

Maintaining the Validity of the National Assessment of Educational Progress in a Common Core Based Environment

Peter Behuniak

August 2015

Commissioned by the NAEP Validity Studies (NVS) Panel

George W. Bohrnstedt, Panel Chair

Frances B. Stancavage, Project Director

The NAEP Validity Studies Panel was formed by the American Institutes for Research under contract with the National Center for Education Statistics. Points of view or opinions expressed in this paper do not necessarily represent the official positions of the U.S. Department of Education or the American Institutes for Research.

The NAEP Validity Studies (NVS) Panel was formed in 1995 to provide a technical review of NAEP plans and products and to identify technical concerns and promising techniques worthy of further study and research. The members of the panel have been charged with writing focused studies and issue papers on the most salient of the identified issues.

Panel Members:

Peter Behuniak
University of Connecticut

George W. Bohrnstedt
American Institutes for Research

James R. Chromy
Research Triangle Institute

Phil Daro
University of California, Berkeley

Richard P. Durán
University of California, Santa Barbara

David Grissmer
University of Virginia

Larry Hedges
Northwestern University

Gerunda Hughes
Howard University

Ina V.S. Mullis
Boston College

Scott Norton
Council of Chief State School Officers

James Pellegrino
University of Illinois at Chicago

Gary Phillips
American Institutes for Research

Lorrie Shepard
University of Colorado at Boulder

David Thissen
University of North Carolina, Chapel Hill

Karen Wixson
University of North Carolina at Greensboro

Project Director:

Frances B. Stancavage
American Institutes for Research

Project Officer:

Grady Wilburn
National Center for Education Statistics

For Information:

Frances Stancavage
NAEP Validity Studies (NVS)
American Institutes for Research
2800 Campus Drive, Suite 200
San Mateo, CA 94403
Phone: 650/ 843-8163
E-mail: fstancavage@air.org

Acknowledgments

The author is grateful to Melissa Deblasio for her help in conducting background research and compiling data for use in this paper. In addition, I thank Frances Stancavage for her thoughtful editing, William Tirre for providing useful documentation regarding NAEP, and Jeffrey Nellhaus for his willingness to share timely details about the current actions and future plans of the PARCC assessment consortium. Finally, I appreciate the many constructive comments and suggestions offered by the members of the NVS Panel.

CONTENTS

Section 1: Purpose	1
Section 2: Context	2
Section 3: NAEP Content Compared to CCSS-Based Assessments	6
Section 4: Discussion of Technical Considerations	9
Section 5: Discussion of Policy Considerations	17
References	19

Section 1: Purpose

The creation and widespread adoption of the Common Core State Standards (CCSS) has ushered in the most significant change in education policy since No Child Left Behind (NCLB). Given the level of investment in this initiative, there will be great interest in tracking its impact on student achievement as reflected in the assessments developed to align with the CCSS. However, throughout this transition to the CCSS, the National Assessment of Educational Progress (NAEP) will surely continue to be a major indicator that will be of interest to educators and policy makers. This white paper examines the issues associated with employing NAEP in the role of monitoring student achievement during a period when states are increasingly implementing CCSS-based assessments.

The focus of the paper is to bring both policy and psychometric considerations to bear in discussing the relationship of NAEP and CCSS-based assessments. The discussion will examine how the traditional roles of NAEP are affected by the introduction of CCSS-based curricula and assessments. The paper will attempt to identify caveats or limitations that may be necessary to consider when interpreting NAEP results over the next several years. The examination of how NAEP will function in a CCSS-based world is organized in five sections:

1. Purpose
2. Context
3. NAEP Content Compared to CCSS-Based Assessments
4. Discussion of Technical Considerations
5. Discussion of Policy Issues

Much attention has already been paid to the potential implications for NAEP of the introduction of the CCSS and CCSS-based assessments. This has taken the form of discussions, white papers, studies, and pilot activities. In some cases, the studies or other activities have been completed or are underway, while in other cases the activities are planned or under consideration for future implementation. This paper will reference these activities or plans as they relate to each topic. Some of the completed white papers and studies have proposed recommendations for NAEP that bear on the focus of this paper. When these recommendations are discussed, there will be an attempt to highlight the relevance of each recommendation to the maintenance of NAEP's validity. Suggestions will be offered where possible to expand or enhance these recommendations or to propose additional recommendations that can help to address the challenges created by the shift toward CCSS-based assessments.

Section 2: Context

NAEP has long served the field of education and the public by addressing multiple purposes. Two of these purposes are particularly relevant to the validity of NAEP in a CCSS-focused era. First, as the only nationwide measure of student achievement, the periodic administrations of NAEP have provided snapshots of how well students across the nation are doing in mathematics, reading, and other content areas. This role has earned NAEP the moniker of the nation's report card. Second, NAEP has tracked changes in student achievement over time, establishing whether trends in student performance are positive, neutral, or negative.

Both of these roles apply at the state level as well as the national level, and this might also become important in judging the implications of the CCSS for NAEP. The state NAEP component provides a measure of how well students are learning in each state. NAEP also provides a common metric across states. This function of NAEP could be significantly affected by states' responses to the CCSS.

A few years ago, these roles of NAEP looked like they might become obsolete. The adoption of the CCSS by 45 states was accompanied by the creation of two huge consortia, the Smarter Balanced Assessment Consortium (SBAC) and the Partnership for Assessment of Readiness for College and Careers (PARCC), which were dedicated to the development of CCSS-based assessments. The plan was that participant states would quickly implement these assessments and rely on them to monitor student achievement on their CCSS-based curricula. The role(s) of NAEP in this scenario was uncertain given the reality that NAEP is based not on the CCSS, but on NAEP frameworks. While NAEP frameworks will be discussed in more detail later in this paper, it is prudent to acknowledge up front that these frameworks were deliberately designed to be broad statements intended to guide the development of NAEP and created to be independent of any particular curriculum.

