AIR®
AMERICAN INSTITUTES FOR RESEARCH®

NATIONAL BENCHMARKS

For State Achievement Standards

February 22, 2016

Gary W. Phillips
Vice President and Institute Fellow
American Institutes for Research

Contents

List of Tables

Executive Summary

State achievement standards represent how much the state expects their students to learn in order to reach various levels of academic proficiency. In this study, the academic subjects are English language arts (ELA) and mathematics. In the past, these achievement standards were used by each state to report adequate yearly progress (AYP) under No Child Left Behind (NCLB) federal legislation, and they are currently being used for federal reporting under the Every Student Succeeds Act (ESSA) of 2015. These standards are also used by the state to monitor progress from year to year and to report on the success of each classroom, school, and district to parents and the public.

This report uses national benchmarking as a common metric to examine state achievement standards and compare how high these standards are compared to the National Assessment of Educational Progress (NAEP) achievement levels. It also compares how much students are expected to learn in some states with how much they are expected to learn in other states. The study uses NAEP grades 4 and 8 reading and mathematics as benchmarks for individual state achievement standards. The study also benchmarks the achievement standards of Smarter Balanced Assessment Consortium (referred to in this study as *Smarter Balanced*), Partnership for Assessment of Readiness for College and Careers (PARCC), and ACT Aspire. Benchmarking Smarter Balanced, PARCC, and ACT Aspire provides a common metric (i.e., the NAEP scale) that can be used to compare the stringency of their achievement standards. The most important findings in the study relate to achievement standards that represent college readiness. Each of these consortium tests in grades 4 and 8 has achievement standards that indicate the student is on track to be college ready. The college-ready standards are Level 3 (Met) for Smarter Balanced, Level 4 (Met) for PARCC, and Level 3 (Ready) for ACT Aspire.

The overall findings in the study are:

1. Smarter Balanced college-ready standards (Level 3) are comparable in difficulty to the NAEP Basic levels.
2. Smarter Balanced college-ready standards (Level 3) are significantly below PARCC college-ready standards (Level 4) by about one-quarter of a standard deviation. In the statistical literature, a standard deviation unit is referred to as an *effect size*. The effect sizes are for ELA grades 4 and 8, and mathematics grades 4 and 8 are −.26, −.28, −.26 and −.36, respectively.
3. Smarter Balanced college-ready grade 8 standards are comparable to ACT Aspire college-ready grade 8 standards. However, for grade 4, the Smarter Balanced college-ready standard is significantly below the ACT Aspire college-ready standard for Reading (effect size = −.26) but significantly above the ACT Aspire college-ready standard for mathematics (effect size = +.29).
4. PARCC college-ready standards (Level 4) are comparable in difficulty to the NAEP Basic level for ELA and comparable to the NAEP Proficient level for mathematics.
5. PARCC college-ready standards (Level 4) are comparable in difficulty to the ACT Aspire college-ready standard for Reading grade 4. However, PARCC standards are significantly

above ACT Aspire college-ready standards for ELA grade 8 (effect size = +.28), mathematics grade 4 (effect size = +.55), and mathematics grade 8 (effect size = +.48).

6. ACT Aspire college-ready standards (Ready) are comparable in difficulty to the NAEP Basic levels.
7. Individual states that have college-readiness standards that map to the NAEP Proficient level are:

    a. ELA grade 4—Florida and New York;

    b. ELA grade 8—Florida, Kansas, and New York;

    c. Mathematics grade 4—Florida and Kansas; and

    d. Mathematics grade 8—Alaska, Florida, Kansas, New York, and Pennsylvania.

Note that Iowa, Nebraska, and Texas have three achievement levels, instead of the usual four levels or five levels in other states. At the time of this report, the author was unable to determine which levels in these states represented college readiness.

Brief History of Common Core–Related Activities

Role of NCLB: Probably the biggest contributor to the development of the Common Core State Standards (CCSS) was the passage of the No Child Left Behind Act of 2001. A fundament problem with NCLB demonstrated the need for the CCSS. NCLB required each state to have challenging content standards and performance standards but left it up to the state to define what "challenging" meant. Some states used low standards in order to report higher levels of proficiency. States with low standards were living in a kind of Lake Wobegon world where more and more students were being reported as proficient but fewer and fewer students were prepared for college. This led the National Governors Association (NGA) and the Council of Chief State School Officers (CCSSO) to see if there was a way to make state standards more competitive and consistent.

Role of NGA and CCSSO: In 2006–2007, Arizona Governor Janet Napolitano chaired the NGA. In order to find a way to make America's educational system internationally competitive, she created a task force of state and national education policy leaders that released a report titled "Benchmarking for Success: Ensuring U.S. Students Receive a World-Class Education" (2008). The state leaders responsible for the report were the NGA and the CCSSO as well as the nonprofit group Achieve. The concepts in this report caught on, and in 2009 state leaders launched CCSS. These three groups obtained the support of other organizations that were critical in the development of the CCSS. These organizations included the American Federation of Teachers, the National Education Association, the National Council of Teachers of Mathematics, the National Council of Teachers of English, and the International Reading Association.

Role of U.S. Federal Government: The CCSS was a state-led effort and was not initiated by the federal government. The NGA and the CCSSO received no financial support from the federal government to develop the CCSS. However, once CCSS was developed, the federal government used the bully pulpit to encourage many states to implement internationally competitive common standards. For example, in 2009 President Obama, in a speech to the U.S. Hispanic Chamber of Commerce, recognized the need for high and consistent standards. He stated:

> *Let's challenge our states to adopt world-class standards that will bring our curriculums into the 21st century. Today's system of 50 different sets of benchmarks for academic success means fourth-grade readers in Mississippi are scoring nearly 70 points lower than students in Wyoming—and getting the same grade.*

The federal government also provided seed money to help states implement common standards. The funding was provided in the 4.35 billion dollar Race to the Top grant as part of the American Recovery and Reinvestment Act of 2009, which was part of the federal economic stimulus package.

