards. Finally, for an item to meet the alignment criterion for B.9.2, the reviewers had to agree with all of the alignment ratings offered by the vendor, which may differentially affect vendors that list more standards compared to those who list fewer.

We recommend the approach used in our implementation—pulling apart alignment from item quality. For this sub-criterion, we believe reviewers should only evaluate items for technical and editorial quality. If other implementers also want to evaluate the extent to which the vendors' alignment ratings are accurate, that could either be presented as a separate index, or as its own sub-criterion with scoring guidance.

Mathematics

Methodological Concerns

C.1.1 This criterion concerns the extent to which the test focuses on the major work of the grade. While it is required that the large majority of the major work clusters be assessed,[87] the methodology does not require that each cluster be adequately addressed. Thus, a test that has just one item per cluster receives full credit, even though the test could not possibly provide a reliable score for that cluster. As a related point, the methodology does not call for a review of the degree to which the items assessing a given cluster address different standards or skills within that cluster or are redundant. As a result, the perception is that tests covering all the clusters could meet this criterion, even if they are quite unbalanced and only measure a small slice of the standards.

We recommend revising this criterion to include a measure of balance of coverage across the major work clusters. For instance, Webb's methodology includes measures of balance that could be applied at the cluster level. This would help ensure that assessments are not overly focused on a small subset of the standards in the major work.

C.2.1 This criterion assesses the extent to which the assessment is adequately balanced across application, conceptual understanding, and procedural fluency. There were a number of difficulties with the implementation of this criterion. First, the requirement of categorizing items by their predominant focus led to a failure to recognize and give credit for items that address two or more categories. Second, and related, items that were coded as measuring "combined" skills were not counted in any way, so assessments with more "combined" items were penalized in accordance with the tentative scoring guidance. Third, the broad definition of "application" (i.e., any item that includes a context) resulted in many items that also assessed conceptual understanding and/or procedural skill/fluency to be categorized as only application because they included use of a context (even if trivial). This resulted in a lowered rating and a failure to recognize the other competencies being addressed.

To address these issues, we recommend that the methodology be revised to allow reviewers to indicate 1, 2, or 3 of the available skill categories, rather than encouraging them to select just one and penalizing the assessment if "combined" is selected. Another approach would allow raters to allocate emphasis across each category (e.g., rate an item as two-thirds procedural and one-third application). If either of these changes were made, the scoring criteria need not be changed—a goal of equal balance of application procedures, and conceptual understanding is still appropriate and could be calculated easily.

87. The major work clusters are the groups of content standards that correspond to the major work of the grade. The clusters are listed here, "Key Instructional Shifts of the Common Core State Standards for Mathematics," http://achievethecore.org/content/upload/Focus%20in%20Math_091013_FINAL.pdf.

C.3.1 This criterion is intended to evaluate the test's measurement of the standards for mathematical practice (SMP). However, as currently implemented, it is inadequate in this regard. The methodology simply requires that items coded to a SMP also be coded to align with *any* content standard—not even grade level content standards. The result is that all assessments either earn a perfect score for this criterion (because they code to the SMPs) or a zero (because they do not). A related concern is that the methodology requires reviewers to accept the program's designations of alignment to SMPs, and these were, at times, found to be generous to aggressive. Finally, the methodology does not require coverage of the SMPs, allowing a program that assesses only one or two SMPs across all items to be awarded a high rating.

We have several recommendations for this criterion. First, reviewers might be asked to evaluate vendors' claims of SMP alignment, rather than accepting them. This will require additional training and time to complete the review. Second, the methodology might allow reviewers to evaluate the extent to which the mathematical practices are adequately covered by the assessment. Again, this could be done using indices of coverage or balance, such as those employed in the Webb alignment methodology. Third, implementers might remove the requirement of content standard coverage from this criterion; every item on every test was reported as aligned to a content standard.

C.4.1 The same issues and recommendations apply in mathematics as for criterion B.4.1 in ELA/Literacy.

Another issue in mathematics was that reviewers were dissatisfied that the difficulty of the item was not captured in the methodology. Some reviewers felt that difficulty was a de facto component of the mathematics DOK levels, with DOK 1 and 2 seen as focusing on the number of steps and DOK 3 and 4 focusing on the level of reasoning. We recommend considering item difficulty as a separate dimension.

A final issue in mathematics (although an issue in both subjects, it was only raised by the mathematics panel) is that some reviewers were uncomfortable with the methodology's requirement that the DOK distribution of the test match the DOK distribution of the standards, since the standards themselves do not call for such a match. As mentioned above in the ELA/Literacy section, one solution to this problem would be to specify an a priori distribution of DOK that is desirable.

C.5.1 The same issues and recommendations apply in mathematics as for criterion B.9.1 in ELA/Literacy.

C.5.2 Although the methodology distinguishes between technical quality and editorial accuracy, it does not distinguish between major quality issues that would impact score accuracy or meaningfulness, and trivial quality issues that would not have such impacts. For example, panelists found numerous instances in which multiple strategies could be used to solve a given test item. If the purpose of the item is to determine whether the student understands and can apply one specific strategy, having multiple could weaken the meaning of the score and, consequently test validity. While this is not a problem for constructed-response items where work is evaluated, it can be for multiple-choice items if the student's work is not provided.

We recommend that additional options be given to allow reviewers to indicate the severity of the item quality problem. Furthermore, we recommend reviewers be allowed to indicate cases in which the manner in which a student solves a problem does not invalidate an item, but does seriously obscure what competencies the item is testing. The panel also recommends that reviewers be given access to the vendor's worked solutions for all test items, so that the intention for student-solution methods can be directly evaluated.

Across Subjects

Finally, we summarize several concerns that apply to both mathematics and ELA/Literacy.

First, reviewers noted that the methodology does not allow them to express enthusiasm for items, only to indicate problems. In both subjects reviewers identified a number of items they viewed as especially high quality and they had no way to differentiate them. We recommend that the methodology allow reviewers to note and describe especially high-quality items based on their professional judgment.

Second, the quality of the discussions during Phase 2 of the methodology (when data are rolled up from test forms to programs) depended in large part on the quality of the comments left by reviewers throughout the process. Because only certain comments populated forward in the Excel review sheets designed for the study, it was not always possible for the Phase 2 reviewers to understand the rationales behind other reviewers' scoring decisions. We recommend that reviewers be required to provide substantive comments at key locations in the coding documents, so that these comments can be carried forward and inform subsequent discussions.

Third, the methodology, as written, required reviewers to examine answer keys to ensure technical accuracy. However, several test vendors had privacy policies that prohibited their releasing answer keys to reviewers, or that limited answer-key release to specific times and places (i.e., in locked rooms using paper answer keys). In practice, this made the review of answer keys impractical. For the purposes of this review we made two changes to the methodology, which we recommend in cases where the answer key is not fully available across all assessments under review:

- Remove the option "item incorrectly keyed" from the list of possible quality issues in B.9.2 and C.5.2; and
- Instruct reviewers to use the option "unintended correct answer" to identify selected-response items where there appear to be more correct answers than allowed (e.g., two correct answers exist though the item calls for just one). While these solutions limit the scope of the item quality review, we see no practical alternative if vendors have strict answer key privacy policies.

Fourth and finally, the methodology relies heavily on test vendors' metadata, especially their standards alignment data. But several programs do not take a one-to-one standards-to-item approach to test construction. While reviewers found a way to work around this problem, the quality of the item alignment determinations may have been less accurate from those vendors that used an evidence-centered design, resulting in weaker scores. Considering how to make the methodology more useful for all programs, regardless of their approach to alignment, may be beneficial.

Summary

Taken together, these suggestions provide a wealth of good counsel from our reviewers about how the methodology could be improved for future implementation. Though there are many recommendations, reviewers were nonetheless quite satisfied overall with the methodology as written and implemented and broadly agreed with the conclusions reached.

We are confident that the methodology, as implemented in this study, represents an important first step in rethinking techniques to evaluate test quality, which should inform future generations of assessments in U.S. schools. With some or all of the proposed revisions incorporated, these methods will provide an even stronger, clearer message about the kind of assessment we need to ensure valid measurement of student mastery of college and career readiness standards.

Appendix A

Depth of Knowledge (DOK) of the Four Assessment Programs as Compared to the CCSS and Other Policy-Relevant Assessments

We also conducted a supplementary analysis of the Depth of Knowledge (DOK) of the four assessment programs against the Common Core State Standards (CCSS). HumRRO had previously conducted a DOK analysis of the CCSS, so we started with those results. Next, we contracted with one subject-area expert in each subject to independently code the standards in terms of DOK (the content standards only, not the Standards for Mathematical Practice). Finally, we contracted with another content-area expert in each subject to adjudicate in any instance where our reviewers disagreed with the HumRRO reviewers. In each instance, reviewers were allowed to place each standard into one or more DOK levels, and the standards were equally weighted to arrive at the final DOK distribution for the standards, as is common in alignment studies.[88] To help contextualize the findings, we compared the DOK of our four assessments against CCSS and against two analyses of a) a set of state tests in the NCLB era regarded as having the highest likelihood of assessing deeper knowledge and b) national and international assessments previously conducted by the RAND Corporation in 2012 and 2014, respectively.[89] The results of this analysis are shown below.

In broad terms, DOK refers to the cognitive demands required by a task to produce an acceptable response. In each subject, there are four levels of DOK. Generally, level 1 refers to rote or reproductive skills (e.g., identifying an obvious detail in a text, conducting a straightforward one-step operation in mathematics). Level 2 refers to skills and concepts such as multi-step operations in mathematics and comprehension across one or more sentences. Level 3 refers to strategic thinking, such as solving a mathematics problem that has multiple possible approaches or identifying complex themes across an entire passage. Level 4 refers to extended thinking, such as extended mathematical investigations or synthesis and analysis across multiple texts.

The results in ELA/Literacy, shown in Table A.1, show that the DOK of the PARCC assessments is the highest of the four studied at both grade levels. This difference is especially large in eighth grade, where nearly 70 percent of PARCC score points are on DOK 3 or 4. In fifth grade, the other three assessments fall short of the CCSS's emphasis on DOK 3+, with Smarter Balanced the lowest as 22 percent. In eighth grade, ACT Aspire has just 19 percent of score points at DOK 3+, as compared to 46 percent in the standards and 36 percent or more for the other assessments. With few exceptions, the assessments studied here have far higher proportions of DOK 3+ score points than was found on the fourteen NCLB-era state tests studied by Yuan and Le.

The results in mathematics, shown in Table A.2, show that ACT Aspire has a much larger proportion of DOK 3+ score points than the other assessments studied in this or the two Yuan and Le studies. The difference is especially notable at grade 5. With the exception of MCAS at grade 5, the assessments studied here meet or exceed the DOK of the standards. They also typically exceed the higher DOK emphasis of the fourteen NCLB-era assessments, and they match up favorably to assessments such as AP and PISA.

88. A.C. Porter, "Measuring the Content of Instruction: Uses in Research and Practice, *Educational Researcher* 31, no. 7: 2002, 3–14.

89. Yuan and Le, 2012.

TABLE A-1

Depth of Knowledge of ELA/Literacy Assessments

Reading	DOK 1	DOK 2	DOK 3	DOK 4
CCSS grade 5	17.6%	36.8%	37.5%	8.1%
ACT Aspire grade 5	34.3%	37.5%	28.1%	0.0%
MCAS grade 5	9.6%	63.5%	26.9%	0.0%
PARCC grade 5	4.5%	45.0%	50.5%	0.0%
Smarter Balanced grade 5	19.0%	58.9%	20.8%	1.2%
CCSS grade 8	10.1%	44.2%	41.8%	3.8%
ACT Aspire grade 8	44.3%	36.8%	15.0%	3.8%
MCAS grade 8	4.8%	58.7%	33.7%	2.9%
PARCC grade 8	1.6%	29.1%	46.4%	22.9%
Smarter Balanced grade 8	15.0%	40.8%	36.7%	7.5%
14 states	31.5%	43.8%	23.3%	1.6%
AP	11.0%	33.0%	56.0%	0.0%
NAEP	20.4%	45.2%	33.3%	1.1%
PISA	37.3%	26.2%	36.5%	0.0%
TIMSS	55.2%	17.8%	26.1%	0.9%

Note: Results for AP, NAEP, PISA, PIRLS and fourteen states from Yuan and Le (2014). IB results not provided because IB does not have a reading assessment. AP, NAEP, PISA, and PIRLS results are for reading assessments, whereas MCAS, PARCC, Smarter Balanced, ACT Aspire, and CCSS results are for combined ELA/Literacy assessments. The programs evaluated in this study are based on percentages of score points, whereas other assessment results are based on percentages of items.

TABLE A-2

Depth of Knowledge of Mathematics Assessments

Math	DOK 1	DOK 2	DOK 3	DOK 4
CCSS grade 5	43.3%	49.6%	7.1%	0.0%
ACT Aspire grade 5	23.0%	40.2%	36.8%	0.0%
MCAS grade 5	40.3%	57.7%	2.0%	0.0%
PARCC grade 5	34.4%	54.5%	11.1%	0.0%
Smarter Balanced grade 5	46.6%	35.6%	15.2%	2.6%
CCSS grade 8	50.9%	39.8%	9.3%	0.0%
ACT Aspire grade 8	19.9%	45.1%	34.3%	0.7%
MCAS grade 8	40.2%	45.8%	14.0%	0.0%
PARCC grade 8	13.3%	62.0%	24.2%	0.5%
Smarter Balanced grade 8	16.1%	74.5%	8.6%	0.8%
14 states	41.2%	51.0%	7.9%	0.0%
AP	30.7%	54.2%	14.0%	1.0%
IB	21.0%	50.0%	28.0%	1.0%
NAEP	38.6%	54.0%	6.7%	0.6%
PISA	32.8%	50.4%	16.8%	0.0%
TIMSS	52.0%	46.5%	1.5%	0.0%

Note: Results for AP, IB, NAEP, PISA, TIMSS and fourteen states from Yuan and Le. The programs evaluated in this study are based on percentages of score points, whereas other assessment results are based on percentages of items.

Appendix B

Key Terminology

Accessibility: The degree to which an assessment allows all test takers, including English language learners and those with disabilities, to demonstrate their mastery of the knowledge, skills, and abilities being tested.

Alignment: The degree of agreement, overlap, or intersection among standards, instruction, and assessments.[90]

Applications: A type of mathematics item that includes a context and requires students to use that context to decide which mathematical skills or concepts to use to solve the problem.[91]

Blueprints: A series of documents that together describe the content and structure of a test.

Calibration: A process used to make sure that evaluators or reviewers apply the scoring standards correctly and uniformly.[92]

Claims: Statements about the knowledge, skills, or abilities of test takers who have attained a specified level of performance on the test.