It is important to note that states' adoption of the CCSS as the standards on which to focus instruction and their determination of whether or not to participate in SBAC or PARCC have proven to be independent decisions. In 2010, it appeared that almost all of the states would first move to adopt the CCSS and then select one of the consortia to obtain access to the CCSS-based assessments. However, a variety of educational, political, and financial factors have led states to rethink their commitments to the CCSS, the consortium assessments developed by SBAC and PARCC, or both.

This atmosphere of shifting perspectives means that the prospect of having the vast majority of states administer one or two common assessments is far less certain today than it was a few years ago. Education Week (Gewertz & Ujjifusa, 2014) reported that the number of states committed to administering SBAC or PARCC assessments has declined steadily from a peak of 45. In fact, as the test administration window for the

first operational assessments opened in March 2015,¹ the states' commitments to CCSS and the CCSS-based assessments continued to change.

The status of each of the 50 states and the District of Columbia as of March 2015 is shown in Table 1 (Gewertz, 2015; Academic Benchmarks, 2015).

Table 1: State Affiliation With the Common Core and Assessment Consortia as of the Start of the First Operational Testing Window (March 2015)

	CCSS States			Non-CCSS States
	SBAC (18)	PARCC (11)	Other (15)	Other (7)
California	Arkansas	Alabama	Alaska	
Connecticut	Colorado	Arizona	Indiana ³	
Delaware	Washington, DC	Florida	Nebraska	
Hawaii	Illinois	Georgia	Oklahoma ⁴	
Idaho	Louisiana	Iowa	South Carolina ⁴	
Maine	Maryland	Kansas	Texas	
Michigan	Mississippi	Kentucky	Virginia	
Missouri	New Jersey	Massachusetts ¹		
Montana	New Mexico	Minnesota ²		
North Dakota	Ohio	North Carolina		
New Hampshire	Rhode Island	New York		
Nevada		Pennsylvania		
Oregon		Tennessee		
South Dakota		Utah		
Vermont		Wyoming		
Washington				
Wisconsin				
West Virginia				
Approximate Percentage of Student Population	28%	18%	35%	18%

¹ Fifty-four percent of districts will take PARCC in 2015; 46 percent will take the Massachusetts Comprehensive Assessment System (MCAS) (Massachusetts Department of Elementary & Secondary Education, 2015).

² Adopted English Language Arts standards only (CCSS Initiative, 2015)

³ Formally withdrew from standards, March 2014 (Fineout & Talley, 2014)

⁴ Repealed standards, June 2014 (Associated Press, 2014; Ujifusa, 2014)

¹ PARCC split its testing window for 2015. The window for performance-based assessments began in mid-February and the window for end of year assessments began in mid-April. In 2016 PARCC plans to have one window that starts at the beginning of March. SBAC had a single testing window in 2015, starting at the beginning of March, and intends to keep that schedule in the future.

There are three observations about these data that are relevant to NAEP. First, the number of jurisdictions administering either the SBAC or PARCC assessments has dropped from 45 to 29. Certainly 29 jurisdictions, including approximately 46 percent of the nation's public school population, represents a large number of students. However, this number of participating states, and the fact that they are a self-selected sample, means they will be insufficient to support the interpretation of the results of administrations of the SBAC and PARCC assessments as indications of the national status of student achievement. Even if all states remaining committed to the CCSS were administering one of the two consortium assessments, there would be concerns about the degree to which the results would provide adequate evidence of nationwide achievement levels due to the potential lack of representativeness of the participating states. With only slightly more than half of the states participating (and less than half of the student population), inferring national achievement levels on the basis of SBAC and PARCC results will not be possible under the current circumstances.

The second observation is that the number of states that have not adopted or have reversed their adoption of the CCSS as of March 2015 has grown from four to seven. This is important to NAEP for at least two reasons. First, the number of students in these states is large, and represents about 18 percent of the country's student population. Second, even the small increase in the number of states seeking to distance themselves from the CCSS could be an indication of further reversals in the future.

The third observation regarding Table 1 has to do with the 15 states that remain committed to the CCSS but have chosen to opt out of the SBAC and PARCC consortia. These states will either be building their own assessments, purchasing commercial assessments that are CCSS-based, or employing a hybrid model that blends these options depending on the grade and content area. This is a scenario that is similar to the situation that existed under NCLB, where the assessments and associated performance standards in use across state lines varied considerably, making state-to-state comparisons difficult.

These developments have profound implications for NAEP. The reduction of state participation in SBAC and PARCC, combined with the increasing discontent with the CCSS, significantly increases the likelihood that NAEP will continue to serve as the nation's report card for the foreseeable future. The administrative and psychometric effort that would be necessary to have the consortia draw nationally representative samples of students or to link SBAC and PARCC assessments with other measures that would enable them to take on this reporting function does not appear likely at this time.

The other role of NAEP, monitoring trends in student performance, also appears secure. Since most states are administering new assessments, whether or not they are members of one of the consortia, it will be years before the results from these new instruments form a trend. Further, once it is at least possible to examine the trend in student performance on consortium instruments (in 2017 and beyond), it is not clear what inferences will be justified. It is likely that intrastate inferences (e.g., determining the trend of a state's performance) will be appropriate, assuming certain assumptions are met. These would include the state's continued adherence to the

CCSS, application of consistent inclusion/exclusion rules, and other basic procedures typically employed to maintain alignment and standardization. However, interstate inferences (e.g., comparing the status or trends across states) could be much more difficult to support. A significant factor that will affect the viability of both of these types of inference is the degree to which states continue to modify their allegiance to the consortium assessments.

Recommendation 1. It is not possible to predict how many states will adjust their focus on the CCSS and their participation in SBAC and PARCC consortia. Since these sorts of changes have implications for NAEP, it is recommended that states' involvement with the CCSS and the related assessments be monitored closely. Annual reviews of the states' status should be useful in determining whether the CCSS and consortium assessments are continuing to lose support or if the acceptance of the CCSS and the associated assessments stabilizes.