Role of Smarter Balanced and PARCC: Part of the Race to the Top grant was awarded to PARCC and Smarter Balanced to develop tests that measure the CCSS. Over several years of development, some states dropped out of the initiative. By spring 2015, 18 states had given the first operational administration of the Smarter Balanced assessment, and 11 states plus the District of Columbia gave the first operational administration of the PARCC assessment. These

are the jurisdictions on which the current consortium results are based. The Virgin Islands were also administered the Smarter Balanced assessment, but they were excluded in this mapping study because they did not participate in the 2015 NAEP assessment.

ACT Aspire: In 2015, ACT Aspire was administered in two states—Alabama and South Carolina—which represents a group of states taking the same assessment. Recognizing that a large portion of students were graduating high school unprepared for college, ACT developed an assessment that was built around college readiness beginning in elementary school. The ACT Aspire replaced the ACT Explore (grades 8 and 9) and ACT Plan (grade 10) and was administered in grades 3–10.

Benchmarking State Achievement Standards

Benchmarking is a way to calibrate the difficulty level of state achievement standards so they can be compared to each other and to national standards. This type of benchmarking is similar to benchmarking in business and industry. For example, the fuel efficiency and quality of American-built cars are often benchmarked against those of cars built in Japan and South Korea. Such benchmarking is important in education if we are to expect our students to compete in a global economy. In this study, we use the NAEP as a national benchmark.

Some terminology clarification is needed in order to navigate through the results of this study. This report is about benchmarking (or comparing) state achievement standards (cut-scores on the state accountability test used to report results to the federal government under ESSA) to the NAEP achievement levels. In some testing programs, achievement standards are referred to as *performance standards*. The comparisons are obtained through equipercentile linking (described in the Appendix). An achievement standard is a specific number, or cut-score, on the scale such as those in Tables 1–3. What this study does is determine the NAEP equivalent of the state achievement standard (or cut-score) and report the NAEP achievement level in which the NAEP equivalent falls. For example, the Smarter Balanced ELA grade 4 cut-score for Level 3 is 2473 (see One caveat in the study is that for Smarter Balanced and PARCC we are mapping *ELA* standards, which include writing, to NAEP *Reading* standards, which do not include writing. This should not make much difference because, generally, the dis-attenuated correlations between reading and writing are very high.

Table 1). The linking analysis shows this is equivalent in difficulty to a NAEP score of 222 (see Table 6). The NAEP equivalent of 222 falls within the range of the NAEP Basic level (208-237; see Table 4).

Grades 4 and 8 Achievement Standards for Smarter Balanced, PARCC, ACT Aspire, and NAEP

Each of the assessments used by groups of states in 2015 has its own achievement standards. In each case, the standards were set through a consortium or national consensus process and represent how much we expect students to know and be able to do at different levels of achievement. Possibly the most important achievement standard is the one that indicates the student is on track to be college ready by the end of high school. For Smarter Balanced this is Level 3, for PARCC this is Level 4, and for ACT Aspire this is Level 3. The achievement standards for each assessment—Smarter Balanced, PARCC, ACT Aspire, and NAEP—are indicated in Tables 1–4.

One caveat in the study is that for Smarter Balanced and PARCC we are mapping *ELA* standards, which include writing, to NAEP *Reading* standards, which do not include writing. This should not make much difference because, generally, the dis-attenuated correlations between reading and writing are very high.

Table 1: Smarter Balanced Achievement Standards

Subject	Grade	Level 2 Nearly Met	Level 3 Met	Level 4 Exceeded

ELA	4	2416	2473	2533
ELA	8	2487	2567	2668
Mathematics	4	2411	2485	2549
Mathematics	8	2504	2586	2653

Table 2: PARCC Performance Standards

Subject	Grade	Level 2 Partially Met	Level 3 Approached	Level 4 Met	Level 5 Exceeded
ELA	4	700	725	750	790
ELA	8	700	725	750	794
Mathematics	4	700	725	750	796
Mathematics	8	700	725	750	801

Table 3: ACT Aspire Achievement Standards

Subject	Grade	Level 2 Close	Level 3 Ready	Level 4 Exceeding
Reading	4	412	417	422
Reading	8	418	424	430
Mathematics	4	411	416	421
Mathematics	8	419	425	431

Table 4: NAEP Achievement Standards

Subject	Grade	Basic	Proficient	Advanced
Reading	4	208	238	268
Reading	8	243	281	323
Mathematics	4	214	249	282
Mathematics	8	262	299	333

Using NAEP as a National Benchmark

NAEP represents probably the best assessment against which to benchmark state achievement standards. First, the NAEP content standards and achievement standards were developed through an elaborate national process that has been exhaustively evaluated. NAEP standards have been demonstrated to be internationally competitive and are often referred to as the gold standard against which other standards can be compared. Second, NAEP provides biennial state representative assessments that can be treated as randomly equivalent to the local state testing population. This facilitates comparisons between local state testing results and state NAEP testing results. Third, because NAEP is administered in each state, the NAEP scale can be used as an anchor test to provide a common metric to compare local state-by-state testing results. This was the strategy used in this study.