Close reading: Close reading of text involves an investigation of a short piece of text, with multiple readings completed over multiple instructional lessons. Through text-based questions and discussion, students are guided to deeply analyze and appreciate various aspects of the text, such as key vocabulary and how its meaning is shaped by context; attention to form, tone, imagery and/or rhetorical devices; the significance of word choice and syntax; and the discovery of different levels of meaning as passages are read multiple times.[93]

Cognitive demand: The type of thinking required to solve a task. (See DOK for additional information.)

College and career readiness: The level of academic preparation (in this case in ELA/Literacy and mathematics) needed by a student to both enroll in and successfully complete entry-level postsecondary collegiate or vocational programs without remedial academic work or assistance.

Computer-adaptive testing (CAT): A type of testing in which the questions presented to the test taker are selected on the basis of the test taker's previous responses. Correct answers by the test taker lead to harder questions; incorrect answers lead to easier questions. The purpose of adaptive testing is to use testing time

90. N. L. Webb, "Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education," Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 6 (Madison, WI: University of Wisconsin–Madison, Wisconsin Center for Education Research, 1997).

91. The National Center for the Improvement of Educational Assessment, Inc. (NCIEA), "Guide to Evaluating Assessments Using the CCSSO Criteria for High Quality Assessments: Focus on Test Content" (Dover, NH: NCIEA, February 2016): http://www.nciea.org/publication_PDFs/Guide%20to%20Evaluating%20CCSSO%20Criteria%20Test%20Content%20020316.pdf.

92. See Educational Testing Service, "Glossary of Standardized Testing Terms," https://www.ets.org/understanding_testing/glossary/.

93. S. Brown and L. Kappes, *Implementing the Common Core State Standards: A Primer on "Close Reading of Text"* (Washington, D.C.: The Aspen Institute, 2012).

more efficiently and to produce more precise scores for students toward the extremes of the performance distribution.[94]

Conceptual Understanding: In mathematics, problems that assess conceptual understanding require students to use their understanding of mathematical concepts, operations and relations, as opposed to memorized facts, to solve problems.[95]

Constructed-response item: A test question that requires the test taker to supply the answer, instead of choosing it from a list of possibilities. These items can take the form of a single word, number, or symbol, to a sentence, equation, or full paragraph.

Cut score: A specific test score used for classifying the test takers into groups on the basis of performance. Scores at or above that point are interpreted to mean something different from scores below that point, such as grade-level proficiency or lack thereof.

Depth: In the context of this study, Depth is a rating of the degree to which a test or testing program assesses the depth and complexity of skills and knowledge called for by college and career readiness standards.

Depth of Knowledge (DOK): The complexity or depth of understanding required to answer or explain an assessment related item, developed by Norman Webb. Level 1 includes basic recall of facts, concepts, information, or procedures; Level 2 includes skills and concepts, such as the use of information (graphs) or requires two or more steps with decision points along the way; Level 3 includes short-term strategic thinking; Level 4 includes extended thinking and, often, the application of concepts. Levels 3 and 4 are also referred to as "higher-order thinking skills."[96]

Evidence-based selected response items: These items feature two parts. The first (Part A) resembles a traditional multiple-choice item, and the second (Part B) calls for students to select/provide evidence to support the answer they chose in the first part. Part B can be multiple-choice, short-answer constructed response, multi-select, or technology-enhanced.

Evidence-centered design: A systematic approach to test development that involves a) the development of claims (see above), b) statements that describe the evidence needed from the student to prove the claims, and c) tasks designed to provide that evidence.[97]

Form: A set of test questions given to a student.

Generalizability: The degree to which the inferences drawn from a sample are representative of the whole population.

High-quality texts: Texts that are content rich, exhibit exceptional craft and thought, and/or provide useful information.[98]

Higher-order thinking skills: Complex thinking skills, including analysis, evaluation, creation, logical reasoning, judgment and critical thinking, as well as problem solving.

94. See Glossary of Education Reform, "Computer-Adaptive Test," updated August 29, 2013, http://edglossary.org/computer-adaptive-test.

95. The National Center for the Improvement of Educational Assessment, Inc. (NCIEA), "Guide to Evaluating Assessments Using the CCSSO Criteria for High Quality Assessments: Focus on Test Content" (Dover, NH: NCIEA, February 2016): http://www.nciea.org/publication_PDFs/Guide%20to%20Evaluating%20CCSSO%20Criteria%20Test%20Content%20020316.pdf.

96. N. L. Webb, "Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education."

97. R. J. Mislevy, et al., "A Brief Introduction to Evidence-Centered Design," CSE Technical Report 632, (Los Angeles, CA: University of California–Los Angeles, 2004).

98. *Ibid.*

Item: A test question, including the question itself, any stimulus material provided with the question, and the answer choices (for a multiple-choice item) or the scoring rules (for a constructed-response item).[99]

Multi-select items: Similar to multiple-choice items, but students can (or must) select more than one correct answer.

Operational items: Test items administered to students as part of their official state assessment.

Performance task: Small clusters of items that address a specific scenario and ask students to perform a task to demonstrate their knowledge, understanding, and proficiency. They can be used to assess several standards or outcomes, and typically require students to apply knowledge and skills to novel situations.

Procedural fluency: As defined in the CCSS, "skill in carrying out procedures flexibly, accurately, efficiently, and appropriately."[100]

Proficiency: The level of performance needed to meet or exceed grade level standards.

Research tasks: These tasks require students to use multiple (minimum of two) informational texts about one topic (may be written, audio, or visual), analyze these multiple texts, and synthesize and/or organize information across texts. Research tasks may include multiple-choice and technology-enhanced items and must include at least one constructed-response item for which the student writes a response.[101]

Rubric: A set of rules for scoring the responses on a constructed-response item. Sometimes called a "scoring guide."

Summative assessment: An assessment administered at or near the conclusion of a grade level or course to monitor student learning and/or to meet accountability requirements.

Task: Also referred to as a "performance task," an activity that requires students to select and apply a set of skills to respond. Tasks typically involve a scenario and a series of test items, one or more of which are constructed-response items.

Technology-enhanced items: Items that require students to perform some action that cannot be executed on a traditional paper-and-pencil test, such as drag-and-drop, highlighting, or sequencing of objects.

Tier 2 vocabulary: According to the CCSS, these are "general academic" words that are far more likely to appear in written text than in speech. They "appear in all sorts of texts: informational texts (words such as *relative, vary, formulate, specificity*, and *accumulate*), technical texts (*calibrate, itemize, periphery*), and literary texts (*misfortune, dignified, faltered, unabashedly*)."[102]

Traditional multiple-choice items (also referred to as selected response): Items that require the student to select one answer from a set of provided options.

Writing to sources: A writing activity that requires students to read or listen to one or more sources (texts, videos, audio files, etc.) and to draw evidence from the source(s) to support a conclusion, generalization, or inference in the written response.[103]

99. See Educational Testing Service, "Glossary of Standardized Testing Terms," https://www.ets.org/understanding_testing/glossary/.

100. CCSS, "Standards for Mathematical Practice," http://www.corestandards.org/Math/Practice/.

101. The National Center for the Improvement of Educational Assessment, Inc. (NCIEA), "Guide to Evaluating Assessments Using the CCSSO Criteria for High Quality Assessments: Focus on Test Content" (Dover, NH: NCIEA, February 2016): http://www.nciea.org/publication_PDFs/Guide%20to%20Evaluating%20CCSSO%20Criteria%20Test%20Content%20020316.pdf.

102. CCSS, "Common Core State Standards for English/Language Arts and Literacy in History/Social Studies, Science, and Technical Subjects," Appendix A, http://www.corestandards.org/assets/Appendix_A.pdf, 33–36.

103. *Ibid*.

Appendix C

The Methodology as Written

ELA/Literacy Criteria

There are nine criteria in English language arts (see Section I, *The Study Criteria*). Underneath these nine criteria are a total of thirty-nine sub-criteria. Of these thirty-nine sub-criteria, twenty-three are outcome sub-criteria (based on test items or forms) and sixteen are documentation sub-criteria.

Criterion B.1
The assessments are English language arts and literacy tests that are based on an aligned balance of high-quality literary and informational texts.

Item review. There are six sub-criteria under B.1. These sub-criteria focus on the texts that students are required to read as part of the tests. In particular, criteria B.1.1 and B.1.4 focus on the balance of informational and literary texts, criteria B.1.2 and B.1.5 focus on the quality of texts, and criteria B.1.3 and B.1.6 focus on the type of informational texts. These correspond to major focuses of the standards—that students should read both informational and literary texts (transitioning to a greater proportion of informational texts in higher grades), that these texts should be high quality and authentic, and that informational texts should be balanced across types (science/technical, history/social science, and literary nonfiction).

For the item review, raters read each text passage on the test and make four judgments about the text passages. First, they decide whether the text passage is literary (fiction) or informational (non-fiction). Second, for informational texts, they decide whether the text is primarily narrative (to tell a story a series of events) or expository (to inform or explain). Third, for informational texts on grades 6–12 assessments, they decide whether the text is science/technical, history/social science, or literary nonfiction (selecting more than one option is acceptable on this criterion). Fourth, they review the test metadata (the details provided by the test vendor to describe the characteristics of test items and passages), indicating whether the text has been previously published, and make a judgment as to whether the text is of publishable quality.

Form roll-up. To reach a rating for each of the criteria for a test form, the ratings from individual items must be rolled up to the test form (i.e., to give a complete picture of each test form on each criterion). For all test form roll-ups, individual reviewers use the tentative scoring guidance offered in the methodology. For instance, middle school criterion B.1.1 reads:

> *Texts are balanced across literary and informational text types and across genres, with more informational than literary texts used as the assessments move up in the grade bands. Goals include: In grades 3–8, approximately half of the texts are literature and half are informational.*

This is operationalized in the tentative scoring guidance as 2 – Meets: 45–55 percent informational texts; 1 – Partially Meets: 33–44 percent or 56–84 percent; and 0 – Does Not Meet: 0–32 percent or 85–100 percent. Reviewers classify each passage on the test form as literary or informational, then calculate the proportion of

text passages that are informational on the entire test form, which is compared against the scoring guidance to create a tentative score. The reviewer is then given the opportunity to consider whether the tentative score matches her assessment of the entire test form. If not, he or she may change the score and provide comments. The reviewer's final score is passed on to Phase 2 of the review. The other two sub-criteria are described in the full scoring guidance in the appendix.

Document review. The documentation review for B.1 criterion is quite similar, as seen in the full scoring template. The proposed scoring guidelines are the same and the three documentation sub-criteria parallel the item review sub-criteria one-to-one.

Criterion B.2
The assessments require appropriate levels of text complexity; they raise the bar for text complexity each year so students are ready for the demands of college- and career-level reading no later than the end of high school. Multiple forms of authentic, previously published texts are assessed, including written, audio, visual, and graphic, as technology and assessment constraints permit.

Criterion B.2 focuses on text complexity. The goal of this criterion is to ensure that students are presented with grade-appropriate texts, as defined by both qualitative and quantitative measures of text complexity. The complexity of assessment passages should increase across grades, just as the complexity of readings in ELA/Literacy classes, such that the student is ready for college- or career-level reading by the end of high school.

Item review. To evaluate sub-criterion B.2.1 for the item review, raters again focus on the individual passages. Here, there are three ratings. First, reviewers indicate whether there is evidence that both quantitative and qualitative measures of text complexity have been used to select the passage. Second, reviewers indicate whether there is evidence that the text has been placed in an appropriate grade band (the CCSS grade bands are 2–3, 4–5, 6–8, 9–10, and 11–12) based on quantitative text complexity data. Finally, reviewers indicate whether there is evidence that the text has been placed in an appropriate grade level based on qualitative text complexity. The scoring at the form level is straightforward: forms are evaluated on the proportion of texts that fall into appropriate grade bands and levels based on the quantitative and qualitative evidence.

Document review. The documentation review for B.2 involves just one criterion, which asks reviewers to evaluate the process that is used to select texts and the extent to which that process adequately uses quantitative and qualitative text complexity data.

Criterion B.4
The assessments require all students to demonstrate a range of higher-order, analytical thinking skills in reading and writing based on the depth and complexity of college- and career-ready standards, allowing robust information to be gathered for students with varied levels of achievement.

While the first two criteria focus on reading passages, the remaining seven focus on test items. Criteria B.4 and B.9 are recommended for reviewers to evaluate first, because these are the only criteria that are evaluated for all items. The other criteria focus on only subsets of items depending on what is being tested (e.g., reading, writing, and language).

Outcome review. Criterion B.4 focuses on the cognitive demand of the test items relative to the standards. To evaluate the test items, reviewers use the Webb Depth of Knowledge (DOK) framework specific to the content area being evaluated.[104] Each item is rated on the DOK framework; items may be placed into one or two of the four available levels.

104. N. L. Webb, "Depth of Knowledge Levels for Four Content Areas," 2002.

Form roll-up. Based on the item-level ratings, the DOK distribution of the entire test is calculated by averaging across items. To ensure an accurate calculation, test items are weighted by the number of score points associated with each item (e.g., on a two-item test where the first item is a one-point item placed at DOK 1 and the second item is a two-point item placed at DOK 3, the DOK distribution for the test would be 33 percent DOK 1 and 67 percent DOK 3).

To reach a 0–1–2 score, the DOK distribution for the test form is compared to the recommended DOK distribution for the grade-level standards. The standards' DOK distribution is obtained in a parallel fashion—each standard is rated for DOK and the results are averaged across all grade-level standards to determine the proportion at each DOK level. In this case each standard is equally weighted.

TABLE C-1

Example DOK Index Calculation

	Standards	Test	Minimum
DOK 1	.25	.40	.25
DOK 2	.25	.40	.25
DOK 3	.25	.20	.20
DOK 4	.25	0	0
			sum = .70

The comparison between the DOK distribution of the test and that of the standards is based on two measures. First, the DOK distributions are compared by creating a DOK index, which is based on the proportional agreement between the test form and the standards in DOK distribution. Mathematically, this is calculated as the sum of the cell-by-cell minima between the two documents. For example, suppose the standards were coded as being 25 percent at each of the four DOK levels, and the test was coded as being 40 percent at DOK 1, 40 percent at DOK 2, and 20 percent at DOK 3. The DOK index would be .70 (25 percent from DOK 1, 25 percent from DOK 2, and 20 percent from DOK 3). (See Table C-1.)

Second, because a key problem of prior-generation assessments was their low overall DOK, the test is compared with the standards specifically on coverage of higher-level DOK (3+), with the goal of ensuring that the proportion of the test on DOK 3+ is not markedly lower than that of the standards.

Document review. The documentation review for B.4 focuses on the extent to which DOK is an explicit part of the test documentation. That is, there is a research-based definition of cognitive demand, a way of operationalizing cognitive demand at the item level, and a rationale for and specification of distribution of cognitive demand for each test form.