Recommendation 2. Conduct a study that examines the specific content of each state's content standards in terms of their similarity to the CCSS. Some states may claim to have independent standards but, in fact, have standards that are very similar to the CCSS. Other states may formally adhere to the CCSS, but modify the standards to a degree that is significant. It would be informative to generate an independent analysis of content standards across states to better understand their relationship to the CCSS.

Section 3: NAEP Content Compared to CCSS-Based Assessments

The CCSS were developed beginning in 2009 in an initiative led by the National Governors' Association (NGA) and the Council of Chief State School Officers (CCSSO) (NGA & CCSSO, 2010). The CCSS, which cover mathematics and English language arts/literacy, have received a great deal of scrutiny from educators across the country. At least in states that continue to support the CCSS, these standards will form the basis for implementing curricula, instruction, and assessments in the coming years.

NAEP instruments are based on NAEP frameworks in mathematics, reading, writing, and other content areas. These frameworks were developed by content experts under the direction of the National Assessment Governing Board. The frameworks are periodically revised and are specifically intended to guide the development of the NAEP instruments. The NAEP frameworks are not intended to be the basis of particular curricula but, instead, are created to fairly measure achievement among students who have been exposed to a variety of curricula and instructional approaches.

Given these differences in origin and purpose, it is to be expected that there would be substantive differences between the NAEP frameworks and the corresponding sections of the CCSS. The primary question of interest is whether these differences are so great as to preclude the use of NAEP as an independent monitor of student achievement in environments in which CCSS-based curricula and assessments are implemented.

Two studies were initiated by the NAEP Validity Studies (NVS) Panel to investigate this question. In the first study, the authors conducted a thorough comparison of the NAEP framework and the CCSS in mathematics (Hughes, Daro, Holtzman, & Middleton, 2013). The second study performed the same examination in reading and writing (Wixson, Valencia, Murphy, & Phillips, 2013). The purpose of this paper is not to reiterate all of the findings of the studies. Instead, the emphasis in the following section is on summarizing the overall findings and discussing the implications for NAEP, with particular attention paid to any threats to the validity of using NAEP as a nationwide monitor of student achievement in an era in which many states are focusing on the CCSS.

The mathematics study conducted by Hughes et al. (2013) focused on the conceptual match between the subtopics and objectives in the NAEP Mathematics Framework and the content standards in the CCSS for Mathematics (CCSS-M) in grades 4 and 8. In order to achieve the best possible understanding of the areas of overlap and any dissimilarity between these two statements of content, the methodology consisted of performing the analysis in both directions: first mapping the CCSS-M to the NAEP Mathematics Framework and then mapping the NAEP Mathematics Framework to the CCSS-M. The key findings were reported as follows:

The study did not find wide areas of content in the NAEP Mathematics Framework that were not covered in the CCSS-M. Similarly, the study did not find wide areas of content in the CCSS-M that were not covered by the NAEP Mathematics Framework. Nevertheless, there were differences in specificity and conceptual understandings between the CCSS-M and the

NAEP Mathematics Framework that are important to note: (1) the CCSS-M have more rigorous content in eighth-grade algebra and geometry; (2) the CCSS-M infuse and distribute the development of mathematical expertise, such as the ability to estimate accurately, throughout the standards for mathematical content, whereas the NAEP Mathematics Framework assesses estimation as a skill in isolation from the vast majority of the content; (3) the CCSS-M attend to developing conceptual understandings of a greater number of mathematical topics (such as unit fractions, patterns, and functions) than does the NAEP Mathematics Framework; and (4) the CCSS-M introduce some mathematics content, such as probability, at higher grades than does the NAEP Mathematics Framework. (pp. 11–12)

These results are encouraging. The content included in the NAEP frameworks was never intended to mirror exactly any set of curriculum standards and, as the study authors pointed out, the CCSS-M are curriculum standards, while the NAEP frameworks are intended to guide assessment development. The design of the NAEP frameworks to broadly cover the content area is what allows NAEP to monitor achievement for students who are engaged in a variety of instructional approaches aligned with many different sets of curriculum standards. Accordingly, the results of the Hughes et al. study support the validity of continuing to use NAEP to monitor achievement in mathematics for students in CCSS-based educational systems.

The Wixson et al. (2013) study also focused on identifying the ways in which the NAEP Reading and Writing Frameworks overlapped with or were dissimilar to the CCSS for English Language Arts (CCSS-ELA).² The reading study examined (1) range of text types, (2) quality of text, and (3) text complexity. In addition to the NAEP Writing Framework, the writing study compared information in the CCSS-ELA to NAEP scoring guides, anchor papers, and prompts in order to better capture the character of the NAEP writing assessment. The study authors offered the following observation as part of their conclusions:

The Reading and Writing Panel members recognize the different purposes of NAEP and CCSS-ELA and feel strongly that NAEP should retain its independence from any particular curriculum and serve as a general assessment of reading and writing performance. Overall, the panels are cautiously optimistic that, with attention to the specific issues identified in this report and a systematic program of special studies to inform future assessments, NAEP could continue to serve as an independent monitor of student achievement in an era of CCSS. (p. 4)

The study also identified 12 specific ways in which the NAEP Reading Framework was similar to or different from the CCSS-ELA and 5 ways in which the NAEP Writing Framework was alike or different from the CCSS-ELA. The findings are generally similar to the results for mathematics. That is, there are differences between

² Note that the CCSS for Listening and Speaking were not included in the study since NAEP does not assess these areas.

the NAEP frameworks and the CCSS, but there is also a significant overlap. It is worth noting again that NAEP frameworks are not intended to be curriculum standards and that NAEP has a long history of successfully measuring academic achievement for students receiving instruction aligned with varying curricula. The main risk in the current situation is that, as many states focus their instruction on the CCSS, NAEP may underestimate student learning in certain under-covered areas of content. That is, students in CCSS states may make academic progress in areas to which NAEP is less sensitive.