National NAEP Benchmarks for Smarter Balanced

In 2015, 18 states and the Virgin Islands administered the Smarter Balanced assessment. Because they all used the same test, a weighted average of the percentage at and above each achievement level for the 18 states was used for the analysis. The weights were based on the student population size in each state. The 18 jurisdictions were California, Connecticut, Delaware, Hawaii, Idaho, Maine, Michigan, Missouri, Montana, Nevada, New Hampshire, North Dakota, Oregon, South Dakota, Vermont, Washington, West Virginia, and Wisconsin. The Virgin Islands were excluded because they did not participate in NAEP in 2015. For ELA, Wisconsin and Missouri were excluded from the weighted average because their administration deviated from the Smarter Balanced blueprint. North Dakota was excluded for both ELA and mathematics because the author was unable to find their results on their state web site. Aggregate NAEP estimates for the Smarter Balanced states were obtained from the NAEP Data Explorer (NDE).

The national NAEP benchmarks for Smarter Balanced are contained in Tables 5–7. The most important Smarter Balanced level to benchmark is Level 3, considered to represent being on track to be college ready. We see in Table 6 that each of the Smarter Balanced Level 3 cut-scores maps to the NAEP Basic achievement level.

Table 5: NAEP Equivalents of Smarter Balanced Achievement Standards for Level 2

Subject	Grade	Smarter Cut-Score for Level 2 Nearly Met	Percent at and Above Smarter Level 2 Nearly Met	NAEP Scaled Score Equivalent of Level 2 Nearly Met	Standard Error of NAEP Equivalent of Level 2 Nearly Met	NAEP Achievement Level Equivalent of Level 2 Nearly Met
ELA	4	2416	66	201	2.0	Below Basic
ELA	8	2487	77	236	1.0	Below Basic
Math	4	2411	74	216	1.0	Basic
Math	8	2504	61	269	1.0	Basic

Table 6: NAEP Equivalents of Smarter Balanced Achievement Standards for Level 3

Subject	Grade	Smarter Cut-Score for Level 3 Met	Percent at and Above Smarter Level 3 Met	NAEP Scaled Score Equivalent of Level 3 Met	Standard Error of NAEP Equivalent of Level 3 Met	NAEP Achievement Level Equivalent of Level 3 Met
ELA	4	2473	44	222	2.0	Basic
ELA	8	2567	48	264	1.0	Basic
Math	4	2485	40	244	1.0	Basic
Math	8	2586	35	294	1.0	Basic

Table 7: NAEP Equivalents of Smarter Balanced Achievement Standards for Level 4

Subject	Grade	Smarter Cut-Score for Level 4 Exceeded	Percent at and Above Smarter Level 4 Exceeded	NAEP Scaled Score Equivalent of Level 4 Exceeded	Standard Error of NAEP Equivalent of Level 4 Exceeded	NAEP Achievement Level Equivalent of Level 4 Exceeded
ELA	4	2533	22	247	2.0	Proficient
ELA	8	2668	13	302	1.0	Proficient
Math	4	2549	15	268	1.0	Proficient
Math	8	2653	17	315	1.0	Proficient

National NAEP Benchmarks for PARCC

In 2015, 11 states and the District of Columbia administered the PARCC assessment. Because they all used the same test, a weighted average of the percentage at and above each achievement level for the 12 jurisdictions was used for the analysis. The weights were based on the student population size in each state. The 12 jurisdictions were Arkansas, Colorado, District of Columbia, Illinois, Louisiana, Maryland, Massachusetts, Mississippi, New Jersey, New Mexico, Ohio, and Rhode Island. The aggregate NAEP estimate for the PARCC jurisdictions was obtained from the NDE.

The national NAEP benchmarks for PARCC are contained in Tables 8–11. The most important PARCC level to benchmark is Level 4, considered to represent being on track to be college ready. We see in Table 10 that each of the PARCC Level 4 cut-scores maps to the NAEP Basic achievement level for ELA and the NAEP Proficient achievement level for mathematics.

Table 8: NAEP Equivalents of PARCC Performance Standards for Level 2

Subject	Grade	PARCC Cut-Score for Level 2 Partially Met	Percent at and Above PARCC Level 2 Partially Met	NAEP Scaled Score Equivalent of Level 2 Partially Met	Standard Error of NAEP Equivalent of Level 2 Partially Met	NAEP Achievement Level Equivalent of Level 2 Partially Met
ELA	4	700	89	179	1.0	Below Basic
ELA	8	700	86	229	1.0	Below Basic
Math	4	700	88	200	1.0	Below Basic
Math	8	700	78	255	1.0	Below Basic

Table 9: NAEP Equivalents of PARCC Performance Standards for Level 3

Subject	Grade	PARCC Cut-Score for Level 3 Approached	Percent at and Above PARCC Level 3 Approached	NAEP Scaled Score Equivalent of Level 3 Approached	Standard Error of NAEP Equivalent of Level 3 Approached	NAEP Achievement Level Equivalent of Level 3 Approached
ELA	4	725	70	205	1.0	Below Basic
ELA	8	725	67	250	1.0	Basic
Math	4	725	62	228	1.0	Basic
Math	8	725	52	282	1.0	Basic

Table 10: NAEP Equivalents of PARCC Performance Standards for Level 4

Subject	Grade	PARCC Cut-Score for Level 4 Met	Percent at and Above PARCC Level 4 Met	NAEP Scaled Score Equivalent of Level 4 Met	Standard Error of NAEP Equivalent of Level 4 Met	NAEP Achievement Level Equivalent of Level 4 Met
ELA	4	750	41	232	1.0	Basic
ELA	8	750	42	273	1.0	Basic
Math	4	750	32	252	1.0	Proficient
Math	8	750	27	307	1.0	Proficient

Table 11: NAEP Equivalents of PARCC Performance Standards for Level 5

Subject	Grade	PARCC Cut-Score for Level 5 Exceeded	Percent at and Above PARCC Level 5 Exceeded	NAEP Scaled Score Equivalent of Level 5 Exceeded	Standard Error of NAEP Equivalent of Level 5 Exceeded	NAEP Achievement Level Equivalent of Level 5 Exceeded
ELA	4	790	7	277	1.0	Advanced
ELA	8	794	7	318	1.0	Proficient
Math	4	796	3	297	2.0	Advanced
Math	8	801	3	358	1.0	Advanced

National NAEP Benchmarks for ACT Aspire

In 2015, two states administered the ACT Aspire test. They were Alabama and South Carolina. Because both states used the same test, a weighted average of the percentage at and above each

achievement level was used for the analysis. The weights were based on the student population size in each state. The aggregate NAEP estimate for the ACT Aspire jurisdictions was obtained from the NDE. The national NAEP benchmarks for ACT Aspire are contained in Tables 12–14. The most important ACT Aspire level to benchmark is Level 3, considered to represent being on track to be college ready. We see in Table 13 that each of the ACT Aspire college-ready cut-scores map to the NAEP Basic achievement level.