Criterion B.9
High-quality items and a variety of types are strategically used to appropriately assess the standard(s).

The next criterion, which is also evaluated for every item on each test form, focuses on overall item and test quality. Reviewers are asked to consider quality along multiple dimensions. For test form quality, the main focus is that a variety of item types are used (and that at least one of these item types requires students to construct, rather than select, a response). For item quality, reviewers are asked to consider issues, such as alignment to standards (whether the vendor's rating of alignment is accurate), editorial accuracy, and whether there are multiple correct answers.

Outcome review. For the item review, the first step is to determine the item type for each item. (In practice, the item type is almost always provided as part of test metadata, so no reviewer judgment is needed in this step.)

Second, reviewers are presented with all of the CCSS listed in the test's metadata as being measured by the item (as many as twelve standards, depending on the vendor). For each item, the reviewers indicate whether they agree with each of the alignment ratings (yes or no). They also are asked to list any additional on- or off-grade standards to which they believe the item is aligned.

Third, reviewers rate each item on six possible quality issues, indicating as many as apply. These are:

1 Item may not yield valid evidence of targeted skill.
2 Item has issues with readability.
3 Item is incorrectly keyed.[105]
4 Content is inaccurate.
5 Item has unintended correct answer.
6 Item has issues with editorial accuracy.

Reviewers are also given the option to indicate there is another problem not included in the six possible issues and provided a text box in which they can provide additional notes.

Form roll-up. These item-level ratings are rolled up to the whole test form for criteria B.9.1 and B.9.2. For B.9.1, the scoring is as follows:

2 – **Meets:** At least two item formats are used, including one that requires students to generate, rather than select, a response (i.e., constructed response, extended writing);

1 – **Partially Meets:** At least two formats (but not including constructed response) are used, including technology-based formats and/or two-part selected response formats; and

0 – **Does Not Meet:** Only a traditional multiple choice format is used. For B.9.2, the scoring is based on the proportion of items that exhibit high quality (none of the quality issues) and alignment to standards, with the cutoffs for receiving a 2 being 95 percent of items having technical quality and 90 percent reflecting alignment to standards.

Document review. The documentation criteria for B.9 mirror the item review criteria, with B.9.3 focusing on item types and B.9.4 focusing on technical quality.

Criterion B.3
Reading assessments consist of test questions or tasks, as appropriate, which demand that students read carefully and deeply and use specific evidence from increasingly complex texts to obtain and defend correct responses.

Criterion B.3 focuses on important dimensions of reading as called for in the CCSS, such as close reading, focusing on central ideas, and citing direct textual evidence. This criterion includes six sub-criteria (four relative to the item review, two relative to documentation).

Outcome review. To evaluate criterion B.3, reviewers are asked to make four judgments about each relevant item:

- First, whether the reading item aligns to the specifics of the reading standard. This rating uses the ratings reviewers made for each individual standard that they did under B.9. In particular, the judgment here is whether the item aligns to one or more reading standards other than standard R.1, which underlies all reading items. (In other words, if an item aligns to only standard R.1, it does not receive credit here. If an item aligns to any other reading standard than R.1, it does receive credit.)
- Second, whether the item requires close reading and analysis—that is, whether it requires the student to carefully read the text, rather than drawing on prior or background knowledge.

105. Some testing programs would not make their answer keys available to all reviewers due to confidentiality reasons. Thus, we removed from the evaluation the determination of "Item incorrectly keyed" in B.9.2 and C.5.2; our review does not verify the accuracy of the answer keys used by the testing programs.

- Third, whether the item focuses on central ideas and important particulars, rather than superficial or peripheral content.
- Fourth, whether the item requires direct use of textual evidence—that is, whether it requires students to provide or cite specific quotes or close paraphrases in order to answer the question.

Form roll-up. These four ratings are simply rolled up to the form level and compared against the tentative guidance, in order to reach score ratings. Sub-criterion B.3.1 is based on the proportion of test items that require close reading. Sub-criterion B.3.2 is based on the proportion of test items that focus on central ideas and important particulars. Sub-criterion B.3.3 is based on the proportion of test items that are aligned to the specifics of the standards. Each of these three sub-criteria are scores on the same scale:

2 – Meets: 90–100 percent;

1 – Partially Meets: 75–89 percent; and

0 – Does Not Meet: 0–74 percent.

The fourth sub-criterion, B.3.4, is based on the proportion of test items that require direct use of textual evidence. The thresholds for this sub-criterion are slightly lower: 51–100 percent, 33–50 percent, and 0–32 percent, respectively.

Document review. The two documentation sub-criteria for B.3 mirror the outcome sub-criteria. Sub-criterion B.3.5 considers the extent to which documentation indicates expectations for close reading, central ideas and important particulars, as well as standards alignment. Sub-criterion B.3.6 considers the extent to which documentation requires text dependency.

Criterion B.5
Assessments emphasize writing tasks that require students to engage in close reading and analysis of texts so that students can demonstrate college- and career-ready abilities.

This criterion focuses on writing. There are four sub-criteria. Sub-criteria B.5.1 and B.5.3 focus on the distribution of writing tasks across exposition, argument, and narrative types. Sub-criteria B.5.2 and B.5.4 focus on the extent to which writing tasks are text-based (that is, they focus on writing to texts).

Item review. The task of evaluating writing items is straightforward for reviewers. First, reviewers determine the type of writing that is called for by the item, choosing from among persuasive/argumentative, narrative, expository, and blended (i.e., a combination of two or more of these types). Second, reviewers indicate whether the writing task requires writing to a text (that is, to confront text or other stimuli directly, to draw on textual evidence, and to support valid inferences from text or stimuli).

Form roll-up. To roll-up to the whole form, reviewers evaluate whether the writing prompts are approximately balanced across the three types in order to rate criterion B.5.1. To evaluate criterion B.5.2, reviewers simply indicate the proportion of writing tasks that require writing to texts.

Document review. The documentation analysis for writing is completely analogous and simply requires evaluating the extent to which the documentation lays out appropriate expectations for the distribution of writing types and the proportion of writing tasks that require writing to texts.

Criterion B.6
The assessments require students to demonstrate proficiency in the use of language, including vocabulary and conventions.

The B.6 criterion focuses on language. This criterion has the largest number of sub-criteria at eight. These comprise four sets of two sub-criteria, each of which has one outcome criterion and one documentation sub-criterion. Criteria B.6.1 and B.6.5 consider vocabulary items and the extent to which they focus on Tier 2

words and phrases, require the use of context, and assess words important to central ideas. Criteria B.6.2 and B.6.6 consider the extent to which language items mirror real-world activities, focus on common errors, and emphasize the conventions most important for readiness. Criteria B.6.3 and B.6.7 consider the emphasis placed on vocabulary on the test (i.e., the proportion of test points), whereas criteria B.6.4 and B.6.8 consider the emphasis placed on language on the test.

Item review. Ratings on criterion B.6 depend on whether the item is a vocabulary item or a language item. If it is a vocabulary item, the reviewers rate it on three dimensions including whether: 1) the item tests a Tier 2 word or phrase; 2) the item tests a word central to understanding the text; and 3) the item requires the use of context. If the item is a language item, reviewers rate it on three separate dimensions: whether it mirrors real-world activities; whether it covers skills in the CCSS language progression skills chart (i.e., the skills considered most important for readiness); whether the item focuses on common student errors.

Form roll-up. The roll-up to the 2/1/0 score for the four outcome criteria is straightforward. For B.6.1, reviewers evaluate the joint distribution of the proportion of Tier 2 words and phrases and the proportion of items assessing words important to central ideas against the thresholds laid out in the tentative cutoffs. For B.6.2, the reviewers consider the proportion of items assessing language meet the three language dimensions. For B.6.3 and B.6.4, reviewers consider the proportion of total score points allocated to vocabulary and language, respectively, where the tentative cutoffs for each are as follows:

2 – Meets: Vocabulary (B.6.3)/language (B.6.4) is reported as a sub-score or $\geq$ 13 percent of score points;

1 – Partially Meets: 10–12 percent of score points; and

0 – Does Not Meet: 0–9 percent of score points.

Document review. The documentation criteria in B.6 exactly parallel the four outcome criteria just described.

Criterion B.7
The assessments require students to demonstrate research and inquiry skills, demonstrated by the ability to find, process, synthesize, organize, and use information from sources.

This criterion asks whether the assessments require students to demonstrate research and inquiry skills, demonstrated by the ability to find, process, synthesize, organize, and use information from sources.

Item review. In order to qualify as a research item, it must reference two or more reading passages. For those items that do reference two or more reading passages, there is only one rating to be made under this criterion: whether the item requires analysis, synthesis, and/or organization of information (mirroring real-world activities).

Form roll-up. To roll this sub-criterion up to the form-level for B.7.1, the index is the proportion of total research points that indeed require analysis, synthesis, and/or organization of information (mirroring real-world activities).

Document review. There are two documentation criteria for B.7. Documentation criterion B.7.3 mirrors B.7.1 and asks the proportion of research items that mirror real-world activities in the ways described. Documentation criterion B.7.2 simply asks whether a research score is reported or whether there is some other indication that the test vendor indicates that research is an important skill.

Criterion B.8
Over time, and as assessment advances allow, the assessments measure the speaking and listening communication skills students need for college and career readiness.

The final ELA/Literacy criterion—measuring speaking and listening—does not apply to the majority of existing assessments. It has four total sub-criteria—two for item review and two for documentation, and these are in matched pairs. Criteria B.8.1 and B.8.3 assess whether the listening "passages" reflect the criteria for passage quality from sections B.1 and B.2, as well as whether the items require active listening skills. Criteria B.8.2 and B.8.4 assess the quality of speaking tasks against several dimensions shown in the full scoring template.

Of the four tests, only Smarter Balanced currently assesses listening items, and none of the tests assess speaking. For these criteria, reviewers were allowed to provide an "insufficient evidence" rating, though as noted previously, these were ultimately scored as a 0 if the test did not offer items that meet them.

Item review. The B.8 criterion requires reviewers to make three ratings. First, raters indicate whether the listening passage meets the definition for high-quality reading passages laid out in criterion B.1. Second, they indicate whether the passage meets the definition for reading passage complexity laid out in criterion B.2. Finally, they rate whether the passage requires active listening skills.

Form roll-up. The roll-up to the form for B.8.1 considers the proportion of items that meet the criteria for high-quality items and active listening skills. The roll-up to the form for B.8.2 considers the proportion of speaking items that meet the criteria for high-quality speaking items.

Document review. The documentation review for B.8.3 and B.8.4 consider the extent to which the documentation indicates attention to the issues raised in B.8.1 and B.8.2.

Mathematics Criteria

There are five mathematics criteria, comprising six item review sub-criteria and seven documentation sub-criteria.

Criterion C.1
The assessments help educators keep students on track to readiness by focusing strongly on the content most needed in each grade or course for later mathematics.

The first mathematics criterion addresses the extent to which the assessment focuses on the content most needed for success in later mathematics (the "major work of the grade"). These are a set of content clusters in each grade's standards that the standards' creators have determined are the most central to success in future mathematics.

Criterion C.1 includes rating assessments' mathematical progressions—descriptions of the development of student understanding of a particular topic organized across grade levels. Because we were only evaluating assessments at grades 5 and 8, we could not consider the coherence of assessments' mathematical progressions across grades. This is an important area for future research using adjacent-grade assessments.

Item review. The task for a reviewer was in three steps:

1. verify the standards listed for each item by the test vendor;
2. add any additional standards that were tapped by the item; and
3. answer yes or no whether the item exclusively covered major work standards.

The index for this criterion is therefore the proportion of items that exclusively cover major work standards.

Document review. The documentation sub-criterion for C.1 focuses on the same issue, but using test blueprints. Again, the question is what proportion of the test content focuses on the major work of the grade.

Criterion C.2
The assessments measure conceptual understanding, fluency and procedural skill, and application of mathematics, as set out in college and career readiness standards.

The second mathematics criterion is based on the claim that the best mathematics assessments will measure a balance of conceptual understanding, procedural fluency, and application. Procedural skill items generally assess the extent to which test-takers can solve a purely mathematical item that has no or purely superficial context. Application items assess the extent to which test-takers can use problem context to choose skills or techniques to solve mathematical problems. Conceptual understanding items generally do not require procedures or

applications, though they may be made easier if students have those skills. Rather, conceptual understanding items typically test a sort of generalization from a particular example to a broader rule or class of problems.

Item review. The task for reviewers is to rate each mathematics item as covering one of these three types of knowledge (or a combination). Then, the form-level averages are compared to pre-established targets (with the goal of approximately an equal distribution across the three item types).

Document review. The two documentation criteria underneath C.2 also relate to the equal distribution of items across these three types. The first (C.2.2) simply asks whether item blueprints indicate there is an approximately equal distribution of items across these types. The second (C.2.3) asks whether all or nearly all students are likely to receive test forms that have an equal balance across the item types. This C.2.3 sub-criterion is designed to ensure that test forms are designed in such a way as to ensure that no form overly focuses on one item type at the expense of others. This may be especially relevant for computer-adaptive tests, where each student may take a unique form.

Criterion C.3
The assessments include brief questions and also longer questions that connect the most important mathematical content of the grade or course to mathematical practices, for example, modeling and making mathematical arguments.

The third mathematics criterion relates to the standards for mathematical practice. Specifically, this criterion seeks to ensure that each item that assesses a standard for mathematical practice (SMP) also assesses a content standard.

Item review. Reviewers simply rate each item that, according to the metadata, assesses an SMP for whether it also assesses a content standard. If the test vendor does not indicate standards for mathematical practice in the metadata, then this criterion cannot be addressed. The scoring guidelines suggest that 90 percent or more of items that assess a standard for mathematical practice also assess a content standard in order to award the highest score of 2.

Document review. The documentation criterion is the same, focusing instead on the extent to which the documentation indicates that items assessing mathematical practices also assess content standards.

Criterion C.4
The assessments require all students to demonstrate a range of higher-order analytical thinking skills in reading and writing based on the depth and complexity of college and career readiness standards, allowing robust information to be gathered for students with varied levels of achievement. Assessments include questions, tasks, and prompts about the basic content of the grade or course as well as questions that reflect the complex challenge of college- and career-ready standards.

As with ELA/Literacy criterion B.4, criterion C.4 focuses on the cognitive demand of the test items relative to the standards. The task for mathematics reviewers is exactly the same as for ELA/Literacy reviewers, with the only difference being that the DOK language is subject-specific for mathematics.

Criterion C.5
High-quality items and a variety of item types are strategically used to appropriately assess the standard(s).

The final four criteria for mathematics are under C.5: ensuring high-quality items and a variety of item types. Again, the procedures for C.5 are exactly parallel to those for B.9.