It is reasonable to conclude that the differences between the NAEP frameworks and the CCSS as noted in these studies are not so large as to prevent NAEP from validly continuing to fill its traditional role of monitoring student achievement. However, more effort is necessary to ensure that the validity of NAEP for measuring student learning remains high across all states, regardless of the different curricular choices made by the states. Thus there are two areas in which further investigation is warranted.

Recommendation 3. The content differences between the NAEP frameworks and the CCSS should be investigated to determine the effects, if any, on the scores produced by NAEP and CCSS-based assessments. This recommendation is consistent with the conclusions reached by the authors of both content studies. The investigation should be designed to identify any parts of the CCSS that are underrepresented on NAEP and which might therefore lead to NAEP underestimating student learning. One goal of this investigation would be to determine whether any of the content differences justify consideration of adjustments to the NAEP frameworks.

Recommendation 4. An examination of the items contained in NAEP and the consortium assessments should be conducted to identify areas of similarity and difference. The consortium-developed assessment items were not available for inspection at the time that the aforementioned content studies were underway. The item-level comparisons should be performed in both directions—mapping NAEP items onto consortium measures and mapping the consortium measures onto NAEP. The suggested item-level comparisons can be useful in at least two ways. First, these comparisons can provide more specific, tangible evidence of differences or similarities in the ways that NAEP and CCSS-based assessments are measuring the constructs of mathematics, reading, and writing. Second, item-level analyses can reveal if there are any areas in which measurement differences exist between NAEP and CCSS-based assessments even though the items are considered aligned with the same (or very similar) content standard. That is, even when a NAEP framework and the CCSS overlap, the test items created to measure the concept or skill could be significantly different.

Section 4: Discussion of Technical Considerations

It is clear that NCES needs to take steps to ensure that NAEP maintains its high technical and psychometric standards regardless of what happens with state adoptions of CCSS-based assessments across the country. As discussed earlier, many states' plans have shifted over the past two years. This shifting of states' affiliations is likely to continue in the coming years given changes in prevailing educational, political, and economic circumstances. Therefore, the planning of studies or other activities intended to preserve NAEP standards must take into account this uncertain and changing environment.

There are at least four potential threats to the validity of interpreting NAEP results as an indicator of student achievement in a CCSS-based educational environment. This section will discuss each of these and suggest actions that might be taken to address these threats.

Content and Item Differences. The basic threats caused by differences between the content and items used in NAEP and in CCSS-based assessments were discussed earlier. These include content differences between the NAEP frameworks and the CCSS as well as differences in the items used in NAEP and the CCSS-based assessments. Recommendations 2 and 3 were intended to address these issues.

There is an additional concern related to the coverage of the mathematics, reading, and writing constructs by NAEP and the CCSS. States have grown accustomed to operating their testing programs with some latitude, as allowed by the U.S. Department of Education. This flexibility, customarily referred to as the 15 percent rule, is reported in the Federal Register (2010) as follows:

A State may supplement the common standards with additional standards provided that the additional standards do not exceed 15 percent of the State's total standards for that content area. (p. 19499)

This flexibility was originally offered in connection with the waiver options implemented by the Department of Education that allowed states to avoid the constraints of NCLB. States not only will continue to make use of this option, they will do so for different reasons. For example, one state may remain committed to the CCSS but use the 15 percent rule to incorporate one or more elements of content the state believes is missing or underemphasized in the CCSS. Another state may make a similar adjustment to its target content but drop the CCSS label and use the additional content to explain why that state does not wish to adopt the CCSS. While these varying state responses may be largely the product of political considerations, the modification of states' target content has educational implications for instruction and assessment.

The specific implications of the 15 percent rule for NAEP depend on the degree to which states employ it. If many states that adopt the CCSS adjust their content standards to the maximum extent allowed, the variation across states could be considerable. It is possible the variation would begin to resemble interstate content differences that existed under NCLB. However, if most states make only minor (or no) adjustments to their CCSS-based content standards, then the rule should have

little effect on either instruction or the assessment being employed. This is a development that bears watching.

Recommendation 5. The trend in states' use of the 15 percent rule to adjust the CCSS should be monitored over the next several years. The purpose is to determine which states have used this rule and the degree to which their effective content standards (i.e., the actual CCSS or the modified versions) vary across these states. It would then be possible to determine how these varied standards compare to NAEP frameworks.

Performance Standards. The percentage of students who attain specific performance levels is the primary reporting mechanism for many audiences of large-scale student assessment results. This has been true for many years at the state, national, and international levels. Usually, the performance levels created for an assessment are considered unique. That is, even if the label used on two tests is the same (e.g., Proficient), there is no expectation that the difficulty of attaining that level on the two tests is psychometrically equivalent. It is certainly possible to design studies to investigate the relationship among performance levels on different assessments, such as the analyses that were done in the NAEP state mapping study that linked each state NCLB assessment and NAEP. However, there are often political, educational, or psychometric reasons that such studies are not conducted.

The comparison of the performance standards across NAEP and the SBAC and PARCC assessments presents some interesting challenges. NAEP results are reported at four levels, labeled Below Basic, Basic, Proficient, and Advanced. The SBAC consortium plans to use four achievement levels but will refer to them as either Levels 1–4 or as Novice, Developing, Proficient, and Advanced depending upon the context (SBAC, 2014a). The PARCC consortium has recently decided to report total test scale scores for mathematics, reading, and writing in relation to five performance levels labeled Minimal, Partial, Moderate, Strong, and Distinguished (PARCC, 2015a, 2015b; Nellhaus, 2015). An additional complication is that, unlike NAEP, both SBAC and PARCC will apply these performance standards to a total English/Language Arts (ELA) score. While the consortium assessments will report subscores for reading and writing as well, the performance standards and performance level descriptors will not be applied to these subscores.