Table 12: NAEP Equivalents of ACT Aspire Achievement Standards Level 2

Subject	Grade	ACT Aspire Cut-Score for Level 2 Close	Percent at and Above ACT Aspire Level 2 Close	NAEP Scaled Score Equivalent of Level 2 Close	Standard Error of NAEP Equivalent of Level 2 Close	NAEP Achievement Level Equivalent of Level 2 Close
Reading	4	412	67	202	2.0	Below Basic
Reading	8	418	72	240	1.0	Below Basic
Math	4	411	91	195	1.0	Below Basic
Math	8	419	59	263	2.0	Basic

Table 13: NAEP Equivalents of ACT Aspire Achievement Standards Level 3

Subject	Grade	ACT Aspire Cut-Score for Level 3 Ready	Percent at and Above ACT Aspire Level 3 Ready	NAEP Scaled Score Equivalent of Level 3 Ready	Standard Error of NAEP Equivalent of Level 3 Ready	NAEP Achievement Level Equivalent of Level 3 Ready
Reading	4	417	35	232	2.0	Basic
Reading	8	424	45	264	1.0	Basic
Math	4	416	49	235	1.0	Basic
Math	8	425	30	290	2.0	Basic

Table 14: NAEP Equivalents of ACT Aspire Achievement Standards Level 4

Subject	Grade	ACT Aspire Cut-Score for Level 4 Exceeding	Percent at and Above ACT Aspire Level 4 Exceeding	NAEP Scaled Score Equivalent of Level 4 Exceeding	Standard Error of NAEP Equivalent of Level 4 Exceeding	NAEP Achievement Level Equivalent of Level 4 Exceeding
Reading	4	422	13	260	2.0	Proficient
Reading	8	430	13	298	2.0	Proficient
Math	4	421	14	266	1.0	Proficient
Math	8	431	14	309	2.0	Proficient

National NAEP Benchmarks for Nonconsortium States

Across most of the nonconsortium states with four achievement levels, Level 3 is considered on track to be college ready. For many states with five achievement levels, Level 4 is considered on track to be college ready. However, this is not universally true. For Indiana, the author was not able to obtain the 2015 state results at the present time.

The results of NAEP benchmarks for ELA grade 4 individual states are reported in

Table 15. The reading grade 4 NAEP achievement level cut-scores are Basic = 208, Proficient = 238, and Advanced = 268. The only state with four achievement levels for which Level 3 maps to the NAEP Proficient level is New York. The only state with five achievement levels for which Level 4 maps to the NAEP Proficient level is Florida.

Table 15: ELA Grade 4 NAEP Benchmarks for Nonconsortium States

State	ELA Grade 4											
	Level 2			Level 3			Level 4			Level 5		
	Percent at and Above Level 2	NAEP Equivalent for Level 2	NAEP Achievement Level Comparable to Level 2	Percent at and Above Level 3	NAEP Equivalent for Level 3	NAEP Achievement Level Comparable to Level 3	Percent at and Above Level 4	NAEP Equivalent for Level 4	NAEP Achievement Level Comparable to Level 4	Percent at and Above Level 5	NAEP Equivalent for Level 5	NAEP Achievement Level Comparable to Level 5
Alaska	59	203	Below Basic	40	224	Basic	9	271	Advanced			
Arizona	59	206	Below Basic	42	223	Basic	6	277	Advanced			
DoDEA	93	193	Below Basic	72	217	Basic	37	243	Proficient			
Florida	79	202	Below Basic	54	224	Basic	27	246	Proficient	8	271	Advanced
Georgia	71	204	Below Basic	37	233	Basic	9	267	Proficient			
Iowa	76	198	Below Basic	29	244	Proficient						
Kansas	88	176	Below Basic	55	217	Basic	11	269	Advanced			
Kentucky	81	199	Below Basic	52	226	Basic	14	263	Proficient			
Minnesota	79	192	Below Basic	58	216	Basic	18	259	Proficient			
Nebraska	81	196	Below Basic	38	237	Basic						
New York	68	206	Below Basic	32	240	Proficient	11	267	Proficient			
North Carolina	77	201	Below Basic	59	218	Basic	47	228	Basic	7	275	Advanced
Oklahoma	85	188	Below Basic	70	205	Below Basic	4	279	Advanced			
Pennsylvania	87	185	Below Basic	59	219	Basic	22	255	Proficient			
Tennessee	88	172	Below Basic	45	224	Basic	14	261	Proficient			
Texas	74	194	Below Basic	21	247	Proficient						
Utah	69	208	Basic	42	233	Basic	13	267	Proficient			
Virginia	97	160	Below Basic	77	202	Below Basic	20	260	Proficient			
Wyoming	85	195	Below Basic	61	219	Basic	18	259	Proficient			

The results of NAEP benchmarks for ELA grade 8 individual states are reported in Table 16. The reading grade 8 NAEP achievement level cut-scores are Basic = 243, Proficient = 281, and Advanced = 323. The states with four achievement levels for which Level 3 maps to the NAEP Proficient level are Kansas and New York. The only state with five achievement levels for which Level 4 maps to the NAEP Proficient level is Florida.

