White Paper to Provide Context for NAEP Achievement Levels by Reviewing State and International Practices

Peter Behuniak
Criterion Consulting, LLC

Denny Way
College Board

February 2022
Commissioned by the NAEP Validity Studies (NVS) Panel

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:

Project Director:

Sami Kitmitto
American Institutes for Research

Project Officer:

Grady Wilburn
National Center for Education Statistics

For Information:

NAEP Validity Studies (NVS) Panel
American Institutes for Research
2800 Campus Drive, Suite 200
San Mateo, CA 94403
Email: skitmitto@air.org

ACKNOWLEDGMENTS

Many individuals made significant contributions to the development of this paper. Thanks are due to Jason Nicholas at Westat and Gina Broxerman at NCES for their roles in helping to create and distribute the state survey. The effort of all of the NAEP State Coordinators to compile responses to the surveys was greatly appreciated. A special thanks goes to two State Coordinators, Renee Savoie in Connecticut and William Donkersgoed in Wyoming, for providing valuable suggestions in the development of the survey. The guidance and assistance by Jack Buckley, Fran Stancavage, Sami Kitmitto, Mikael Rae, and their colleagues at AIR were instrumental throughout the course of this project. Finally, the suggestions and comments by the NVS panelists helped greatly to improve the paper.

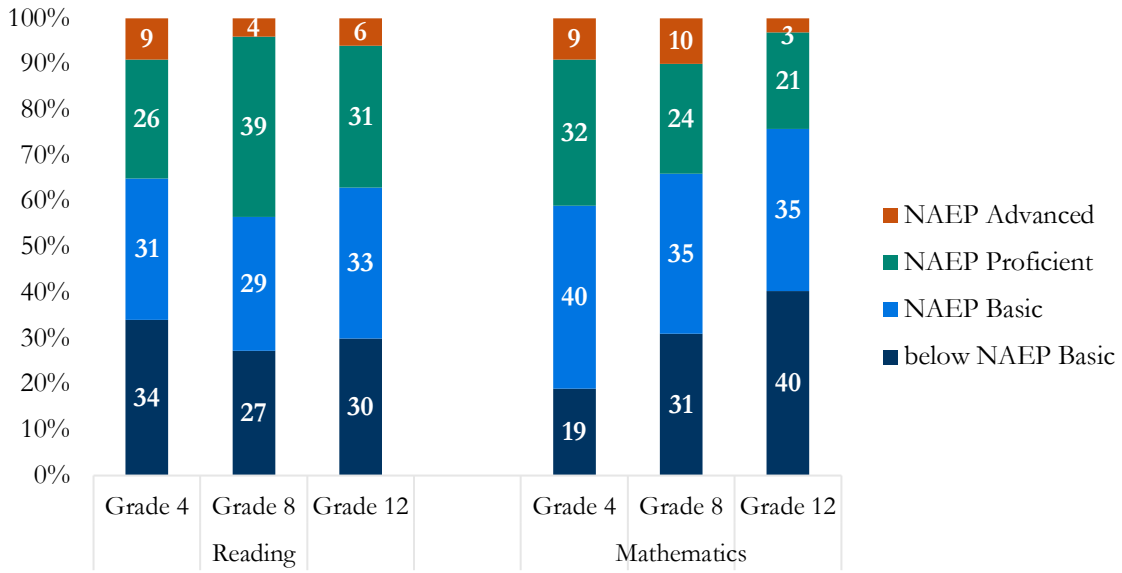
CONTENTS

INTRODUCTION

The National Assessment Governing Board (NAGB) has recently turned increased attention towards the NAEP achievement levels (ALs) and associated achievement level descriptors (ALDs). There has already been a considerable amount of discussion and deliberation. The process began with the work by the National Academies of Sciences, Engineering, and Medicine (NAS) that led to the publication of the *Evaluation of the Achievement Levels for Mathematics and Reading on the National Assessment of Educational Progress* (National Assessment Governing Board, 2017). The conclusions and recommendations in this report were reviewed by NAGB; their responses and proposed next steps are summarized in the *NAGB Achievement Levels Work Plan* (National Assessment Governing Board, 2020b) and *Update to Achievement Levels Work Plan* (National Assessment Governing Board, 2020c).