It is not only the labels of the performance standards and the scores to which they are applied that differ across these assessments. There will be differences in the percentages of students who will be able to attain these levels of performance. At this time, only SBAC results can be compared to NAEP because PARCC will not conduct standard-setting activities until the summer of 2015. The SBAC results represent estimates of the percentage of students by performance level based on the 2014 trial administration in SBAC member states. Mathematics presents a more straightforward comparison between NAEP and SBAC since both report total mathematics scores by performance levels. These data are shown in Table 2. Note that there are two differences in the timing of when the data used to produce these estimates were collected for the two programs. First, NAEP was administered in 2013 and SBAC in 2014. Second, the high school tests were given in grade 12 for NAEP and grade 11 for SBAC. However, these differences in timing should not

eliminate the usefulness of the comparisons since there are usually not large changes in student performance from one year to the next or for adjacent grades.

Table 2: Comparison of SBAC and NAEP Achievement Level Scores in Mathematics¹

Level	Performance Level Descriptor		Grade 4		Grade 8		High School ²	
	SBAC	NAEP	SBAC	NAEP	SBAC	NAEP	SBAC	NAEP
4	Advanced	Advanced	13	8	13	9	11	3
3	Proficient	Proficient	24	34	19	27	22	23
2	Developing	Basic	36	41	30	38	27	39
1	Novice	Below Basic	27	17	38	26	40	35

¹ SBAC data are from 2014 in member states; NAEP data are from 2013 in all states (SBAC, 2014b; NCES, 2014).

² "High School" refers to grade 11 for SBAC and grade 12 for NAEP.

There are some patterns of interest evident in the data. The SBAC results are more evenly spread across the levels. In grade 4, for example, the percentage of students by level for SBAC varies from a low of 13 percent (level 4) to a high of 36 percent (level 2). The corresponding results for NAEP are 8 percent to 41 percent, also for levels 4 and 2, respectively. This pattern, which is also evident in grade 8 and high school, produces a flatter distribution across the performance levels for the SBAC results. Both programs have the lowest percentage of students in the highest level. However, the greatest percentage of students scored in level 2 for NAEP at all three grades, while SBAC estimates the greatest percentage of students will be in level 1 for grade 8 and high school and level 2 for grade 4.

Interpreting the comparison of SBAC and NAEP is more difficult in the area of ELA because of the differences in the way the two programs report results. However, we can compare the results for ELA to those observed in mathematics. The ELA data are shown in Table 3. It is evident that the SBAC results again create a flatter distribution with a more even spread across performance levels. The patterns of which levels hold the highest and lowest percentages of students are also fairly similar to mathematics, with the majority of students falling in the first two levels for both SBAC and NAEP.

Table 3: Comparison of SBAC and NAEP Achievement Level Scores in English Language Arts/Literacy^{1,2}

Level	Performance Level Descriptor		Grade 4		Grade 8			High School ³		
	SBAC	NAEP	SBAC	NAEP Reading	SBAC	NAEP Reading	NAEP Writing	SBAC	NAEP Reading	NAEP Writing
4	Advanced	Advanced	18	8	9	4	3	11	5	3
3	Proficient	Proficient	23	27	32	32	24	30	32	24
2	Developing	Basic	22	33	31	42	54	31	37	52
1	Novice	Below Basic	37	32	28	22	20	28	25	21

¹ SBAC data are from 2014 in member states; NAEP Reading data are from 2013 and NAEP Writing data are from 2011 in all states (SBAC, 2014b; NCES, 2012, 2014; The Nation's Report Card, 2013).

² SBAC reports "English Language Arts/Literacy." NAEP reports "Reading" and "Writing" separately, and does not currently assess "Writing" in Grade 4.

³ "High School" refers to Grade 11 for SBAC and Grade 12 for NAEP.

All of these differences between NAEP and the consortia will complicate attempts to compare results. The major differences include how performance standards were set, how the standards are applied, and how test results are reported. A cautious psychometric position would be to declare that the results of the separate assessment programs are not comparable. This position, however, is impractical. The largest audiences of these testing results—educators, students, parents, policy makers and the public—will see the similarities between the programs more than they will recognize the differences. These stakeholders will point out that these programs measure the same content areas, overlap in at least some grades, profess to establish students' proficiency levels, and use some of the same metrics and performance standards to report test results. These audiences will expect the results to make sense and, if the results do not, they will expect someone to explain it to them.

Fortunately, NAEP has been dealing with many of these types of issues since long before the CCSS and CCSS-based assessments existed. NAEP has operated effectively for decades despite changing state and federal legislation, shifting curricula, and the implementation of assessments and associated performance standards that vary enormously across states. The primary difference between pre-CCSS NAEP and the current situation is the magnitude of the CCSS focus. The support for the CCSS is too great to be ignored but too fluid to predict with any degree of certainty. Accordingly, it seems prudent to take steps now to ensure that NAEP will be in a position to satisfactorily explain its results over the next several years regardless of what happens to the support for the CCSS nationwide.

It is highly desirable to create a psychometric link between NAEP and the consortium measures. This would support conclusions regarding the relative rigor of the performance standards on these assessments with accuracy that would not be possible absent such a link. One aspect of the ESEA legislation due to be reauthorized by Congress that would facilitate such a link is a requirement that states administer NAEP reading and mathematics in grades 4 and 8 (Every Child Achieves Act, 2015). If this occurs, student performance on the consortium assessments could

be mapped onto NAEP scales. An alternate strategy would be to negotiate with the consortia to design a study employing embedded items on the assessments.

Recommendation 6. A study or series of studies should be designed and implemented to examine how test results are reported in relation to performance levels by NAEP and the consortia. This research should explore the relationship among these reporting practices, and would be greatly enhanced by establishing a psychometric link between these assessments. The creation of such a link may be facilitated by the reauthorization of ESEA and would be enhanced by obtaining the cooperation of the consortia. If such a link is not possible, this research should focus on examining relevant, available data in order to gain a better understanding of the meaning of the results, including the identification of areas of inconsistency.