Table 16: ELA Grade 8 NAEP Benchmarks for Nonconsortium States

State	ELA Grade 8											
	Level 2			Level 3			Level 4			Level 5		
	Percent at and Above Level 2	NAEP Equivalent for Level 2	NAEP Achievement Level Comparable to Level 2	Percent at and Above Level 3	NAEP Equivalent for Level 3	NAEP Achievement Level Comparable to Level 3	Percent at and Above Level 4	NAEP Equivalent for Level 4	NAEP Achievement Level Comparable to Level 4	Percent at and Above Level 5	NAEP Equivalent for Level 5	NAEP Achievement Level Comparable to Level 5
Alaska	80	228	Below Basic	31	279	Basic	2	337	Advanced			
Arizona	61	253	Basic	35	276	Basic	8	311	Proficient			
DoDEA	96	230	Below Basic	79	256	Basic	41	283	Proficient			
Florida	78	239	Below Basic	55	259	Basic	29	281	Proficient	11	303	Proficient
Georgia	76	238	Below Basic	39	272	Basic	8	312	Proficient			
Iowa	75	246	Basic	24	291	Proficient						
Kansas	78	241	Below Basic	29	285	Proficient	2	333	Advanced			
Kentucky	79	241	Below Basic	54	264	Basic	18	299	Proficient			
Minnesota	75	248	Basic	56	265	Basic	20	299	Proficient			
Nebraska	79	244	Basic	36	281	Basic						
New York	60	254	Basic	22	291	Proficient	7	317	Proficient			
North Carolina	79	231	Below Basic	53	257	Basic	42	269	Basic	10	309	Proficient
Oklahoma	87	226	Below Basic	75	241	Below Basic	16	295	Proficient			
Pennsylvania	89	225	Below Basic	58	262	Basic	15	306	Proficient			
Tennessee	91	221	Below Basic	50	265	Basic	11	306	Proficient			
Texas	78	234	Below Basic	23	286	Proficient						
Utah	66	256	Basic	42	276	Basic	15	304	Proficient			
Virginia	96	207	Below Basic	75	244	Basic	11	309	Proficient			
Wyoming	79	244	Basic	52	268	Basic	12	305	Proficient			

The results of NAEP benchmarks for mathematics grade 4 individual states are reported in Table 17. The mathematics grade 4 NAEP achievement level cut-scores are Basic = 214, Proficient = 249, and Advanced = 282. The only state with four achievement levels for which Level 3 maps to the NAEP Proficient level is Kansas. The only state with five achievement levels for which Level 4 maps to the NAEP Proficient level is Florida.

Table 17: Mathematics Grade 4 NAEP Benchmarks for Nonconsortium States

State	Mathematics Grade 4											
	Level 2			Level 3			Level 4			Level 5		
	Percent at and Above Level 2	NAEP Equivalent for Level 2	NAEP Achievement Level Comparable to Level 2	Percent at and Above Level 3	NAEP Equivalent for Level 3	NAEP Achievement Level Comparable to Level 3	Percent at and Above Level 4	NAEP Equivalent for Level 4	NAEP Achievement Level Comparable to Level 4	Percent at and Above Level 5	NAEP Equivalent for Level 5	NAEP Achievement Level Comparable to Level 5
Alaska	86	202	Below Basic	39	245	Basic	8	279	Proficient			
Arizona	72	220	Basic	42	244	Basic	10	276	Proficient			
DoDEA	88	218	Basic	66	237	Basic	39	255	Proficient			
Florida	77	222	Basic	59	236	Basic	31	256	Proficient	12	275	Proficient
Georgia	80	212	Below Basic	40	244	Basic	9	275	Proficient			
Iowa	79	219	Basic	29	260	Proficient						
Kansas	85	211	Below Basic	35	252	Proficient	8	282	Advanced			
Kentucky	80	219	Basic	49	243	Basic	16	270	Proficient			
Minnesota	85	217	Basic	70	233	Basic	36	261	Proficient			
Nebraska	77	223	Basic	24	263	Proficient						
New York	73	219	Basic	43	242	Basic	19	262	Proficient			
North Carolina	79	221	Basic	56	239	Basic	49	245	Basic	18	270	Proficient
Oklahoma	90	206	Below Basic	72	224	Basic	27	256	Proficient			
Pennsylvania	75	222	Basic	44	248	Basic	17	273	Proficient			
Tennessee	85	211	Below Basic	50	240	Basic	21	264	Proficient			
Texas	73	227	Basic	17	271	Proficient						
Utah	71	226	Basic	51	242	Basic	26	261	Proficient			
Virginia	97	193	Below Basic	84	218	Basic	28	263	Proficient			
Wyoming	88	214	Basic	51	246	Basic	13	278	Proficient			

The results of NAEP benchmarks for mathematics grade 8 individual states are reported in

Table 18. The mathematics grade 8 NAEP achievement level cut-scores are Basic = 262, Proficient = 299, and Advanced = 333. The only states with four achievement levels for which Level 3 maps to the NAEP Proficient level are Alaska, Kansas, New York, and Pennsylvania. The only state with five achievement levels for which Level 4 maps to the NAEP Proficient level is Florida.

In some states, some of the grade 8 students took the Algebra 1 test. In this benchmarking study, this factor could have had the effect of making the grade 8 mathematics standards appear higher.