NAEP ALs and ALDs have been in use for many years. NAEP uses three ALs, and the associated ALDs are designated as *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced* (see Appendix A for ALDs for Reading Grade 4 and links to ALDs for other grades and subjects). The primary purpose for the NAEP ALs and ALDs is to aid educators, parents, legislators, and other stakeholders in their interpretation of NAEP results (see 2019 results for reading and mathematics in Figure 1). One of the goals of the review by NAGB is to enhance the effectiveness of the ALs and ALDs in future NAEP administrations.

Figure 1. NAEP Achievement Level Results 2019 in Reading and Mathematics



Note. Detail may not sum to totals because of rounding.
Source. NAEP Report Card: Reading (https://www.nationsreportcard.gov/reading/nation/achievement/?grade=12) and *NAEP Report Card: Mathematics* (https://www.nationsreportcard.gov/mathematics/nation/achievement/?grade=12).

This white paper offers two specific contributions for NAGB to consider as they move forward with plans to respond to the NAS report and to develop evidence to support removal of the trial status of the NAEP ALs. NAEP operates in an educational environment that includes the use and dissemination of information from many different educational assessments. Each state annually administers its own battery of assessments. Thus, the

release of NAEP results to state educators and other stakeholders occurs in a context in which the results of other assessments must also be considered. It may be useful to NAGB in pursuing the achievement levels work plan to have some insight into state assessment practices and how NAEP results fit into those practices.

A second opportunity for insight may come from considering international assessments that are administered on cyclical bases similar to NAEP and that have adopted approaches for establishing and maintaining ALs and ALDs that differ from NAEP. In particular, the international assessments do not provide any policy definitions for their achievement levels, but rather establish ALs by defining benchmarks along their reporting scale, and through item mapping and scale anchoring processes that are descriptive in nature. As NAEP endeavors to facilitate clear, accurate, and informative reporting of NAEP achievement level results to the public, these international assessment practices may have increased relevance.

The paper is structured in two main parts. The first section is focused on activities that have taken place in the states with regard to each state's own assessments as well as their perceptions and uses of NAEP results. The information presented in this section is based on a survey of state practices that was distributed to all states and jurisdictions involved in administering NAEP. In the second section, we examine the practices used by international assessments to establish achievement levels and achievement level descriptors. We contrast the practices of different international assessments, note that they were largely patterned after approaches used by NAEP prior to 1990, and present a case for their relevance to NAEP ALs and ALDs as the work plan in response to the NAS report proceeds.

State Assessment Practices

Background

This section focuses on the assessment practices used by the 50 states, Washington, DC, and Puerto Rico, and how these jurisdictions perceive and employ NAEP results in concert with their own assessment systems. These jurisdictions have a long history of conducting educational assessments in multiple content areas to monitor the achievement of their students and to address other concerns—such as to satisfy accountability requirements and judge the effectiveness of established curriculum. Many years ago the nature of these local assessment systems was largely determined by state educational leaders or state legislatures. Over the past several decades, federal requirements have exerted an increasing influence on the nature of these systems.

The introduction in 2010 of two nationwide assessment consortia, Smarter Balanced and PARCC, created the most recent shift in the landscape of educational achievement testing across the country. Many jurisdictions joined one of these consortia and began the process of collaboration as encouraged by the U.S. Department of Education. While the influence of the consortia has waned in the last few years, as some jurisdictions have adjusted or eliminated their affiliations, the effects of the consortium-developed procedures and assessments are still evident in many places.

It is in this context that NAEP operates. Every release of NAEP results to educators and other stakeholders in these jurisdictions is considered in conjunction with the results of local assessments. There is great variation across jurisdictions, not only in the features of their assessment systems, but also in regard to other educational and political issues and practices. It is our hope that, by examining how these jurisdictions perceive and reconcile the NAEP program and their own local assessment systems, we can shed some light on the environment in which NAEP operates.