Test Format and Presentation. Both the CCSS-based assessments provided by the SBAC and PARCC consortia and NAEP are intended to be presented on computers, tablets, or other devices. Yet all of these programs currently are operating in an environment that is transitioning from paper-and-pencil administrations. This transitional period is presenting the field with many challenges involving decisions about hardware, software, and various issues related to test administration, including the degree to which assistance (e.g., calculators, graphing tools, word check) should be provided to examinees and what should be the optimal form of such assistance. Eventually, many of these issues will be resolved and certain protocols likely will become standard operating procedure. However, over the next several years, there are two areas in which this transitional environment increases the threat to NAEP's ability to continue to achieve its dual goals of validly reporting on the status of student achievement for the nation and maintaining trend lines.

The move to digitally based assessment (DBA) requires careful examination of changes to the assessment that may be introduced, whether intentionally or not. Note that these potential changes would be of concern to NAEP irrespective of the emergence of the CCSS. That is, it would have been necessary for NAEP to be cautious during this transition period even if the CCSS never existed. A recent white paper titled *NAEP's Transition to Digitally-Based Assessment* (Oranje, Moran, Campbell, Dresner, Persky, Dion, & Scalise, 2015) provides an excellent discussion of the issues involved in transitioning from paper and pencil to DBA administration and the actions needed to understand how the transition affects the assessments.

The overall design of the transition from paper-and-pencil to DBA incorporates both studying the effect of the mode transition and deciding how and when to add new content. As a result, we will conduct two separate types of studies as part of the transition plan for each subject-area: DBA start-up assessments and DBA pilots. The first type of study (DBA start-up) is designed to transition the existing assessment items from paper to device delivery, and to allow us to both measure and potentially adjust the metric in which results are reported for differences caused by this mode change. Note that while this is basically an administration change, it is more complicated than simply moving items in their current form from paper to computer. Many of the features that make items function well in a paper-based administration are unnecessary or even counterproductive in a digital delivery mode.

The second type of study (DBA pilot) is meant to try-out new items that make effective use of technology as part of the assessments. In all cases, these will measure skills consistent with the existing assessment frameworks. However, they will be targeted to enhance measurement in areas where we have been limited by paper-based administration, or to allow for measurement of elements of the framework that have previously been unavailable to us. The goal of these studies will be to determine the degree to which the addition of new item types to the assessment changes the construct measured by the assessment—and hence introduces perturbations and changes to the meaning of the trend results—and to allow for consideration of various approaches to analysis and reporting should such changes be detected. (p. 5)

This represents a sensible approach that should provide valuable information for NAEP that will be particularly relevant to NAEP's need to maintain trend lines. Recognizing this, the Oranje et al. paper (2015) goes on to describe two possible enhancements to the proposed transition design studies that would increase their utility by expanding pilot activities in 2017. One of these enhancement options would be accomplished by adding a national paper-based assessment sample and the other would utilize a paper-based validation component with samples drawn from each state. The national sample study would allow the mode effect to be tracked nationally for two years, providing information regarding the stability of the link. The state sample study would provide information about the stability of the link nationally and also at the state level. Both of these options would be useful in different ways, but it is highly desirable that at least one of them be implemented.

An additional set of transition-related issues that could affect NAEP arises when the transition plans of the SBAC and PARCC consortia are considered. These issues are related to the many decisions that test developers must make during this transitional period. As NAEP, SBAC, and PARCC work through the various hardware and software options, there are likely to be many ways in which these decisions will take the programs down somewhat different paths. Examples would include the type of device on which the assessment is administered, the format of the assessment platform, the layout of icons for features like text-to-speech, and the types of assistance available to students. Some of the differences that occur will be trivial. Others may be significant. It is not difficult to imagine students in a state with a strong commitment to one of the consortium assessments becoming confused or anxious if they find certain elements of the NAEP administration to be meaningfully different from the DBA format in use by SBAC or PARCC.

This set of transition issues is important to the role of NAEP as the nation's report card. Over the years, NAEP has remained relevant by adjusting its policies, methods, and content whenever it was determined that doing so was necessary to be consistent with major shifts in American public schools. The states' transition from paper-based assessments to DBA is another such shift that should be continually monitored to ensure that NAEP methods and procedures do not differ too greatly from those adopted by states.

Recommendation 7. The enhancements to the transition plan outlined in the Oranje et al. (2015) paper should be given due consideration. Strengthening the design of the proposed studies by adding a paper-based pilot component in 2017 is the best way to investigate transitional effects and obtain information useful for maintaining trend lines.

Recommendation 8. The DBA decisions made by test developers leading the SBAC and PARCC consortia should be regularly monitored over the next several years and compared to the corresponding NAEP decisions. Differences in approaches should be evaluated to determine the extent to which they might affect student performance. It may be useful to consider collecting feedback from teachers and students regarding their perceptions and experiences with NAEP and the consortium assessments.

Participation Rates. NAEP has always needed to attend to variations in participation rates across states. Various approaches have been used to avoid threats to the valid interpretation of NAEP scores, such as flagging scores of high exclusion states and exploring the use of full population estimates (McLaughlin, 2005). Some aspects of this issue are ongoing and are affected little or not at all by the introduction of the CCSS and the SBAC and PARCC assessments. However, there is a possibility that states' responses to the CCSS and the associated assessments will alter the states' participation rates on their own tests. This change, if it occurs, would be of concern to NAEP.