Table 18: Mathematics Grade 8 NAEP Benchmarks for Nonconsortium States

State	Mathematics Grade 8											
	Level 2			Level 3			Level 4			Level 5		
	Percent at and Above Level 2	NAEP Equivalent for Level 2	NAEP Achievement Level Comparable to Level 2	Percent at and Above Level 3	NAEP Equivalent for Level 3	NAEP Achievement Level Comparable to Level 3	Percent at and Above Level 4	NAEP Equivalent for Level 4	NAEP Achievement Level Comparable to Level 4	Percent at and Above Level 5	NAEP Equivalent for Level 5	NAEP Achievement Level Comparable to Level 5
Alaska	89	236	Below Basic	26	304	Proficient	1	361	Advanced			
Arizona	59	275	Basic	34	298	Basic	14	323	Proficient			
DoDEA	96	236	Below Basic	78	267	Basic	45	295	Basic			
Florida	71	256	Below Basic	45	280	Basic	18	308	Proficient	7	328	Proficient
Georgia	75	254	Below Basic	37	291	Basic	12	321	Proficient			
Iowa	75	262	Below Basic	24	312	Proficient						
Kansas	62	274	Basic	22	309	Proficient	4	344	Advanced			
Kentucky	85	242	Below Basic	44	283	Basic	11	320	Proficient			
Minnesota	80	264	Basic	58	287	Basic	27	317	Proficient			
Nebraska	68	270	Basic	22	313	Proficient						
New York	60	271	Basic	22	308	Proficient	7	334	Advanced			
North Carolina	70	262	Below Basic	43	288	Basic	36	295	Basic	11	328	Proficient
Oklahoma	79	248	Below Basic	53	272	Basic	11	314	Proficient			
Pennsylvania	62	271	Basic	30	304	Proficient	8	338	Advanced			
Tennessee	81	246	Below Basic	54	274	Basic	29	299	Basic			
Texas	75	261	Below Basic	6	336	Advanced						
Utah	70	267	Basic	41	294	Basic	14	325	Proficient			
Virginia	93	235	Below Basic	74	265	Basic	9	336	Advanced			
Wyoming	84	255	Below Basic	47	289	Basic	10	327	Proficient			

Comparing Achievement Standards for Smarter Balanced, PARCC, and ACT Aspire

One of the advantages of mapping state achievement standards to NAEP is that the NAEP scale can serve as a common metric with which to compare the achievement standards of Smarter Balanced, PARCC, and ACT Aspire. The strategy is to obtain the NAEP equivalent of each consortium achievement standard and then compare their NAEP equivalents. The procedure used in this report is to compare their NAEP equivalents by using a two-tailed Z test with $p < .05$. The standard error used in the Z test is described in the Appendix.

The most important comparisons are between the college-ready standards of the group assessments. Comparing Smarter Balanced versus PARCC in Table 19 we find that

1. Smarter Balanced college-ready standards (Level 3) are comparable in difficulty to the NAEP Basic levels, and
2. Smarter Balanced college-ready standards (Level 3) are significantly below PARCC college-ready standards (Level 4) by about one-quarter of a standard deviation.

Table 19: Smarter Balanced Versus PARCC

		Smarter Balanced			PARCC			Difference	
Subject	Grade	NAEP Equivalent Level 3 Met	Standard Error NAEP Equivalent	NAEP Achievement Level	NAEP Equivalent Level 4 Met	Standard Error NAEP Equivalent	NAEP Achievement Level	Significant Difference ($p < .05$)	Effect Size Difference Smarter Minus PARCC
ELA	4	222	2	Basic	232	1	Basic	YES	−.26
ELA	8	264	1	Basic	273	1	Basic	YES	−.28
Math	4	244	1	Basic	252	1	Proficient	YES	−.26
Math	8	294	1	Basic	307	1	Proficient	YES	−.36

We can also compare the achievement standards of Smarter Balanced to those of ACT Aspire. When we compare the college-ready standards in Table 20, we find that

1. both Smarter Balanced and ACT Aspire college-ready standards (Ready) are comparable in difficulty to the NAEP Basic level; and
2. Smarter Balanced college-ready grade 8 standards are statistically comparable to ACT Aspire college-ready grade 8 standards. However, for grade 4, the Smarter Balanced college-ready standard is significantly below the ACT Aspire college-ready standard for ELA and reading (effect size = −.26) but significantly above the ACT Aspire college-ready standard for mathematics (effect size = +.29).

Table 20: Smarter Balanced Versus ACT Aspire

		Smarter Balanced			ACT Aspire			Difference	
Subject	Grade	NAEP Equivalent Level 3 Met	Standard Error NAEP Equivalent	NAEP Achievement Level	NAEP Equivalent Level 4 Met	Standard Error NAEP Equivalent	NAEP Achievement Level	Significant Difference ($p < .05$)	Effect Size Difference Smarter Minus ACT Aspire
ELA/ Reading	4	222	2	Basic	232	2	Basic	YES	−.26
ELA/ Reading	8	264	1	Basic	264	1	Basic	NO	
Math	4	244	1	Basic	235	1	Basic	YES	.29
Math	8	294	1	Basic	290	2	Basic	NO	

Similarly, we can compare PARCC and ACT Aspire college-ready standards. From Table 21, PARCC college-ready standards (Level 4) are statistically comparable in difficulty to the ACT Aspire college-ready standard for ELA and reading grade 4. However, PARCC standards are significantly above ACT Aspire college-ready standards for ELA and reading grade 8 (effect size = +.28), mathematics grade 4 (effect size = +.55), and mathematics grade 8 (effect size = +.48).

Table 21: PARCC Versus ACT Aspire

		PARCC			ACT Aspire			Difference	
Subject	Grade	NAEP Equivalent Level 4 Met	Standard Error NAEP Equivalent	NAEP Achievement Level	NAEP Equivalent of Ready Level	Standard Error NAEP Equivalent	NAEP Achievement Level	Significant Difference ($p < .05$)	Effect Size Difference PARCC Minus ACT Aspire
ELA/ Reading	4	232	1	Basic	232	2	Basic	NO	
ELA/ Reading	8	273	1	Basic	264	1	Basic	YES	.28
Math	4	252	1	Proficient	235	1	Basic	YES	.55
Math	8	307	1	Proficient	290	2	Basic	YES	.48

Conclusion

There are essentially three overall findings in this study.