Appendix D

Author Biographies

Nancy Doorey, project manager and report co-author, has been deeply involved in educational reform for more than thirty years, serving as a teacher, state and district policymaker, program director, and consultant in the areas of assessment, teacher quality, and leadership. She has authored reports for several national organizations regarding advances in educational assessment, the six federally funded assessment consortia, and education governance. In 2009, Nancy co-led the creation of the Center for K–12 Assessment & Performance Management at the Educational Testing Service (ETS), which was charged with serving as a catalyst for advances in K–12 testing to support student learning, and authored its widely utilized series "Coming Together to Raise Achievement: New Assessments for the Common Core State Standards." As the director of programs, she formulated the agendas and managed five national research symposia and six webcasts regarding advances and challenges in K–12 assessment. Nancy received a master's degree in elementary education from Simmons College and a Certificate of Advanced Studies in computer science from Harvard University Extension; she also completed doctoral studies in educational leadership at Columbia University (ABD).

Morgan Polikoff, alignment expert and report co-author, is an assistant professor of education at the University of Southern California's Rossier School of Education. His areas of expertise include K–12 education policy; college and career readiness standards; assessment policy; alignment; and the measurement of classroom instruction. Recent work has investigated teachers' instructional responses to content standards and critiqued the design of school and teacher accountability systems. Ongoing work focuses on the implementation of college and career readiness standards and the influence of curriculum materials and assessments on implementation. He is an associate editor of *American Educational Research Journal* and serves on the editorial boards for *AERA Open* and *Educational Administration Quarterly*. His research is currently supported by the National Science Foundation, the WT Grant Foundation, and the Institute of Education Sciences, among other sources. Polikoff received his doctorate from the University of Pennsylvania's Graduate School of Education in 2010 with a focus on education policy and his bachelor's in mathematics and secondary education from the University of Illinois at Urbana-Champaign in 2006.

Appendix E

Review Panelist Biographies

ELA/Literacy and Mathematics Review Panelists

Philip M. Anderson is a Professor in the Department of Secondary Education at Queens College and also in the PhD program in Urban Education at the Graduate Center, City University of New York. Since earning a PhD from the University of Wisconsin–Madison's Department of Curriculum and Instruction, he has published widely on developmental aesthetics, literacy and the literature curriculum, and cultural theory in curriculum design. Dr. Anderson has also served as an expert reviewer for the Massachusetts Comprehensive Assessment System (MCAS) English Language Arts Assessment, grades 3–10, for more than a decade.

Elizabeth Angell is an English language arts teacher at Atlantic Middle School in Quincy, Massachusetts. She has been a middle school language arts teacher for the last fourteen years and also served as a reading interventionist. Eight years ago, Ms. Angell began working with the Massachusetts Department of Elementary and Secondary Education (in collaboration with Measured Progress) as a member of the Assessment Development Committee for Grade 8 ELA. Ms. Angell also served on the Grade 8 ELA Assessment Development Committee for the MCAS for the Massachusetts Department of Elementary and Secondary Education.

Melody Arabo is a third-grade teacher at Keith Elementary in Walled Lake Consolidated School District. She is a National Education Association (NEA) Master Teacher, a Michigan Educator Voice Fellow, Oakland County's Elite 40 under 40 Winner, and the 2015 Michigan Teacher of the Year. Mrs. Arabo is also the author of *Chaldean for Kids* and *Diary of a Real Bully* picture books, and served as an External Reviewer for ACT Aspire in 2014.

Jonathan Budd is K–12 Director of Curriculum, Instruction, and Assessments for Connecticut's Trumbull Public School District. His particular expertise is literacy, with a focus on text complexity. A classroom teacher of nineteen years prior to becoming a district administrator, Dr. Budd holds a BA in English and Music from Connecticut College, an MA in English from Trinity College (Hartford), a Sixth-Year Diploma in Educational Leadership from the University of Connecticut, and a PhD in English Education from Teachers College, Columbia University. Dr. Budd also served as a content-area expert on SBAC's Range-Finding Committee and Pre-Range-Finding Committee for ELA in 2013 and as a content-area expert on two Item Review Committees for Measured Progress in 2013.

Amber Chandler is a National Board English language arts teacher and has taught at the Frontier Central School District in New York for the last thirteen years. A 2014 New York Educator Voice Fellow, Ms. Chandler is an active educational blogger, certified school building leader, and serves in a staff developer role through her local teacher center on topics such as "Danielson's Domains," "Preventing Plagiarism Using the I-Search Paper," and "Differentiation." Ms. Chandler has also served as director of Frontier Central Summer School. Ms. Chandler has also worked with the New York Department of Education as a final eyes reviewer and test item developer.

Wally Etterbeek is an emeritus professor of mathematics at California State University Sacramento, where he was a faculty member for thirty-five years and department chair for twelve. During his early tenure, Dr. Etterbeek taught grades 2–6 for Project SEED, and more recently, taught gifted high school students in the San Juan District and Higher Level International Baccalaureate Mathematics. For four decades, Dr. Etterbeek has also been a

participating member and statistician for the Mathematics Diagnostic Texting Program (MDTP), a joint University of California/CSU project that develops readiness tests for grade 7 through first-year college mathematics courses. Dr. Etterbeek has also served as a member of UCLA's National Center for Research on Evaluation, Standards, and Student Testing (CRESST) panel that reviewed the alignment of the ACT Aspire Quality Core Algebra I and Geometry I courses and the Cambridge mathematics courses for grades 9–10 to the CCSS.

Shannon Garrison has been a fourth- and fifth-grade teacher at Solano Avenue Elementary School in downtown Los Angeles for the past eighteen years. She is a National Board Certified Teacher and was honored as a 2008 Milken Educator in California. In 2010, she was appointed to the National Assessment Governing Board, which sets policy for the National Assessment of Educational Progress (also known as the Nation's Report Card). Ms. Garrison also serves as the Chair of the Assessment Development Committee and was just reappointed for a second term by Secretary Duncan.

Jocelyn Gukeisen has been a middle school math educator in independent schools for fifteen years. Ms. Gukeisen currently serves as the Middle School Director and Director of Studies at The McGillis School in Salt Lake City where she coordinates and facilitates curriculum development across school divisions and across disciplines, and oversees the development of various programs and pedagogical strategies, as well as implementation of standardized testing. She has also served as an external item writer and consultant for ACT Aspire since June 2004.

Joey Hawkins was a middle school humanities public school teacher for over twenty years. A founding member of the Vermont Writing Collaborative and co-author of *Writing for Understanding: Using Backward Design to Help All Students Write Effectively*, she now provides professional development for K–12 teachers, focusing on the integration of literacy into content-based thinking and writing.

Roger Howe is the William Kenan Jr. Professor of Mathematics at Yale University. He has been teaching and conducting research in the Mathematics Department at Yale University for over forty years. His research interests focus on applications of symmetry, including representation theory, automorphic forms, harmonic analysis, and invariant theory. Dr. Howe has held visiting positions at many universities and research institutes in the United States, Europe, and Asia. In 1997–98, he served as a Phi Beta Kappa Visiting Scholar. He has served on many committees, including the NRC Study Committee for *Adding It Up*. He is currently a member of the U.S. National Commission on Mathematics Instruction and the Executive Committee of the International Commission on Mathematics Instruction. He is also a member of the American Academy of Arts and Sciences and the National Academy of Sciences.

Melisa R. Howey is an Education Specialist for Area Cooperative Educational Services (ACES) in Hamden, CT. She graduated from the University of Connecticut with a degree in Mathematics, received a Master's in Education from Quinnipiac University, and a sixth-year degree in Administration from Sacred Heart University. Ms. Howey has previously taught middle school math in both West Hartford and East Hartford, CT, and served as a summer school principal, interim math department head, and the K–6 district math specialist in East Hartford, CT. She also was part of a team who created grade 5 performance tasks for the Smarter Balanced Assessment.

Patricia Ford Kavanagh has been a mathematics teacher at Manchester-Essex Regional School District, Massachusetts since 1982. From 2008–2014, she developed and reviewed test items on the eighth-grade Math Assessment Development Team for the Massachusetts Department of Elementary and Secondary Education.

Bowen Kerins is a senior curriculum designer at Education Development Center (EDC). He develops current techniques and procedures used in the design, development, and implementation of curriculum, curriculum-based professional development, instruction, and assessments. Before joining EDC, Mr. Kerins was a high school mathematics teacher in Massachusetts. He has a BS in mathematics from Stanford University and an MA in teaching secondary mathematics from Boston University. Mr. Kerins also reviewed publicly released items for PARCC from 2012–2014.

David Kirshner is a Professor of Mathematics Education and co-Director of the Gordon A. Cain Center for STEM Literacy at Louisiana State University. He holds a doctoral degree in mathematics education and a master's degree in mathematics, both from University of British Columbia. Dr. Kirshner has served as president of the

North American Chapter of the International Group for the Psychology of Mathematics Education (PME-NA) and Chair of the Editorial Board of the Journal for Research in Mathematics Education. He recently co-edited the third edition of the *Handbook of International Research in Mathematics Education*.

Shelli Klein is a freelance consultant in educational assessment. She has more than twenty-five years of experience in education and educational publishing. As a former Publishing Director for an assessment company, she led large-scale assessment projects, developed and analyzed test blueprints, and trained the English language arts staff in interpreting and implementing the Common Core State Standards. In addition, Ms. Klein has experience teaching multiple grade levels, from elementary school through community college. She has also performed senior reviews of ELA field-test items for both the PARCC and the Smarter Balanced consortia, focusing on alignment of items to the specifications and standards in 2013–14.

Lianne Markus is a fourth- and fifth-grade ELA/Math teacher at Hope Township School, Hope, New Jersey. She received her undergraduate degree in journalism from the University of South Carolina and her NJ Teaching Certificate from Centenary College. As a member of the PARCC NJ Educator Leader Cadre, she represents the consortium at local, state, and national events. Ms. Markus participated in the PARCC Communications Boot Camp and served as a voiceover for Advance Illinois PARCC 101 Explainer Video, and also worked on the PARCC bias and sensitivity state educator review panel. She was nominated as NJ Governor's Teacher/Educational Professional of the Year in 2011, and nominated as Teacher of the Year 2010, 2012, 2013, 2014, and 2015.

Joe McGonegal has taught high school English and college composition for fifteen years in public and private settings in Wisconsin, Florida, New Hampshire, and Massachusetts before returning to school to study journalism. Since 2013, he has been a writer and strategist for MIT Technology Review and the MIT Alumni Association, where he currently serves as director of alumni education. Mr. McGonegal has also consulted on the ACT Aspire reading assessment since 2007.

Jennifer McPartland is a reading specialist at Gordon W. Mitchell Middle School in East Bridgewater, Massachusetts. She has worked in public education for seventeen years. She began her career teaching fifth grade, and then moved to serve as the school's reading specialist, where she has spent the last ten years. In Massachusetts, Ms. McPartland holds certifications in Elementary Education, Reading, and Administration. Since 2007, she has served as a member of the Massachusetts Department of Elementary and Secondary Education's MCAS Assessment Development Committee (ADC) for grade 6 ELA.

Robert Noreen is a professor and former Chair of the Department of English at California State University (CSU), Northridge. Previously, he was the Chair of the California English Placement Test Development Committee (1989–2003) and CSU English Assessment Program. Dr. Noreen was a member of the U.S. Department of Education committee that reviewed early drafts of the Common Core assessments in English. He has also served as a scoring leader and/or as a member of the item development team for the Graduate Management Admission Test (GMAT), National Teacher Examination (NTE), AP-English exam, Test of Written English/Test of English as a Foreign Language (TWE/TOEFL), and essay portion of the American Medical Colleges Admission Test (AMCAT).

Lynne Olmos is a National Board-certified teacher currently teaching English and drama in rural Mossyrock, Washington. Mrs. Olmos specializes in assessment preparation and creation, having served on district, state, and national committees to create, review, range-find and score assessment materials, for both students and teachers. Beginning in 2012, she worked on Washington State-managed committees for item writing and passage selection for Smarter Balanced. In the fall of 2014, Mrs. Olmos served as an item reviewer during Smarter Balanced's online achievement level setting. Her most recent projects involve art and the Common Core State Standards, including membership in Washington's Arts Cadre for arts educators concerned with standards and overall student achievement.

Charles Perfetti is the Distinguished University Professor of Psychology and Director of the Learning Research and Development Center at the University of Pittsburgh. Dr. Perfetti's research has addressed components of reading skill—reading comprehension, word identification, word learning, learning from multiple texts—and comparisons of reading across writing systems. His work has been published in several books and over 200

articles and book chapters. He has also received the Distinguished Scientific Contribution Award from both the Society for the Scientific Study of Reading and the Society for Text and Discourse.

Michael Roach is a visiting clinical assistant professor in mathematics education at Indiana University-Purdue University Indianapolis. He has worked as a high school mathematics teacher, a mathematics specialist at the Indiana Department of Education (IDOE), and a mathematics coach for Indianapolis Public Schools. Dr. Roach has a wide variety of experiences with large-scale assessment, including coordinating the development of an online testing program while at the IDOE and conducting secondary analyses of National Assessment of Educational Progress data.

Wilfried Schmid is the Dwight Parker Robinson Professor of Mathematics at Harvard University. He received his BA from Princeton University in 1964 and his PhD in 1967 from the University of California, Berkeley. Prior to joining the Harvard faculty in 1978, Dr. Schmid taught at Berkeley and at Columbia University. He helped to revise the Mathematics Curriculum Framework for Massachusetts; served as Mathematics Advisor to the Massachusetts Department of Elementary and Secondary Education; as member of the Steering Committee of Mathematics NAEP; as member of the Program Committee of the International Congress of Mathematics Education 2004; and as member of the National Mathematics Advisory Panel. Dr. Schmid assisted with the final review of MCAS mathematics test items from 2003–2004, and participated in the final review of mathematics SAT test items in 2008–09.

Colin Starr is a Professor of Mathematics at Willamette University, where he has taught since 2003. Dr. Starr has been involved in numerous projects that combine mathematics and teaching, including the Algebraic Thinking Project and serving as a co-principal investigator on two National Science Foundation Research Experiences for Undergraduates grants. He has reviewed MCAS items since 2005 as an expert reviewer under work facilitated by Measured Progress and worked with various other groups examining the teaching of mathematics.

Rose Switalski is currently an eighth-grade math teacher at an art integrated charter school in Las Vegas. She completed her bachelor's degree in secondary mathematics education at the University of Nevada. Over the past eight years, she has taught grades 6–12.

Erin Thompson is a secondary English language arts assessment specialist at the Indiana Department of Education (IDOE), where she has worked for the past five years. Prior to working at the IDOE, Ms. Thompson taught English/language arts at grades 6–9 for nine years. She has also participated in PARCC's Operational Work Group for English language arts, where she was involved in passage and assessment review, policy decisions, and creation of performance level descriptors.