Two factors should be considered as possibly increasing the likelihood of states changing their practices for obtaining student participation. First, the CCSS are widely accepted as more academically demanding than the curricula previously used in many states. Certainly, the degree to which the CCSS raises the bar for students will differ from one state to another. We know from years of NAEP state-by-state results that there is large variance in student achievement across states. Under NCLB, states with lower achievement frequently masked the situation by administering easier assessments, setting lower performance standards, or both. As these states move from their NCLB assessments to the more demanding consortium assessments, they will experience greater changes in difficulty, complexity, and comprehensiveness of the assessments. One possible response will be to re-examine educational policies dealing with student participation in CCSS-based assessments. Typical techniques used by states in the past have included allowance of waivers, administration of alternate assessments, and creation of alternate paths. Parental involvement may also play a role, as there is a growing interest in families opting out of participation in testing for political or educational reasons (Klein, 2015). Some schools have reported that over half of their students have missed assessments due to the opt-out phenomenon (Brody, 2015). If these trends escalate and enough states start to modify their policies or practices, there could be a significant effect on participation rates in states using the SBAC and PARCC assessments. A related concern is that, if the opt-out phenomenon evolves into a widespread backlash against testing, NAEP may begin to experience difficulty in obtaining student participation.

A second factor is related to the transition to DBA discussed in the previous section. Most states are currently in the middle of this transition. Accordingly, it is too early

in the process to know with any certainty what obstacles might be encountered. Will there be many students for whom the DBA format is problematic? Will this number vary much from state to state? Will the problems decrease over time? The answer to these questions will eventually be evident. Until then, it is advisable to pay attention to state trends in participation rates.

It should be noted that the concerns regarding participation rates apply to all students. It is true that students with disabilities are often granted a variety of accommodations to facilitate their participation in test administrations. However, the factors identified here will affect all students to varying degrees.

Recommendation 9. It is recommended that a state-by-state study be conducted that tracks the trend in participation rates and policies. It is suggested that the study look backward several years to create a baseline of practices and participation rates that existed prior to the implementation of the CCSS and the associated assessments. The study should continue to compile results annually until there is evidence that state participation policies and rates have stabilized.

Section 5: Discussion of Policy Considerations

The major policy issue facing NAEP concerns the ways in which the program can and should change in the coming years. This includes deciding the nature of changes to be made, the manner in which such changes should be implemented, and the pacing of such implementation. If NAEP changes too much or too rapidly, the program may sacrifice some of its capacity to effectively track student achievement over time. If NAEP changes too little or too slowly, the program may lose relevance. Achieving the optimal balance in the coming years will require the consideration of how educational assessment is changing in the country.

The reaction of states to the CCSS has evolved over the past five years. In 2010, when 45 states and the District of Columbia adopted the Common Core as their target content standards, the CCSS appeared destined to become a monolithic, all-encompassing force in America's public schools. It still might. But as the SBAC and PARCC consortia began to build the assessments that would measure students' achievement on the CCSS, states got a closer look at the implications of their embrace of the CCSS. They gained a better understanding of the nature and structure of the CCSS. As the assessments moved through the development and pilot testing phases, states were able to examine how the tests compared with what they previously were using. They started to consider the costs associated with adoption, both to purchase the assessments and to support the professional development efforts needed to familiarize their professional staff with the CCSS.

This rethinking of states' positions relative to the CCSS and the CCSS-based assessments has led some to modify their commitments. The most common response for these states has been to remain committed to the CCSS but to seek alternate assessment strategies, with 15 states choosing this course of action. Additionally, three states have joined the original four that chose to not adopt the CCSS. But with more than half of the states using the SBAC or PARCC assessments this year, and the vast majority of states remaining committed to the CCSS, the Common Core continues to be hugely influential in American education. The signs of weakening support are noticeable and significant, but they have not progressed to the point of undermining the potential of the CCSS to shape the instruction that will be delivered to millions of public school students in the coming years.

The challenge to NAEP is that it cannot wait to see how states' reactions to the CCSS evolve. At the 10-year anniversary of the CCSS in 2020, we will be able to look back and conclude in what ways and to what degree the implementation of the CCSS was successful. We will know how influential the CCSS have been. And we will know how NAEP's role in measuring the achievement of this nation's students has been affected. Unfortunately, waiting for this future knowledge to arrive is not a reasonable option.

There are two strategies that could help solve the dilemma of needing to act while being faced with many uncertainties. First, the focus should be placed on identifying those steps that are necessary regardless of how state reactions to the CCSS may change. A good example of this is the need to maintain achievement trends throughout the transition to DBA. The suggested studies will be just as important

whether state support for the CCSS is maintained in the coming years or continues to erode. Another activity that meets this criterion is the examination of the items and tasks in the SBAC and PARCC assessments. The comparison of these assessments with NAEP is justified as long as the consortia continue to function. These activities should receive the highest priorities.

A second useful strategy is to adopt a five year plan that guides how NAEP will deal with the CCSS transitional period. From a state's perspective, embracing the CCSS is not a one-time activity. Most of the states that support the CCSS are in the second or third year of coming to terms with the new standards. A typical progression in these states would have been a formal adoption followed by the introduction of the CCSS to the professional staff. This would have been followed by a year of getting familiar with the chosen CCSS-based assessments, which is the phase most states are currently in. The next several years will involve activities such as the provision of additional professional development to help teachers implement the CCSS, analysis of data from the CCSS-based assessments to evaluate student achievement, and implementation of additional features provided by the consortia, such as the formative assessments.

This progression means that the CCSS-based environment that NAEP is facing will be evolving over the next several years. This is different than some of the changes that NAEP has had to address, such as when a NAEP framework is revised. In this environment, it is likely that some states will move closer to the CCSS year by year while other states might move away. NAEP can establish a strategy of identifying changes in state practices that have occurred and those that are likely to occur over the next several years. This information can be used to plan how NAEP should best respond over the next five years in order to ensure that the traditional roles of the program are being fulfilled.

The final observation concerns NAEP's allocation of resources over the coming years. NAEP budgets are not likely to allow implementation of all of the studies and activities discussed in this paper. Accordingly, it will be necessary to choose from among the possibilities and implement those that address the greatest areas of concern. The final, two-part recommendation to prioritize the issues and to establish a plan of action is an attempt to provide the National Center for Education Statistics with a strategy for remaining focused on the most important concerns while expending available resources as efficiently as possible.