1. A handful of nonconsortium states have college-ready standards that are at least as stringent as the NAEP Proficient level. These are
    a. ELA grade 4—Florida and New York;
    b. ELA grade 8—Florida, Kansas, and New York;
    c. Mathematics grade 4—Florida and Kansas; and
    d. Mathematics grade 8—Alaska, Florida, Kansas, New York, and Pennsylvania.
2. For the group-based assessments, only PARCC mathematics, grades 4 and 8, have college-ready standards comparable in difficulty to the NAEP Proficient level.
3. The Smarter Balanced achievement standards are about one-quarter of a standard deviation lower than the PARCC performance standards.

The benchmarking study reported here should give policy makers insight into what states are expecting from their students. Some states expect more, and some expect less. The study is intended to provide a way to benchmark and compare state achievement standards and benchmark and compare the achievement standards of Smarter Balanced, PARCC, and ACT Aspire. The study does not intend to evaluate state achievement standards or make policy recommendations.

Caveats

There are several caveats that are important to note in this study. First, the results in this report do not provide final and complete information about each state. The author was unable to obtain the results for several states, and some states have reported their results as preliminary. In the future, the National Center for Education Statistics (NCES) will conduct their biennial state mapping study. By that time, the NCES should be able to provide a more definitive and comprehensive mapping study.

Second, in some states, some of the grade 8 mathematics students took an end-of-course test, such as Algebra 1. In this benchmarking study, this factor could have had the effect of making the state grade 8 mathematics standards appear higher.

Third, this study maps state achievement standards to NAEP achievement levels and highlights those state standards that reach the NAEP Proficient level. This should not be interpreted to mean that NAEP's Proficient levels in grades 4 and 8 are the gold standards for deciding whether our students are on track to be ready for college. No evidence has been presented by NAEP that the proficient standard in grades 4 and 8 predicts college success. It is the case that NAEP used 12th grade college-ready cut-scores (2013) to report that about 38% of students have the reading skills, and 39% have the math skills that make them ready for college. The cut-scores were 302 for reading and 163 for mathematics. The reading college-ready cut-score was equal to the reading proficient standard, and the mathematics cut-score was just below the mathematics proficient standard.

Fourth, there are some interpretive nuances related to the methodology used in this study. This report uses statistical linking to map state achievement standards onto the NAEP scale. Holland (2007) has outlined three broad categories of linking. These are equating, scale alignment, and prediction. A fundamental difference among the three methods is related to the degree to which they assume the two tests measure the same content and have the same administrative procedures.

- In equating, both tests must be constructed to measure the same identical content, be equally reliable, and both tests must use the same administrative procedures.
- In scale alignment, both tests measure similar but not identical content, may not be equally reliable, and there can be variation in administrative procedures. Scale alignment can provide a good ballpark estimate of how scores line up, but is less precise than equating.
- In prediction, there are no assumptions at all about content, reliability or administrative procedures.

This report uses the second type: scale alignment. The scales we are aligning will not measure identical constructs[1], will not be equally reliable, and will not use identical administrative

[1] A recent study for mathematics by the NAEP Validity Study (NVS) panel found that 79% of NAEP items were matched to content in the CCSS in the 4th grade and 87% in the 8th grade (Daro, Hughes and Stancavage, 2015).

procedures. The method of alignment is equiprecentile linking based on the aggregate reporting of NAEP and the state assessments. It is the scales of the total aggregate distributions that are aligned, so the linking should not be used for disaggregated reporting of individual students or demographic subgroups (such as race/ethnicity or gender) or subpopulations (such as schools). Also, the reader should be aware that the concordance between NAEP and the state assessments established in this report for 2015 may not be applicable in subsequent years.

Fifth, this report does not, in any way, address or evaluate the quality of the CCSS. The CCSS are *content* standards, while this report deals only with *achievement* standards. Content standards represent the curriculum that teachers should teach, and the scope and sequence of what students should learn in school. Achievement standards are cut-scores on the state test that represent performance expectations. For example, what level of performance on the test do we think represents being on track to be college ready.

References

Daro, P., Hughes, G. B., and Stancavage, F. (2015). Study of the Alignment of the 2015 NAEP Mathematics Items at Grade 4 and 8 to the Common Core State Standards (CCSS) for Mathematics. (see http://www.air.org/project/naep-validity-studies-nvs-panel)

Holland, P. W. (2007). A framework and history for score linking. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and aligning scores and scales*, New York, NY: Springer.

National Governors Association, Council of Chief State School Officers, Achieve. (2008). *Benchmarking for success: Ensuring U.S. students receive a world-class education*. Washington, DC: National Governors Association.

NAEP as an Indicator of Students' Academic Preparedness for College. U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2013 Mathematics and Reading Assessments.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).

Obama, B. (2009, March 10). President Obama's remarks to the Hispanic Chamber of Commerce. *New York Times*.

Appendix: Methodology

This study uses equipercentile linking to benchmark state achievement standards against NAEP achievement levels. The derivations described below make two assumptions. First, we assume the state test scores and the NAEP test scores are normal distributions. Second, we assume the NAEP examinee sample is randomly equivalent to the population of examinees who took the state test.

NAEP scores are assumed to have a normal distribution $N(\hat{\mu}_N, \hat{\sigma}_N^2)$, where the standard error of $\hat{\mu}_N$ is estimated by $\hat{\sigma}_{\hat{\mu}_N}$, the standard error of $\hat{\sigma}_N$ is $\hat{\sigma}_{\hat{\sigma}_N}$, and the covariance between $\hat{\mu}_N$ and $\hat{\sigma}_N$ is $\hat{\sigma}_{\hat{\mu}_N,\hat{\sigma}_N}$, usually 0 if from a normal sample.