Howard Tuttman is a K–8 Instructional Math Coach in Everett, Massachusetts, an urban district just north of Boston. For several years, he has served on the Massachusetts Department of Elementary and Secondary Education's Assessment Development Committee (ADC), reviewing fourth-grade math items for the Massachusetts Comprehensive Assessment System (MCAS). He has also served on the PARCC's Math Item Review Committee, reviewing questions for the grades 3–5, and participated in PARCC's Mathematics Performance Level Setting Committee for grades 5–6.

Tad Watanabe is a Professor of Mathematics Education at Kennesaw State University (KSU), where he has taught mathematics content courses for prospective elementary school teachers since 2006. Prior to coming to KSU, he taught four years at the Pennsylvania State University and ten years at Towson University in Towson, Maryland. Dr. Watanabe is interested in various mathematics education practices in Japan, and has written several articles on Japanese lesson study and elementary school mathematics curriculum. He has also worked with numerous lesson study groups throughout the United States.

Rosalynn Wolfe is a graduate student at the University of Iowa, where she is studying Educational Measurement and Statistics. Prior to this Rosalynn was a math teacher for ten years, mostly grades 6–12, in Alaska, Arizona, and Iowa. Ms. Wolfe has worked on curriculum development for Tucson Unified School District and several charter schools in Arizona. She has also reviewed and written items for alignment with academic standards for ACT Aspire for the past three years.

Christopher Yakes began a position as Associate Professor of Mathematics at Salem State University in September 2015. For the previous nine years, he was a faculty member at California State University (Chico) where he taught mathematics courses to undergraduate mathematics majors and pre-service teachers. While at Chico State, he also worked extensively with in-service teachers through the California Mathematics Project and other grant-funded projects, which included holding professional development workshops, facilitating lesson study groups, and serving as a math coach. For the past three years, Dr. Yakes has also worked with the California Department of Education on rewriting the California Mathematics Framework for Public Schools.

Amy Youngblood is the founder of EduOptimus, which provides professional development to school districts in Missouri and across the Midwest. She began her career as a kindergarten teacher in Southwest Missouri, and has been an elementary teacher, a K–8 gifted teacher, a 9–12 gifted teacher, federal programs director, Missouri School Improvement Program (MSIP) Coordinator, and a K–12 curriculum director. Additionally, Ms. Youngblood has served as the President for the Missouri Affiliate of the Association for Supervision and Curriculum Development (ASCD), a member of the ASCD Leadership Council, a member of the Missouri Expert Curriculum Review Team for MSIP, and a member of the Staff Development Leadership Council. Currently, she is a Lead Facilitator for EdReports and a member of the Educators Evaluating the Quality of Instructional Products (EQuIP) Peer Review Panel, which was created by Achieve to identify high-quality instructional materials aligned to the Common Core State Standards.

Generalizability Review Panelists

Heather Goodwin-Nelson is a Special Education Instructional Facilitator for Utah Virtual Academy. She received her BS in Elementary Education from Brigham Young University-Hawaii in 1994. During her undergraduate work, Ms. Goodwin-Nelson taught various grade levels before receiving a master's degree in Special Education from Brigham Young University.

Roger Howe is the William Kenan Jr. Professor of Mathematics at Yale University. He has been teaching and conducting research in the Mathematics Department at Yale University for over forty years. His research interests focus on applications of symmetry, including representation theory, automorphic forms, harmonic analysis, and invariant theory. Dr. Howe has held visiting positions at many universities and research institutes in the U.S., Europe, and Asia. In 1997–98, he served as a Phi Beta Kappa Visiting Scholar. He has served on many committees, including the NRC Study Committee for *Adding It Up*. He is currently a member of the U.S. National Commission on Mathematics Instruction and the Executive Committee of the International Commission on Mathematics Instruction. He is also a member of the American Academy of Arts and Sciences and the National Academy of Sciences.

Shelli Klein is a freelance consultant in educational assessment. She has more than twenty-five years of experience in education and educational publishing. As a former publishing director for an assessment company, she led large-scale assessment projects, developed and analyzed test blueprints, and trained the English language arts staff in interpreting and implementing the Common Core State Standards. In addition, Ms. Klein has experience teaching multiple grade levels, from elementary school through community college. She has also performed senior reviews of ELA field-test items for both the PARCC and the Smarter Balanced consortia, focusing on alignment of items to the specifications and standards in 2013–14.

Tim LaVan is a math teacher in the Oil City Area School District (Pennsylvania) in his twenty-fifth year of teaching. He is highly experienced in current techniques and procedures used in the design, development, and implementation of curriculum, instruction, and assessments, being a member of the Mathematics Advisory Committee for Pennsylvania. Dr. LaVan has been a member of PARCC advisory committees representing Pennsylvania as well. He holds a BS in mathematics education from Clarion University of PA, an MS in Computational Science from Michigan State University, and completed doctoral work in Curriculum and Instruction at the University of Pittsburgh.

Robert Noreen is a professor and former Chair of the Department of English at California State University (CSU), Northridge. Previously, he was the Chair of the California English Placement Test Development Committee (1989–2003) and CSU English Assessment Program. Dr. Noreen was a member of the U.S. Department of

Education committee that reviewed early drafts of the Common Core assessments in English. He has also served as a scoring leader and/or as a member of the item development team for the Graduate Management Admission Test (GMAT), National Teacher Examination (NTE), AP-English exam, Test of Written English/Test of English as a Foreign Language (TWE/TOEFL), and essay portion of the American Medical Colleges Admission Test (AMCAT).

Tabitha Pacheco is a National Board-certified teacher in exceptional needs, a 2015 National Teaching Fellow for the Hope Street Group, and serves on the Practitioners Advisory Group for The Centers on Great Teachers and Leaders. For the past nine years, Ms. Pacheco has worked with students with disabilities and is an expert on accommodating and scaffolding the CCSS to meet the needs of all learners. She has a bachelor's degree from Brigham Young University in Family Life and a post-baccalaureate degree in special education from Brigham Young University.

Michael Roach is a visiting clinical assistant professor in mathematics education at Indiana University-Purdue University Indianapolis. He has worked as a high school mathematics teacher, a mathematics specialist at the Indiana Department of Education (IDOE), and a mathematics coach for Indianapolis Public Schools. Dr. Roach has a wide variety of experiences with large-scale assessment, including coordinating the development of an online testing program while at the IDOE and conducting secondary analyses of National Assessment of Educational Progress data.

Charlie Wayne has been an assessment specialist in mathematics with the Pennsylvania Department of Education (PDE) for over seventeen years. He has been Pennsylvania's mathematics representative with PARCC (Pennsylvania is a participating state in PARCC and Smarter Balanced). He has also participated in various alignment studies with Dr. Norman Webb, Achieve, the American Association for the Advancement of Science (AAAS), and HumRRO. Mr. Wayne has a BS in Economics and an MS in Mathematics along with a graduate certificate in Large-Scaled Assessment Education. Prior to coming to PDE, he taught in elementary and middle schools, at the post-secondary level, and at a business institute.

Appendix F

Full ELA/Literacy and Math Ratings and Summary Statements (Grades 5 and 8)

OVERALL SUMMARY STATEMENTS FOR ELA/LITERACY	
Program	**Summary Statements**
ACT Aspire	In ELA/Literacy, ACT Aspire receives a Limited to Good Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. The combined set of ELA/Literacy tests (reading, writing, and English) requires close reading and adequately evaluates language skills. More emphasis on assessment of writing to sources, vocabulary, and research and inquiry, as well as increasing the cognitive demands of test items, will move the assessment closer to fully meeting the criteria. Over time, the program would also benefit by developing the capacity to assess speaking and listening skills.
MCAS	In ELA/Literacy, MCAS receives a Limited to Good Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. The test requires students to closely read high-quality texts and a variety of high-quality item types. However, MCAS does not adequately assess several critical skills, including reading informational texts, writing to sources, language skills, and research and inquiry; further, too few items assess higher-order skills. Addressing these limitations would enhance the ability of the test to signal whether students are demonstrating the skills called for in the standards. Over time, the program would also benefit by developing the capacity to assess speaking and listening skills.
PARCC	In ELA/Literacy, PARCC receives an Excellent Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. The tests include suitably complex texts, require a range of cognitive demand, and demonstrate variety in item types. The assessments require close reading, assess writing to sources, research, and inquiry, and emphasize vocabulary and language skills. The program would benefit from the use of more research tasks requiring students to use multiple sources, and, over time, developing the capacity to assess speaking and listening skills.
Smarter Balanced	In ELA/Literacy, Smarter Balanced receives a Good to Excellent Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. The tests assess the most important ELA/Literacy skills of the CCSS, using technology in ways that both mirror real-world uses and provide quality measurement of targeted skills. The program is most successful in its assessment of writing and research and inquiry. It also assesses listening with high-quality items that require active listening, which is unique among the four programs. The program would benefit by improving its vocabulary items, increasing the cognitive demand in grade 5 items, and, over time, developing the capacity to assess speaking skills.

ELA/LITERACY CONTENT CRITERIA OVERVIEW

Criteria	Program and Rating	Summary Statements
I. Assesses the content most needed for College and Career Readiness.	L **ACT Aspire** LIMITED/UNEVEN MATCH	ACT Aspire receives a Limited/Uneven Match to the CCSSO Criteria for Content in ELA/Literacy. The assessment program includes an emphasis on close reading and language skills. However, the reading items fall short on requiring students to cite specific textual information in support of a conclusion, generalization, or inference and in requiring analysis of what has been read. In order to meet the criteria, assessing writing to sources, vocabulary, and research and inquiry need to be strengthened.
	L **MCAS** LIMITED/UNEVEN MATCH	MCAS receives a Limited/Uneven Match to the CCSSO Criteria for Content in ELA/Literacy. The assessment requires students to read closely well-chosen texts and presents test questions of high technical quality. However, the program would be strengthened by assessing writing annually, assessing the three types of writing called for across each grade band, requiring writing to sources, and placing greater emphasis on assessing research and language skills.
	E **PARCC** EXCELLENT MATCH	PARCC receives an Excellent Match to the CCSSO Criteria for Content in ELA/Literacy. The program demonstrates excellence in the assessment of close reading, vocabulary, writing to sources, and language, providing a high-quality measure of ELA content, as reflected in college and career ready standards. The tests could be strengthened by the addition of research tasks that require students to use two or more sources and, as technologies allow, a listening and speaking component.
	E **Smarter Balanced** EXCELLENT MATCH	Smarter Balanced receives an Excellent Match to the CCSSO Criteria for Content in ELA/Literacy. The program demonstrates excellence in the areas of close reading, writing to sources, research, and language. The listening component represents an important step toward adequately measuring speaking and listening skills—a goal specifically reflected in the standards. Overall, Smarter Balanced is a high-quality measure of the content required in ELA and literacy, as reflected in college and career readiness standards. A greater emphasis on Tier 2 vocabulary would further strengthen these assessments relative to the criteria.

ELA/LITERACY CONTENT CRITERIA

Criteria	Program and Rating	Summary Statements
B.3 Reading: Tests require students to read closely and use specific evidence from texts to obtain and defend correct responses.	L **ACT Aspire** LIMITED/UNEVEN MATCH	On "requiring students to read closely and use evidence from texts," the rating is Limited/Uneven Match. Although most reading items require close reading of some kind, too many can be answered without analysis of what was read. Items that purport to require specific evidence from text often require only recall of information from text. To meet this criterion, the test items should require students to cite specific text information in support of some conclusion, generalization, or inference drawn from the text.
	G **MCAS** GOOD MATCH	On "requiring students to read closely and use evidence from texts," the rating is Good Match. Most reading items require close reading and focus on central ideas and important particulars. Some questions, however, do not require the students to provide direct textual evidence to support their responses. In addition, too many items do not align closely to the specifics of the standards.
	E **PARCC** EXCELLENT MATCH	On "requiring students to read closely and use evidence from texts," the rating is Excellent Match. Nearly all reading items require close reading, the understanding of central ideas, and the use of direct textual evidence.
	E **Smarter Balanced** EXCELLENT MATCH	On "requiring students to read closely and use evidence from texts," the rating is Excellent Match. Nearly all reading items align to the reading standards requiring close reading, the understanding of central ideas, and use of direct textual evidence in support of a conclusion, generalization, or inference.

ELA/LITERACY CONTENT CRITERIA

Criteria	Program and Rating	Summary Statements
B.5 Writing: Tasks require students to engage in close reading and analysis of texts. Across each grade band, tests include a balance of expository, persuasive/ argument, and narrative writing.	L **ACT Aspire** LIMITED/UNEVEN MATCH	On "assessing writing," the rating is Limited/Uneven. Although the program documentation shows that a balance of all three writing types is required across each grade band, the writing prompts do not require writing to sources. As a result, the program insufficiently assesses the types of writing required by college and career readiness standards.
	W **MCAS** WEAK MATCH	On "assessing writing," the rating is Weak Match. Writing is assessed at only one grade level per band, and there is insufficient opportunity to assess writing of multiple types. In addition, the writing assessments do not require students to use sources. As a result, the program inadequately assesses the types of writing required by college and career readiness standards.
	E **PARCC** EXCELLENT MATCH	On "assessing writing," the rating is Excellent Match. The assessment meets the writing criterion, which requires writing to sources. Program documentation shows that a balance of all three writing types is required across each grade band.
	E **Smarter Balanced** EXCELLENT MATCH	On "assessing writing," the rating is Excellent Match. The writing items are of high quality, and the writing prompts all require the use of textual evidence. Program documentation shows that a balance of all three writing types is required across each grade band.
B.6 Vocabulary and language skills: Tests place sufficient emphasis on academic vocabulary and language conventions as used in real-world activities.	G **ACT Aspire** GOOD MATCH	On "emphasizing vocabulary and language skills," the rating is Good Match. Language items meet the criterion for being tested within writing activities, though more items are needed that are embedded in real world tasks such as editing. The vocabulary items do not meet the criterion because there are too few of them and not enough assess Tier 2 words.
	L **MCAS** LIMITED/UNEVEN MATCH	On "emphasizing vocabulary and language skills," the rating is Limited/Uneven Match. Vocabulary items are sufficient and generally aligned to the criterion; however, the grade 5 items need more words at the Tier 2 level. Furthermore, a lack of program documentation means that the quality of vocabulary assessments cannot be substantiated across forms. MCAS does not meet the criterion for assessing language skills, which call for them to be assessed within writing assessments that mirror real-world activities including editing and revision.
	E **PARCC** EXCELLENT MATCH	On "emphasizing vocabulary and language skills," the rating is Excellent Match. The test contains an adequate number of high-quality items for both language use and Tier 2 vocabulary and awards sufficient score points, according to the program's documentation, to both of these areas.
	G **Smarter Balanced** GOOD MATCH	On "emphasizing vocabulary and language skills," the rating is Good Match. Language skill items are contained in a sub-score and meet the criterion for being assessed within writing and mirroring real-world activities such as editing and revision. The number of items that test vocabulary is a bit low; further, items coded as vocabulary too often did not test Tier 2 vocabulary words.