Recommendation 10. NAEP should implement a two-part strategy to deal with the uncertainty of a CCSS-based environment that will be changing and shifting over the coming years. First, identify and prioritize those activities that are important to NAEP's success regardless of how state support of the CCSS may change. Second, develop a five-year plan of action that takes into account the fact that states' responses to the CCSS will be evolving over the next several years.

References

- Academic Benchmarks. (2015). *Common Core State Standards Adoption Map*. Retrieved July 8, 2015, from <http://academicbenchmarks.com/common-core-state-adoption-map/>
- Associated Press. (2014). *Oklahoma Repeals Common Core Education Standards*. Retrieved July 8, 2015, from <http://www.foxnews.com/us/2014/06/05/oklahoma-repeals-common-core-education-standards/>
- Brody, L. (2015). *N.Y. Testing Is Put to the Test*. *The Wall Street Journal*, V 265, No. 87, April 15, pp. A15, A18.
- Common Core State Standards Initiative. (2015). *Standards in Your State*. Retrieved July 8, 2015, from <http://www.corestandards.org/standards-in-your-state/>
- Every Child Achieves Act. (2015). S. 1177, 114th Congress. Retrieved July 8, 2015, from http://www.help.senate.gov/imo/media/S_EveryChildAchievesActof2015.pdf
- Federal Register. (2010). Part III: Department of Education. *Overview Information; Race to the Top Fund; Notice Inviting Applications for New Awards for Fiscal Year (FY) 2010; Notice*. V 75, No. 71, April 14, 2010.
- Fineout, G., and Talley, T. (2014). *Indiana becomes first state to drop Common Core*. Retrieved July 8, 2015, from <http://wane.com/2014/03/24/pence-approves-withdrawal-from-common-core/>
- Gewertz, C., and Ujifusa, A. (2014). *State Plans for Testing Fragmented*. *Education Week*, V 33, No. 32, May 21, pp.1,16.
- Gewertz, C. (2015). *Final Testing Count: More Than Half of Students Will Take Non-consortium Exams*. Retrieved July 8, 2015, from http://blogs.edweek.org/edweek/curriculum/2015/02/more_than_half_of_students_.html?cmp=ENL-EU-NEWS2
- Hughes, G. B., Daro, P., Holtzman, D., and Middleton, K. (2013). *A Study of the Alignment Between the NAEP Mathematics Framework and the Common Core State Standards for Mathematics (CCSS-M)*. A publication of the NAEP Validity Studies Panel. San Mateo, CA: American Institutes for Research.
- Klein, A. (2015). *How Will the Education Department Handle Schools With High Opt-Out Rates?* *Education Week*. Retrieved July 8, 2015, from http://blogs.edweek.org/edweek/curriculum/2015/03/opting_out_of_testing_a_befuddling_mix_of_state_rules.html
- Massachusetts Department of Elementary & Secondary Education. (2015). *Partnership for Assessment of Readiness for College and Careers: PARCC Overview*. Retrieved July 8, 2015, from <http://www.doe.mass.edu/parcc/>

- McLaughlin, D. H. (2005). *Properties of NAEP Full Population Estimates*. San Mateo, CA: American Institutes for Research.
- National Center for Education Statistics. (2012). *Writing 2011: National Assessment of Education Progress at Grades 8 and 12*. Institute of Education Sciences: U.S. Department of Education. NCES 2012-470.
- National Center for Education Statistics. (2014). *A First Look: 2013 Mathematics and Reading: National Assessment of Educational Progress at Grades 4 and 8*. Institute of Education Sciences: U.S. Department of Education. NCES 2014-451.
- National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards*. Washington, DC: Author. Retrieved on July 8, 2015, from <http://www.corestandards.org/>
- Nellhaus, J. (2015). Personal communication in a telephone conversation and email to the author. March 13, 2015.
- Oranje, A., Moran, R., Campbell, J., Dresher, A., Persky, H., Dion, G., and Scalise, K. (2015). *NAEP's Transition to Digitally Based Assessment*. A paper shared by email with the author. February 27, 2015.
- Partnership for Assessment of Readiness for College and Careers (PARCC). (2015a). *PARCC Assessment - Assessment System - Contents of the Grade- and Subject-Specific Performance Level Descriptors: Mathematics*. Retrieved January 2, 2015, from <http://www.parcconline.org/math-plds>
- Partnership for Assessment of Readiness for College and Careers (PARCC). (2015b). *PARCC Assessment - Assessment System - Contents of the Grade- and Subject-Specific Performance Level Descriptors: ELA/Literacy*. Retrieved July 8, 2015, from <http://www.parcconline.org/ela-plds>
- Smarter Balanced Assessment Consortium (SBAC). (2014a). *Smarter Balanced - Assessment System - Achievement Level Descriptors*. Retrieved July 8, 2015, from <http://www.smarterbalanced.org/wordpress/wp-content/uploads/2014/11/Interpretation-and-Use-of-Scores.pdf>
- Smarter Balanced Assessment Consortium. (2014b). *Smarter Balanced States Approve Achievement Level Recommendations*. Retrieved July 8, 2015, from <http://www.smarterbalanced.org/news/smarter-balanced-states-approve-achievement-level-recommendations/>
- The Nation's Report Card. (2013). *2013 Mathematics and Reading: Grade 12 Assessments*. Retrieved July 8, 2015, from http://www.nationsreportcard.gov/reading_math_g12_2013/#/what-knowledge
- Ujifusa, A. (2014). *S.C. Governor Signs Bill Requiring State to Replace Common Core*. Retrieved July 8, 2015, from

http://blogs.edweek.org/edweek/state_edwatch/2014/06/south_carolina_gov_haley_signs_bill_to.html?cmp=SOC-SHR-TW

Wixson, K. K., Valencia, S. W., Murphy, S., and Phillips, G.W. (2013). *A Study of NAEP Reading and Writing Frameworks and Assessments in Relation to the Common Core State Standards in English Language Arts*. A publication of the NAEP Validity Studies Panel. San Mateo, CA: American Institutes for Research.