If the state-level proportion at and above the cut c is $\hat{p}_c$ with standard error of $\hat{\sigma}_{\hat{p}_c}$, the corresponding NAEP equivalent score, $\hat{s}_N$ assuming random equivalent group tests, can be estimated by solving the equation

$$1 - \hat{p}_c = \int_{-\infty}^{\hat{s}_N} \frac{Exp\left(-\frac{(x-\hat{\mu}_N)^2}{2\hat{\sigma}_N^2}\right)}{\sqrt{2\pi}\hat{\sigma}_N} dx.$$

Let $y = \frac{x-\hat{\mu}_N}{\hat{\sigma}_N}$, and making the change of variable, we obtain

$$1 - \hat{p}_c = \int_{-\infty}^{\frac{\hat{s}_N-\hat{\mu}_N}{\hat{\sigma}_N}} \frac{Exp\left(-\frac{y^2}{2}\right)}{\sqrt{2\pi}} dy,$$

or

$$\frac{\hat{s}_N - \hat{\mu}_N}{\hat{\sigma}_N} = \Phi^{-1}(1 - \hat{p}_c).$$

So

$$\hat{s}_N = \hat{\mu}_N + \hat{\sigma}_N \Phi^{-1}(1 - \hat{p}_c).$$

Using delta method, the variance of the NAEP equivalent score $\hat{s}_N$ can be estimated by

$$Var(\hat{s}_N) = Var(\hat{\mu}_N) + 2\Phi^{-1}(1-\hat{p}_c)Cov(\hat{\mu}_N, \hat{\sigma}_N) + Var(\hat{\sigma}_N)\left(\Phi^{-1}(1-\hat{p}_c)\right)^2 + \hat{\sigma}_N^2 Var\left(\Phi^{-1}(1-\hat{p}_c)\right),$$

or

$$Var(\hat{s}_N) = \hat{\sigma}_{\hat{\mu}_N}^2 + 2\Phi^{-1}(1-\hat{p}_c)\hat{\sigma}_{\hat{\mu}_N,\hat{\sigma}_N} + \left(\Phi^{-1}(1-\hat{p}_c)\right)^2 \hat{\sigma}_{\hat{\sigma}_N}^2 + \hat{\sigma}_N^2 \left(\varphi\left(\Phi^{-1}(1-\hat{p}_c)\right)\right)^{-2} \hat{\sigma}_{\hat{p}_c}^2.$$

The standard error of the NAEP equivalent score $\hat{s}_N$ is then estimated by

$$\hat{\sigma}_{\hat{s}_N} = \sqrt{\hat{\sigma}^2_{\hat{\mu}_N} + 2\Phi^{-1}(1-\hat{p}_c)\hat{\sigma}_{\hat{\mu}_N,\hat{\sigma}_N} + \left(\Phi^{-1}(1-\hat{p}_c)\right)^2\hat{\sigma}^2_{\hat{\sigma}_N} + \hat{\sigma}^2_N\left(\varphi\left(\Phi^{-1}(1-\hat{p}_c)\right)\right)^{-2}\hat{\sigma}^2_{\hat{p}_c}}$$

where Φ is the cumulative distribution function of the standard normal distribution and φ is the probability density function of the standard normal distribution. Assuming $Cov(\hat{\mu}_N, \hat{\sigma}_N) = 0$, this is simplified to

$$\hat{\sigma}_{\hat{s}_N} = \sqrt{\hat{\sigma}^2_{\hat{\mu}_N} + \left(\Phi^{-1}(1-\hat{p}_c)\right)^2\hat{\sigma}^2_{\hat{\sigma}_N} + \hat{\sigma}^2_N\left(\varphi\left(\Phi^{-1}(1-\hat{p}_c)\right)\right)^{-2}\hat{\sigma}^2_{\hat{p}_c}}.$$

The values of $\hat{\sigma}_{\hat{s}_N}$ were rounded up to the nearest NAEP scaled score unit.

For Smarter Balanced, PARCC, and ACT Aspire, the aggregate state-level proportion at and above the cut c is $\hat{p}_c$ with standard error $\hat{\sigma}_{\hat{p}_c}$ and was based on the weighted average of the states and jurisdictions within the consortium. The weights were the population sizes within each state. For Smarter Balanced, PARCC, and ACT Aspire, the state NAEP aggregate scores $\hat{s}_N$ were estimated with the NCES NDE (http://nces.ed.gov/nationsreportcard/naepdata/).

- Aggregate Smarter Balanced results are based on the weighted average of 18 states: California, Connecticut, Delaware, Hawaii, Idaho, Maine, Michigan, Missouri, Montana, Nevada, New Hampshire, North Dakota, Oregon, South Dakota, Vermont, Washington, West Virginia, and Wisconsin. For ELA, Missouri and Wisconsin were excluded because they did not follow the Smarter Balanced blueprint.
- Aggregate PARCC results are based on the weighted average of 12 jurisdictions: Arkansas, Colorado, District of Columbia, Illinois, Louisiana, Maryland, Massachusetts, Mississippi, New Jersey, New Mexico, Ohio, and Rhode Island. In grade 8 mathematics, in some PARCC states, some students took the Algebra 1 test. In the mapping study, this factor could have had the effect of making the grade 8 mathematics PARCC standards appear higher.
- Aggregate ACT Aspire results are based on the weighted average of two jurisdictions: Alabama and South Carolina.

AMERICAN INSTITUTES FOR RESEARCH®

1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
202.403.5000 | TTY: 877.334.3499

www.air.org

Making Research Relevant