ELA/LITERACY CONTENT CRITERIA

Criteria	Program and Rating	Summary Statements
B.7 Research and inquiry: Assessments require students to demonstrate the ability to find, process, synthesize, and organize information from multiple sources.	L **ACT Aspire** LIMITED/UNEVEN MATCH	On "assessing research and inquiry," the rating is Limited/Uneven Match. Although the one item at each grade level involving research and inquiry did indeed require analysis and organization of information, this single item is insufficient to provide a quality measure of research and inquiry.
	W **MCAS** WEAK MATCH	On "assessing research and inquiry," the rating is Weak Match. The assessment has no test questions devoted to research.
	E **PARCC** EXCELLENT MATCH	On "assessing research and inquiry," the rating is Excellent Match. The research items require analysis, synthesis, and/or organization and the use of multiple sources, therefore meeting the criterion for Excellent.
	E **Smarter Balanced** EXCELLENT MATCH	On "assessing research and inquiry," the rating is Excellent Match. The research items require analysis, synthesis, and/or organization, and the use of multiple sources.
B.8 Speaking and listening: Over time, and as assessment advances allow, the assessments measure speaking and listening communication skills.	W **ACT Aspire** WEAK MATCH	On assessing "speaking and listening," the rating is Weak Match. The program does not assess speaking or listening at this time. Because this criterion is to be met "over time, as assessment advances allow," this rating is not included in the overall rating for Content.
	W **MCAS** WEAK MATCH	On assessing "speaking and listening," the rating is Weak Match. The program does not assess speaking or listening at this time. Because this criterion is to be met "over time, as assessment advances allow," this rating is not included in the overall rating for Content.
	W **PARCC** WEAK MATCH	On assessing "speaking and listening," the rating is Weak Match. The program does not assess speaking or listening at this time. Because this criterion is to be met "over time, as assessment advances allow," this rating is not included in the overall rating for Content.
	L **Smarter Balanced** LIMITED/UNEVEN MATCH	On assessing "speaking and listening," the rating is Limited/Uneven Match. Listening is tested with high-quality items that assess active listening skills. Speaking is not assessed. Because this criterion is to be met "over time, as assessment advances allow," this rating is not included in the overall rating for Content.

ELA/LITERACY DEPTH CRITERIA OVERVIEW

Criteria	Program and Rating	Summary Statements
II. Assesses the depth that reflects the demands of College and Career Readiness.	G **ACT Aspire** GOOD MATCH	ACT Aspire receives a rating of Good Match for Depth in ELA/Literacy. The program's assessments are built on high-quality test items and texts that are suitably complex. To fully meet the CCSSO Criteria, at both grade levels more cognitively demanding test items are needed at both grade levels, as well as additional literary narrative text, as opposed to literary informational texts.
	G **MCAS** GOOD MATCH	MCAS receives a rating of Good Match for Depth. The assessments do an excellent job in presenting a range of complex reading texts. To fully meet the demands of the CCSSO Criteria, however, the test needs more items at higher levels of cognitive demand, a greater variety of items to test writing to sources and research, and more informational texts, particularly those of an expository nature.
	E **PARCC** EXCELLENT MATCH	PARCC receives a rating of Excellent Match for Depth in ELA/Literacy. The PARCC assessments meet or exceed the depth and complexity required by the criteria through a variety of item types that are generally high quality. A better balance between literary and informational texts would further strengthen the assessments in addressing the criteria.
	G **Smarter Balanced** GOOD MATCH	Smarter Balanced receives a rating of Good Match for Depth in ELA/Literacy. The assessments use a variety of item types to assess student reading and writing to sources. The program could better meet the depth criteria by increasing the cognitive demands of the grade 5 assessment and ensuring that all items meet high editorial and technical quality standards.

ELA/LITERACY DEPTH CRITERIA

Criteria	Program and Rating	Summary Statements
B.1 Text quality and types: Tests include an aligned balance of high-quality literary and informational texts.	G **ACT Aspire** GOOD MATCH	On "the balance of high-quality literary and informational texts," the rating is Good Match. The texts are of high quality, and the proportion of informational texts meets the criterion. The assessment would better align to the criterion, however, with additional literary narrative text, as opposed to literary informational text.
	G **MCAS** GOOD MATCH	On "the balance of high-quality literary and informational texts," the rating is Good Match. The quality of the texts is very high. Regarding the balance of text types, some forms had too few informational texts.
	G **PARCC** GOOD MATCH	On "the balance of high-quality literary and informational texts," the rating is Good Match. Although the passages are consistently of high quality, the tests would have better reflected the criterion with additional literary nonfiction passages.
	E **Smarter Balanced** EXCELLENT MATCH	On "the balance of high-quality literary and informational texts," the rating is Excellent Match. Overall text quality is high, and among informational texts there is a high proportion of expository text types.

ELA/LITERACY DEPTH CRITERIA

Criteria	Program and Rating	Summary Statements
B.2 Complexity of texts: Test passages are at appropriate levels of text complexity, increasing through the grades, and multiple forms of authentic, high-quality texts are used.	G **ACT Aspire** GOOD MATCH	The rating for "text complexity" is Good Match. It is based solely on the review of program documentation, which is determined to have met the criterion. The test blueprints and other documents clearly and explicitly require texts to increase in complexity grade-by-grade and for texts to be placed in grade bands and grade levels based on appropriate quantitative and qualitative data.
	G **MCAS** GOOD MATCH	The rating for "use of appropriate levels of text complexity" is Good Match. It is based solely on the review of program documentation, which is determined to have met the criterion. The test blueprints and other documents clearly and explicitly require texts to increase in complexity grade-by-grade and for texts to be placed in grade bands and grade levels based on appropriate quantitative and qualitative data.
	G **PARCC** GOOD MATCH	The rating for "use of appropriate levels of text complexity" is Good Match. It is based solely on the review of program documentation, which is determined to have met the criterion. The test blueprints and other documents clearly and explicitly require texts to increase in complexity grade-by-grade and for texts to be placed in grade bands and grade levels based on appropriate quantitative and qualitative data.
	G **Smarter Balanced** GOOD MATCH	The rating for "use of appropriate levels of text complexity" is Good Match. It is based solely on the review of program documentation, which is determined to have met the criterion. The test blueprints and other documents clearly and explicitly require texts to increase in complexity grade-by-grade and for texts to be placed in grade bands and grade levels based on appropriate quantitative and qualitative data.
B.4 Cognitive demand: The distribution of cognitive demand for each grade level is sufficient to assess the depth and complexity of the standards.	W **ACT Aspire** WEAK MATCH	On "requiring a range of cognitive demand," the rating is Weak Match. To better reflect the depth and complexity of the standards, both grade-level tests should require more items with higher cognitive demands, although this problem is greater at grade 8.
	L **MCAS** LIMITED/UNEVEN MATCH	On "requiring a range of cognitive demand," the rating is Limited/Uneven Match. More items that measure the higher levels of cognitive demand are needed to sufficiently assess the depth and complexity of the standards.
	E **PARCC** EXCELLENT MATCH	On "requiring a range of cognitive demand," the rating is Excellent Match. The test is challenging overall; indeed the cognitive demand of the grade 8 test exceeds that of the CCSS.
	G **Smarter Balanced** GOOD MATCH	On "requiring a range of cognitive demand," the rating is Good Match. The cognitive demand of items cover a sufficient range and, in grade 8, the percentage of more demanding items (DOK 3 and 4) correspond well to the demand of the standards. However, the grade 5 test needs more items at higher levels of cognitive demand to reflect fully the depth and complexity of the standards.

ELA/LITERACY DEPTH CRITERIA

Criteria	Program and Rating	Summary Statements
B.9 High-quality items and variety of item types: Items are of high technical and editorial quality and each test form includes at least two items types, including at least one that requires students to generate rather than select a response.	E **ACT Aspire** EXCELLENT MATCH	On "ensuring high-quality items and a variety of item types," the rating is Excellent Match. The test includes items that exhibit high technical quality and editorial accuracy. Multiple item formats are used, including student-constructed responses.
	E **MCAS** EXCELLENT MATCH	On "ensuring high-quality items and a variety of item types," the rating is Excellent Match. Multiple item formats are used, including student-generated response items. The items exhibit high technical quality and editorial accuracy. The paper-and-pencil format precludes the use of technology-enhanced items, but the criterion for multiple item types is met.
	E **PARCC** EXCELLENT MATCH	On "ensuring high-quality items and a variety of item types," the rating is Excellent Match. The tests use multiple item formats, including student-constructed responses.
	G **Smarter Balanced** GOOD MATCH	On "ensuring high-quality items and a variety of item types," the rating is Good Match. The tests use multiple formats and technology-enhanced items including constructed responses. However, editorial or technical issues, including readability, were noted in a number of items.

OVERALL SUMMARY STATEMENTS FOR MATHEMATICS

Program	Summary Statements
ACT Aspire	In mathematics, ACT Aspire receives a Limited/Uneven to Good Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. Some of the mismatch with the criteria is likely due to intentional program design, which requires that items be included from previous and later grade(s). The items are generally high quality and test forms at grades 5 and 8 have a range of cognitive demand, but in each case the distribution contains significantly greater emphasis at DOK 3 as reflected by the standards. Thus, students who score well on the assessments will have demonstrated strong understanding of the standard's more complex skills. However, the grade 8 test may not fully assess standards at the lowest level of cognitive demand. The tests would better meet the CCSSO Criteria with an increase in the number of items focused on the major work of the grade and the addition of more items at grade 8 that assess standards at DOK 1.
MCAS	In mathematics, MCAS receives a Limited Match to the CCSSO Criteria for content and an Excellent Match for depth relative to assessing whether students are on track to meet college and career readiness standards. The MCAS mathematics test Items are of high technical and editorial quality. Additionally, the content is distributed well across the breadth of the grade level standards, and test forms closely reflect the range of cognitive demand of the standards. Yet, the grade 5 tests have an insufficient degree of focus on the major work of the grade. While mathematical practices are required to solve items, MCAS does not specify the assessed practices(s) within each item or their connections to content standards. The tests would better meet the criteria through increased focus on major work at grade 5 and identification of the mathematical practices that are assessed—and their connections to content.
PARCC	In mathematics, PARCC receives a Good Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. The assessment is reasonably well aligned to the major work of each grade. At grade 5, the test includes a distribution of cognitive demand that is similar to that of the standards. At grade 8, the test has greater percentages of higher-demand items (DOK 3 and 4) than reflected by the standards, such that a student who scores well on the grade 8 PARCC assessment will have demonstrated strong understanding of the standard's more complex skills. However, the grade 8 test may not fully assess standards at the lowest level (DOK 1) of cognitive demand. The test would better meet the CCSSO Criteria through additional focus on the major work of the grade, the addition of more items at grade 8 that assess standards at DOK 1 and increased attention to accuracy of the items, primarily editorial, but in some instances mathematical.
Smarter Balanced	In mathematics, Smarter Balanced has a Good Match to the CCSSO Criteria relative to assessing whether students are on track to meet college and career readiness standards. The test provides adequate focus on the major work of the grade, although it could be strengthened at grade 5. The tests would better meet the CCSSO Criteria through increased focus on the major work at grade 5, an increase in the number of items on the grade 8 tests assessing standards at DOK 1, and attention to serious mathematical or editorial flaws in some items.

MATHEMATICS CONTENT CRITERIA OVERVIEW

Criteria	Program and Rating	Summary Statements
I. Assesses the content most needed for College and Career Readiness.	L **ACT Aspire** LIMITED/UNEVEN MATCH	ACT Aspire provides a Limited/Uneven Match to the CCSSO Criteria for Content in Mathematics. The program does not focus exclusively on the major work of the grade, but rather, by design, assesses material from previous and later grade(s). This results in a weaker match to the criteria. The tests could better meet the criteria at both grades 5 and 8 by increasing the number of items that assess the major work of the grade.
	L **MCAS** LIMITED/UNEVEN MATCH	MCAS provides a Limited/Uneven Match to the CCSSO Criteria for Content in Mathematics. While the grade 8 assessment focuses strongly on the major work of the grade, the grade 5 assessment does not, as it samples more broadly from the full range of standards for the grade. The tests could better meet the criteria through increased focus on the major work of the grade on the grade 5 test.
	G **PARCC** GOOD MATCH	PARCC provides a Good Match to the CCSSO Criteria for Content in Mathematics. The test could better meet the criteria by increasing the focus on the major work at grade 5.
	G **Smarter Balanced** GOOD MATCH	Smarter Balanced provides a Good Match to the CCSSO Criteria for Content in Mathematics. The tests could better meet the criteria by increasing the focus on the major work in grade 5.

MATHEMATICS CONTENT CRITERIA

Criteria	Program and Rating	Summary Statements
C.1 Focus: Tests focus strongly on the content most needed in each grade or course for success in later mathematics (i.e., Major Work).	W **ACT Aspire** WEAK MATCH	On "focusing strongly on the content most needed for success in later mathematics," the rating is Weak Match. ACT Aspire forms do not consistently place sufficient emphasis on the major work of the given grade, due in part to intentional test design, which requires inclusion of selected content from earlier and later grades. Still, many of the items coded to standards from lower grades do not address the major work of the relevant grade.
	L **MCAS** LIMITED/UNEVEN MATCH	On "focusing strongly on the content most needed for success in later mathematics," the rating is Limited/Uneven Match. The grade 8 assessment is focused on the major work of the grade. The grade 5 assessment is significantly less focused on the major work of the grade than called for by the criterion, as it samples content across the full set of standards for the grade.
	G **PARCC** GOOD MATCH	On "focusing strongly on the content most needed for success in later mathematics," the rating is Good Match. While the grade 8 tests focus strongly on the major work of the grade, the grade 5 tests fall short of the threshold required for the top rating.
	G **Smarter Balanced** GOOD MATCH	On "focusing strongly on the content most needed for success in later mathematics," the rating is Good Match. While the grade 8 tests focus strongly on the major work of the grade, the grade 5 tests fall short of the threshold required for the top rating.

MATHEMATICS CONTENT CRITERIA		
Criteria	**Program and Rating**	**Summary Statements**
***C.2 Concepts, procedures, and applications: Assessments place balanced emphasis on the measurement of conceptual understanding, fluency and procedural skill, and the application of mathematics.**	**ACT Aspire** N/A	The test forms contain items that assess conceptual understanding, procedural fluency, and application. However, forms contain few test problems in which procedural skill/fluency is the predominant emphasis and therefore may under-assess these skills.
	MCAS N/A	The test forms contain items that assess conceptual understanding, procedural fluency, and application. Particularly in fifth grade, however, test forms are overly focused on procedural skill/fluency and application relative to conceptual understanding.
	PARCC N/A	The test forms contain items that assess conceptual understanding, procedural fluency, and application. Some of the application problems, however, have shallow contexts that are not necessary or important to the problem.
	Smarter Balanced N/A	The test forms contain items that assess conceptual understanding, procedural fluency, and application. However, the percentage of each problem type varies across forms, with some forms having a relative excess of application problems and other forms an excess of procedural skill/fluency problems.

*All four programs require, in their program documentation, the assessment of conceptual understanding, procedural skill/fluency, and application, although most do not clearly distinguish between procedural skill/fluency and conceptual understanding. Also, specific balance across these three types is not required. Due to variation across reviewers in how this criterion was understood and implemented, final ratings could not be determined with confidence. Therefore, only qualitative observations are provided.

MATHEMATICS DEPTH CRITERIA OVERVIEW

Criteria	Program and Rating	Summary Statements
II. Assesses the depth that reflects the demands of College and Career Readiness.	G **ACT Aspire** GOOD MATCH	ACT Aspire provides a Good Match to the CCSSO Criteria for Depth in Mathematics. The items are well crafted and clear, with only rare instances of minor editorial issues. The ACT Aspire tests include proportionately more items at high levels of cognitive demand (DOK 3) than the standards reflect, and proportionately fewer at the lowest level (DOK 1). This finding is both a strength, in terms of promoting strong skills, and a weakness, in terms of ensuring adequate assessment of the full range of cognitive demand within the standards. While technically meeting the criterion for use of multiple item types, the range is nonetheless limited, with the large majority comprising multiple-choice items. The program would better meet the criteria for Depth by including a wider variety of item types and relying less on traditional multiple-choice items.
	E **MCAS** EXCELLENT MATCH	MCAS provides an Excellent Match to the CCSSO Criteria for Depth in Mathematics. The assessment uses high-quality items and a variety of item types. The range of cognitive demand reflects that of the standards of the grade. While the program does not code test items to math practices, mathematical practices are nonetheless incorporated within items. The program might consider coding items to the mathematical practices and making explicit the connections between specific practices and specific content standards.
	G **PARCC** GOOD MATCH	PARCC provides a Good Match to the CCSSO Criteria for Depth in Mathematics. The tests include items with a range of cognitive demand, but at grade 8 that distribution contains a higher percentage of items at the higher levels (DOK 2 and 3) and significantly fewer items at the lowest level (DOK 1). This finding is both a strength, in terms of promoting strong skills, and a weakness, in terms of ensuring adequate assessment of the full range of cognitive demand within the standards. The tests include a variety of item types that are largely of high quality. However, a range of problems (from minor to severe) surfaced relative to editorial accuracy and, to a lesser degree, technical quality. The program could better meet the Depth criteria by ensuring that all items meet high editorial and technical standards and by ensuring that the distribution of cognitive demand on the assessments provides sufficient information across the range.
	G **Smarter Balanced** GOOD MATCH	Smarter Balanced provides a Good Match to the CCSSO Criteria for Depth in Mathematics. The exam includes a range of cognitive demand that fairly represents the standards at each grade level. The tests have a strong variety of item types including those that make effective use of technology. However, a range of problems (from minor to severe) surfaced relative to editorial accuracy and, to a lesser degree, technical quality. A wide variety of item types appear on each form, and important skills are assessed with multiple items, as is sound practice. Yet, individual forms sometimes contained two or three items measuring the same skill that were nearly identical, with only the numerical values changed in the item stem and a different set of answer choices. Such near-duplication may not impact the accuracy of the score, but a greater variety of question stems/scenarios is desirable. The program could better meet the Depth criteria by ensuring that all items meet high editorial and technical standards and that a given student is not presented with two or more virtually identical problems.

MATHEMATICS DEPTH CRITERIA

Criteria	Program and Rating	Summary Statements
C.3 Connecting practice to content: Test questions meaningfully connect mathematical practices and processes with mathematical content.	E **ACT Aspire** EXCELLENT MATCH	On "connecting practice to content," the rating is Excellent Match. All items that are coded to mathematics practices are also coded to one or more content standard.
	E **MCAS** EXCELLENT MATCH	On "connecting practice to content," the rating is Excellent Match. Although no items are coded to mathematical practices, the practices were nonetheless assessed within items that also assessed content.
	E **PARCC** EXCELLENT MATCH	On "connecting practice to content," the rating is Excellent Match. All items that are coded to mathematics practices are also coded to one or more content standard.
	E **Smarter Balanced** EXCELLENT MATCH	On "connecting practice to content," the rating is Excellent Match. All items that are coded to mathematics practices are also coded to one or more content standard.
C.4 Cognitive demand: The distribution of cognitive demand for each grade level is sufficient to assess the depth and complexity of the standards.	L **ACT Aspire** LIMITED/UNEVEN MATCH	On "requiring a range of cognitive demand," the rating is Limited/Uneven Match. At both grades 5 and grade 8, the test forms include significantly more items of high cognitive demand (DOK 3) than reflected in the standards, and proportionately fewer at the lowest level (DOK 1). While these items increase the challenge of the tests, standards that call for the lowest level of cognitive demand (DOK 1) may be under-assessed.
	E **MCAS** EXCELLENT MATCH	On "requiring a range of cognitive demand," the rating is Excellent Match. At each grade level, the distribution of cognitive demand closely reflects that of the standards.
	G **PARCC** GOOD MATCH	On "requiring a range of cognitive demand," the rating is Good Match. The distribution of cognitive demand of items reflects that of the standards very well at grade 5, while the grade 8 test includes proportionately more items at the higher levels of cognitive demand (DOK 2 and 3). As a result, Grade 8 standards that call for the lowest level of cognitive demand may be under-assessed.
	G **Smarter Balanced** GOOD MATCH	On "requiring a range of cognitive demand," the rating is Good Match. The distribution of cognitive demand of items reflects that of the standards very well at grade 5. At grade 8, the test includes proportionately fewer items at the lowest levels of cognitive demand (DOK 1) than in the standards, and proportionately more items at the mid-level of cognitive demand (DOK 2). As a result, grade 8 standards that call for the lowest level of cognitive demand may be under-assessed.

MATHEMATICS DEPTH CRITERIA

Criteria	Program and Rating	Summary Statements
C.5 High-quality items and variety of item types: Items are of high technical and editorial quality, are aligned to the standards, and each test form includes at least two items types including at least one that requires students to generate rather than select a response.	E **ACT Aspire** EXCELLENT MATCH	On "ensuring high-quality items and a variety of item types," the rating is Excellent Match. The program uses multiple item types, including constructed response on the Extended Task items. These items, although they carry high point values, are limited in number; the rest of the items are predominantly multiple-choice. The large majority of items are of high technical and editorial quality, with only very minor issues of editing, language, or accuracy. At the grade 8 level, some items appear to be susceptible to simplification by use of calculators, which are allowed on all items at grade 8, in contrast to the other programs that allow them on a restricted set of items.
	E **MCAS** EXCELLENT MATCH	On "ensuring high-quality items and a variety of item types," the rating is Excellent Match. Both grade 5 and grade 8 forms include multiple item types, including constructed-response. The items are of high technical and editorial quality, with very minor issues of editing, language, and accuracy at grade 8.
	G **PARCC** GOOD MATCH	On "ensuring high-quality items and a variety of item types," the rating is Good Match. The program includes a wide variety of item types, including several that require student-constructed responses. However, there are a number of items with quality issues, mostly minor editorial but sometimes mathematical.
	L **Smarter Balanced** LIMITED/UNEVEN MATCH	On "ensuring high-quality items and a variety of item types," the rating is Limited/Uneven Match. The program includes a wide variety of item types, many of which make effective use of technology. The program could be improved by ensuring that virtually identical items are not presented to individual students. Further, a good deal of variability across forms and grades is observed, with some forms fully meeting the item quality criterion and others only partially meeting it. Issues exist with the editorial quality and mathematical accuracy of individual items, most of which are minor, but some of which could impact assessment of the targeted skill, resulting in a rating of Limited/Uneven.

Appendix G

Testing Program Responses to Study and Descriptions of Test Changes for 2015–2016

ACT Aspire

Response to Report

One primary purpose of this study was to identify areas for improvement in each of the four evaluated testing programs. The study's findings include insights that promise to advance the industry standards for assessment quality. ACT Aspire is taking this opportunity to listen to the findings and implement adjustments to the assessment.

There are also aspects of the assessment that we have identified as areas for change through our own internal analyses, and we have already made design adjustments that will improve ACT Aspire in 2015–16. Both ACT Aspire's response to the study and the 2015–16 design adjustments will be discussed here.

Response to Study Findings about ACT Aspire

ACT Aspire is planning changes to two key elements for which the study found limited alignment with the CCSSO Criteria in English Language Arts:

1. Writing
   - Although the ACT Aspire Writing test was intentionally designed to have writing tasks that do not contain the heavy reading load of "writing to sources" tasks, we are currently exploring updated designs that would supplement the current items with tasks that measure these valuable literacy skills. These tasks would also improve coverage of the "Assessing research and inquiry" criterion in the CCSSO framework.
2. Reading
   - In response to the findings about distribution of Depth of Knowledge (DOK), ACT Aspire has already increased the percentage of upper-level DOK items. This effort will build on changes already in effect for the 2016 assessments (DOK 3 items in grades 5 and 8 will increase from 31 percent in 2015 to 38 percent in 2016).
   - ACT Aspire is adding new technology-enhanced item designs that emphasize selecting evidence directly from the passage to support claims and interpretations. While some of these new TE items will be operational in 2016, ACT Aspire is continuing to explore new ways to assess student use of evidence from texts.

ACT Aspire would also like to make a clarification about terminology used to classify texts on the Reading test:

- The study findings indicate that while ACT Aspire is a Good Match in Depth, the tests should have "additional literary narrative text, as opposed to literary informational texts." The study's ELA/Literacy panel has made a different interpretation than ACT Aspire of CCSSO criterion B.1 that this finding refers to. Differences of interpretation around genre definitions are understandable, but it is important to note the effects on the study outcomes. The CCSSO criteria B.1 refers to texts that are "balanced across literary and informational text types and genres" and does not specify a balance of fiction and nonfiction in the literary category. The study's panel interprets "literary" text types as only including literary narrative *fiction*. ACT Aspire, however, interprets "literary" to include both literary fiction and literary nonfiction passages that have a narrative structure. In accordance with B.1 ("In all grades, informational texts are primarily expository rather than narrative in structure"), ACT Aspire does not include texts that have a primarily narrative structure in the informational category. Aspire's interpretation of criterion B.1 results in a stronger match to the specified balance of text types.

In math, ACT Aspire is enacting changes in response to three of the study findings:

1. Range of item types
   - The report distinguished levels of use of multiple-choice items of 50 percent and 75 percent and preferred 50 percent or less. With regard to technology-enhanced items, the report also recommended "using them strategically (read sparingly)" in order to use resources wisely. ACT Aspire will apply its research to make high-quality items of all types and look to expand what is possible in directions that involve technology.
2. Content focus
   - The report recommended "an increase in the number of items focused on the major work of the grade." We will be working to increase this focus and gathering data to understand the balance in terms of promoting college and career readiness.
3. Depth of Knowledge
   - The study's review panel recommended "the addition of more items at grade 8 that assess standards at DOK 1." ACT Aspire currently has a plan in place that will increase the number of DOK 1 questions.

2015–16 Test Program Changes

In an effort to continuously improve ACT Aspire, we have already made adjustments in the following three categories for 2015–16:

1. Timing Adjustments – Based on customer feedback and in order to allow all students a better opportunity to show what they know and can do, we will be adjusting the time per test by five to ten minutes (Writing will not change). (See Tables G-1 and G-2 for more information on timing and point adjustments by grade and category.)
2. Adjustments to English Test – Adding six multiple-choice items for grades 3, 4, and 5.
3. Adjustments to Math Test – Adding six multiple choice items for grades 3, 4, and 5; removing one constructed-response (CR) item from grades 3, 4, and 5.

TABLE G-1

Timing Adjustments

ACT Aspire Summative Testing Time Adjustments (in minutes)								
Grade	**English (Current)**	**English (New)**	**Math (Current)**	**Math (New)**	**Reading (Current)**	**Reading (New)**	**Science (Current)**	**Science (New)**
3	30	**40**	55	**65**	60	**65**	55	**60**
4	30	**40**	55	**65**	60	**65**	55	**60**
5	30	**40**	55	**65**	60	**65**	55	**60**
6	35	**40**	60	**70**	60	**65**	55	**60**
7	35	**40**	60	**70**	60	**65**	55	**60**
8	35	**40**	65	**75**	60	**65**	55	**60**
***EHS**	40	**45**	65	**75**	60	**65**	55	**60**

*Early High School

TABLE G-2

English and Mathematics: Number of Points by Reporting Category

	GRADE									
	3 (new)	**3 (old)**	**4 (new)**	**4 (old)**	**5 (new)**	**5 (old)**	**6**	**7**	**8**	**EHS**
Reporting Category	*# of Points*									
English										
Production of Writing	12–14	9–11	8–10	6–8	8–10	6–8	11–13	9–11	9–11	12–14
Knowledge of Language			3–5	2–4	3–5	2–4	2–4	4–6	4–6	6–8
Conventions of Standard English	17–19	14–16	17–19	14–16	17–19	14–16	19–21	19–21	19–21	29–31
Total for English	**31**	**25**	**31**	**25**	**31**	**25**	**35**	**35**	**35**	**50**
Mathematics										
Number & Operations in Base 10		5–7	5–8	3–5	5–8	3–5	1–3	1–3	1–3	0–2
Number & Operations - Fractions	3–5	2–4	6–8	4–6	6–8	4–6	1–3	1–3	1–3	0–2
The Number System							3–5	3–5	2–4	1–3
Number & Quantity										1–3
Operations & Algebraic Thinking	6–8	3–5	4–6	3–5		3–5	1–3	1–3	0–2	0–2
Expressions & Equations							3–5	3–5	5–7	2–4
Ratios & Proportional Reasoning							3–5	3–5	0–2	1–3
Algebra										2–4
Functions									3–5	3–5
Measurement & Data (measurement)							0–2	0–2	1–3	1–3
Geometry		3–5		3–5	4–6	3–5	5–7	4–6	6–8	5–7
Measurement & Data	5–7	3–5		3–5		3–5				
Measurement & Data (data)							0–2	1–3	1–3	1–3
Statistics & Probability							3–5	3–5	4–6	4–7
Justification & Explanation	12	16	12	16	12	16	16	16	20	20
Total for Mathematics	**39**	**37**	**39**	**37**	**39**	**37**	**46**	**46**	**53**	**53**

MCAS

Response to Report

Our goal as a Commonwealth is to ensure that every Massachusetts student is prepared to succeed in postsecondary education and compete in the global economy. We have been administering annual assessments in Massachusetts since 1998 as our way of holding ourselves accountable for our progress toward this goal. The Massachusetts Comprehensive Assessment System (MCAS) tests are generally considered the gold standard of state assessments. They hold students to high expectations—in most cases, equivalent to the proficiency standard on the National Assessment of Educational Progress (NAEP)—and use a variety of question formats to ensure that we assess the full range of student abilities. Over the years we have refined the assessments to adapt to changes in the curriculum frameworks, most notably the incorporation of the Common Core State Standards into our 2010 frameworks, and to improve the quality of the assessment over time.

Our students and educators have accomplished incredible things under this system. Massachusetts' NAEP scores have moved from middle of the pack to leading the nation, and our students have scored well on international assessments. We have also made substantial progress toward closing the proficiency gaps between student subgroups, and we have dramatically reduced our dropout rate and increased our cohort graduation rate. That success would not have been possible without a high-quality assessment providing feedback on student, school, district, and state achievement and progress.

The Massachusetts Comprehensive Assessment Systems was a terrific twentieth-century assessment—but it has reached a point of diminishing returns. In 2015, MCAS was administered for the eighteenth year. We have a better understanding now than we did a decade or two ago about learning progression in mathematics, text complexity and the interplay of reading and writing, and the academic expectations of higher education and employers. And we now know that nearly one-third of our public high school students who go on to enroll in Massachusetts public colleges take at least one remedial course in their first semester, suggesting that the curriculum and assessments they have experienced have not adequately prepared them for the world beyond high school. Indeed, MCAS was never designed to be an indicator of college and career readiness. We joined the Partnership for Assessment of Readiness for College and Careers (PARCC) consortium specifically in order to partner with other states in developing an assessment that is more closely aligned to these expectations.

Thus, we were not surprised by this report's conclusion that the MCAS does not always measure well what's most important today. This report also confirms that in many ways, PARCC sets a higher bar than MCAS for student performance. This is particularly true as students move up the grades into middle and high school. This higher bar is not simply about being harder: PARCC provides more opportunities for critical thinking, applying knowledge, research, and making connections between reading and writing. More and more schools have upgraded curriculum and instruction to align with our 2010 frameworks. While we adjusted MCAS to test those frameworks, PARCC was built around them. Classroom instruction is now increasingly focused on the knowledge and skills in the frameworks, rather than how to pass a test.

We are proud of what we have accomplished in Massachusetts in the nearly two decades that we have been administering the MCAS. Now that we have the benefit of that experience and have revised our curriculum frameworks to reflect our upgraded learning expectations, it is time to upgrade our assessments too. Our state Board of Elementary and Secondary Education voted in November 2015 to do exactly that.

2015–16 Test Program Changes

Over the next few years, we will transition to a new statewide assessment system that will take much of what this report identifies as the strengths of PARCC—high-quality content aligned strongly to college and career ready standards—and combine it with elements of MCAS in the context of a Massachusetts-specific governance system that will allow us to set our own policies on test content, administration, and reporting. With this approach, we will continue to benefit from a high-quality, next-generation assessment while ensuring that the test will reflect the Commonwealth's unique needs and concerns. Most importantly, our students will be better prepared for success after high school—our ultimate goal.

PARCC

Response to Report

PARCC would like to thank the study authors and review panelists for a comprehensive, strong study. There are two areas where we would like to present a few additional comments.

ELA/Literacy Content Rating

Panel Recommendation:

The tests could be strengthened by the addition of research tasks that require students to use two or more sources and, as technologies allow, a listening and speaking component.

PARCC Response:

Every PARCC assessment in grades 3–8 requires students to complete a research simulation task where the student reads two or three sources and must integrate or synthesize the ideas in a written essay. All students also read two literary texts and write a literary analysis (literary research) essay.

The PARCC assessment measures many aspects that are key to the Speaking and Listening standards. PARCC uses multimedia texts to measure comprehension for all students taking its tests online (providing students with opportunities to demonstrate strengths and needs in comprehending audio and audiovisual texts). The CCSS build coherence across the ELA strands and identify similar skills built into both the reading comprehension standards (standards RI.7 and RL.7) and the listening standards. PARCC chose to report students' speaking and listening performance in relation to the reading standards.

The PARCC assessment system includes a robust set of Speaking and Listening tools. All schools administering PARCC in 2015–2016 have access to a comprehensive set of formative assessments and instructional tools to support educators, parents, and students in better understanding students' strengths and needs in speaking and listening. Further information about the PARCC Speaking and Listening tools can be found on PARCC's Partnership Resource Center: https://prc.parcconline.org/library/speaking-and-listening-overview.

Cognitive Demand

Panel Recommendation:

The program could better meet the Depth criteria by ensuring that the distribution of cognitive demand on the assessments provides sufficient information across the range.

PARCC Response:

It is important to note that students who meet Level 1/Level 2 Depth of Knowledge (DOK) for items situated at higher DOK levels are given partial credit points for demonstrating skills that require lower cognitive complexity. Reviewers did not consider the possibility that scoring, rather than adding more Level 1 or Level 2 items, could allow for the balance of item complexities. For more information on the PARCC scoring rubrics and to view released items, visit: https://prc.parcconline.org/assessments/parcc-released-items.

The PARCC assessment uses a cognitive complexity framework that was developed by the PARCC consortium to more accurately reflect the demands of the CCSS. This framework received recognition from AERA (2014 Outstanding Contribution to Practice in Cognition and Assessment award). An article detailing the innovations of this framework and potential next steps in research around cognitive complexity has been published in a new book titled *The Next Generation of Testing: Common Core Standards, Smarter-Balanced*.[106]

106. H. Jiao and R. Lissitz, eds., *The Next Generation of Testing: Common Core Standards, Smarter-Balanced, PARCC, and the Nationwide Testing Movement* (Charlotte, NC: Information Age Publishing, Inc., 2016).

2015–16 Test Program Changes

In May 2015, the chief state school officers from the PARCC states unanimously voted to streamline the assessment. They accomplished this goal while retaining all the key elements of the test—a strong commitment to quality and reliability, measurement of the full range of the standards, and the ability to get results back to teachers and parents quickly, so that they can help meet the needs of students for the coming school year. The following changes to the test design will be instituted in the 2015–16 school year:

- The two testing windows (the performance-based and end-of-year components) in mathematics and English language arts/literacy (which includes reading and writing) will be consolidated into one. The single testing window will simplify administration of the test for states and schools. Schools will have up to thirty school days to administer the test, and the testing window will extend from roughly the 75 percent mark to the 90 percent mark of instructional time. Most schools will complete testing in one to two weeks during that window.

  The testing time for students will be reduced by about ninety minutes overall (sixty minutes in mathematics; thirty minutes in English language arts/literacy). The result will be that the total testing time for ELA/Literacy and mathematics will be approximately 8.5 hours at grades 3–5, 9.2 hours at grades 6–8, and 9.7 hours at grade 11. There will also be more uniformity of test unit times, allowing for easier scheduling in schools.

- Each PARCC assessment is administered in multiple sections, called units. The number of test units was reduced for all students, and includes three units in English language arts/literacy and three or four units in mathematics.

  The testing time was shortened by reducing the number of score points and items in both subject areas. The tables below show a comparison of score points between the previous test design and the redesign.

TABLE G-3

Comparison of Score Points in the Previous ELA/Literacy Design and the Redesign

		Previous Two Administrations	Adopted Single Administration
Grade 3	Reading Points	64	58
	Writing Points	36	36
	Total Points	100	94
	Units	4	3
	Total Testing Time	4.75 hours	4.25 hours*
Grades 4–5	Reading Points	70	62
	Writing Points	36	36
	Total Points	106	98
	Units	4	3
	Total Testing Time	5.0 hours	4.5 hours**
Grades 6–11	Reading Points	94	76
	Writing Points	45	45
	Total Points	139	121
	Units	5	3
	Total Testing Time	5.75 hours	5.2 hours***

* Add 1.5 hours for field test unit

** Add 1.5 hours for field test unit

***Add 1.8 hours for field test unit

TABLE G-4

Comparison of Score Points in the Previous Mathematics Design and the Redesign

		Previous Two Windows	Adopted Single Administration
Grade 3-8	Short Items	56 pts	40 pts
	Reasoning Items	14 pts	14 pts
	Modeling Items	12 pts	12 pts
	Total Points	82 pts	66 pts
	Units	4 @ varies	4 @ 60 min.
	Total Time on Task	5 hours	4 hours
Algebra I, Geometry, Algebra II, and Integrated Math I, II, III	Short Items	65 pts	49 pts
	Reasoning Items	14 pts	14 pts
	Modeling Items	18 pts	18 pts
	Total Points	97 pts	81 pts
	Units	4 @ varies	3 @ 90 min.
	Total Time on Task	5.3 - 5.5 hours	4.5 hours

- Standalone field testing will be eliminated. As with all similar testing, field test items—items that could be used in future years—are embedded in each student's test. Because the performance tasks in English language arts/literacy are longer, a sampling of students had to take a standalone field test unit for these tasks in spring 2015. To further streamline the testing process for all schools, the PARCC field test will now be wrapped into the testing window. Each year, a small percentage of students will take an additional ELA/Literacy unit. Schools and classrooms selected in one year—per the process determined in their state—will in almost all cases not have to field test again for several years.
- The test design changes do not result in the loss of any performance tasks in English language arts/literacy (there are still three performance tasks). Additionally, there are now two or three text sets included in the units, depending on the grade level (one text set was removed for grades 6–11).

  The test design changes do not result in the loss of any reasoning and modeling mathematics items, with the exception of Algebra II and Integrated Math III at the high school level. Short answer items were removed.

  For more information, visit http://parcconline.org/assessments/test-design/design-changes.

Smarter Balanced

Response to Report

Tony Alpert, Executive Director
Luci Willits, Deputy Executive Director

December 9, 2015

Thank you to the Thomas B. Fordham Institute and HumRRO for its diligent work to evaluate the quality of the Smarter Balanced summative assessment and its alignment to the Common Core State Standards.

While this report focused on the end-of-year test, Smarter Balanced is more than a summative assessment: it's a system to improve teaching and learning. Our system includes optional and flexible interim assessments available throughout the year to help teachers monitor student progress, as well as a Digital Library with thousands of educator-approved classroom resources. Nearly 5,000 educators from across the country helped build the Smarter Balanced system. Smarter Balanced assessments are designed to be administered online and are customized for every child using built-in accessibility resources.

This report recognizes many of these strengths and gives Smarter Balanced an Excellent or Good Match in all but one subcategory. In addition, the report recognizes that Smarter Balanced is the only assessment that measures students' listening skills. We are proud of these ratings. We also recognize that there is always room for improvement. However, one of the greatest strengths of Smarter Balanced, the computer-adaptive feature of the summative assessment, is not addressed in this report. Because it is an adaptive test that is customized for each student, it is difficult to compare the Smarter Balanced summative assessment on an item-per-item basis to a fixed form test that is static. In addition, this study did not consider some other important features of the Smarter Balanced assessment, including the ability of states to work with the service provider of their choice. Finally, Smarter Balanced is arguably the most accessible large-scale assessment system that includes supports for over 90 percent of the consortium's English language learners' primary languages. Individually, these elements are historic; collectively, they are unprecedented.

It is important to note that due to the timing of this study, reviewers were not able to access all of the interactive features that are available to all students during a live test. For instance, reviewers did not interact with features such as highlighting text in passages and test questions, zooming in and out of test pages, making notes about a test question in the notepad, and using strikethrough for answer options. In addition, the study's version of the system did not provide some of the helpful built-in student tools, such as error messages when students use incorrect keys, the ability to mark items, or move forward in the test without answering all the questions on a page.

2015–16 Test Program Changes

For the 2015–2016 summative tests, Smarter Balanced members have the flexibility to determine whether classroom activities will be given prior to the administration of performance tasks.

The following table describes the differences between the system used in the study versus the one actually used for students:

TABLE G-5

Differences in Student and Study Interface

Students' Actual Version:	Study Version:
Calculators available in grades 6 through high school for items when not measuring computation.	No calculators.
Tutorials to show students how to use the available tools and to interact with all of the different types of items they might see on a test.	No tutorials.
Grade appropriate and item-specific English glossaries are available for mathematics and English language arts.	No glossaries.
For mathematics, grade appropriate item-specific translated glossaries are available in ten different languages, plus dialects.	No glossaries.
Error messages given to students when they try to enter characters that aren't allowed.	No error messages when equation editor and fill-in blank items are incorrectly populated.
Verdana Font (14 pt)	Times New Roman font (12 pt)
Formatted for best results according to student cognitive labs and field testing.	Format not consistent.
All the research-based tools available as appropriate to the content area and item (as shown in the practice tests).	Limited availability of tools. For example, notepad, underline, highlight, etc. were not available.

As part of the development process, Smarter Balanced collaborated with national experts and local teachers to determine how to best measure critical thinking and problem solving skills as part of college and career ready standards. For example, at times, it is most appropriate to ask students to solve engaging items within a real-world scenario; while at other times, presenting students with an equation to solve is a better way to measure student knowledge. This is reflected in the test blueprint.

It is important to note that with the adaptive test, Smarter Balanced can measure more complex skills for low- and high-performing students alike. In mathematics, Smarter Balanced chose to emphasize the more complex skill sets with the understanding that students must have the procedural knowledge to do well on the test. With English language arts, we will discuss the report's findings with our membership and consider changes.

Smarter Balanced is committed to including only high-quality questions on our tests. We were disappointed to see that reviewers found a handful of questions that needed improvement and received a rating of Limited/Uneven Match. Smarter Balanced has an extensive process for question development to ensure each test item

is extensively reviewed prior to being included on a student's test. Educators including national mathematical, English language arts, and accessibility experts write questions and review them for content accuracy as well as for any potential bias or lack of sensitivity. Questions that do not meet a very high standard are revised or are removed. However, we will use this study to improve our item development review processes. Immediately, Smarter Balanced will initiate a detailed review of the existing test questions based on the feedback from this report.

In addition to this positive review of Smarter Balanced, we were pleased to note that the National Network of State Teachers of the Year echoed complimentary feedback in their report as well. That review looked at many of the same questions as this review. The nation's best teachers said Smarter Balanced provides a better picture of student performance, is grade-level appropriate, and supports great teaching and learning throughout the year.

Thank you again for your review and for the opportunity to provide more context into the reviewers' findings.

Sincerely,

Anthony Alpert
Executive Director
Smarter Balanced Assessment Consortium