Methodology

Information regarding state assessment practices was collected by means of a survey developed by the authors. The data collection plan was to distribute the survey electronically to NAEP State Coordinators in each of the 50 states, Washington, DC, and Puerto Rico. A draft of the survey questions was reviewed by the Westat coordinator for NAEP and two NAEP State Coordinators, and revised based on their feedback. The final survey (Appendix B) was shared with NAEP State Coordinators during a teleconference on January 22, 2021. The Coordinators were assigned the primary responsibility for responding to the survey questions, with instructions to request assistance as necessary from other knowledgeable individuals in their agencies, such as state assessment directors.

The survey was distributed during the first week of February using the SurveyMonkey platform. The distribution instructions requested a response by the end of February. Various state agency activities during that period caused a delay of several weeks in some jurisdictions, resulting in the last responses being submitted in late March. The final response rate was 98.1 percent, with 51 of the 52 jurisdictions successfully completing the survey. The one state that did not complete the survey was affected by the fact that it was operating without a NAEP State Coordinator.

Survey Results

The first question asked each jurisdiction to confirm its current AL and ALD. It should be noted that the use of the terms AL and ALD for the purposes of the survey carry specific meaning. The use of AL refers to the number of standards applied to each assessment while ALD refers to the labels attached to each AL. Many jurisdictions provide further elaborations of each ALD, which range from a phrase to a paragraph or more. While it is recognized that some jurisdictions consider the elaborated descriptions to be part of their ALD, the purpose of this study was necessarily limited to the labels associated with each AL.

There were many different variations of ALD selected, but three clusters were identifiable. The first and most common was the method selected by 15 jurisdictions: using the labels "Advanced" and "Proficient" to describe their highest two AL (Table 1). This is the set of ALDs that most resembles the ALDs used by NAEP. However, the ALD for the lower AL varied considerably across jurisdictions. (A more detailed version of Table 1 is provided in Appendix C.)

Table 1. Achievement Level Descriptor Themes

Scheme Type	N
Advanced/Proficient	15
Exceeds/Meets	11
Numbered Levels	11
Other	15
Total	52

Note. For the one state that did not respond to the survey, we collected information about the ALD scheme from its website.

The other two clusters of jurisdictions elected to use Exceeds/Meets or numbered levels to describe their highest two ALs. Both of these approaches were adopted by 11 jurisdictions. As with the first cluster, the ALD for the lower AL varied considerably across jurisdictions. The remaining 15 jurisdictions used a variety of ALDs to report scores at all ALs. Overall, it is clear that there are many different variations of ALDs employed across the country.

Question 2 focused on the degree of influence that NAEP ALs and ALDs had on the standard-setting activities used by each jurisdiction when they established the ALs and ALDs for their own assessments. The results are presented in Table 2; they indicate that the NAEP ALs had greater influence on local standard-setting efforts than did the NAEP ALDs. Over half of the jurisdictions, almost 65 percent, reported that NAEP ALs influenced the setting of their own ALs "a little" or to a "considerable" degree. Far fewer jurisdictions, under 30 percent, indicated that the NAEP ALDs were influential for setting their state's ALDs.

Table 2A. Was the Setting of Your State's Current Achievement Levels (Cut Scores) Influenced at All by the NAEP Achievement Levels and/or ALDs? [Question 2]

Response Options	Number of Responses	Percentage of Responses
No	18	35.3%
Yes, a little	27	52.9%
Yes, to a considerable degree	6	11.8%
Total	**51**	**100.0%**

Table 2B. Was the Setting of Your State's Current ALDs Influenced at All by the NAEP Achievement Levels and/or ALDs? [Question 2]

Response Options	Number of Responses	Percentage of Responses
No	34	70.8%
Yes, a little	13	27.1%
Yes, to a considerable degree	1	2.1%
Total	**48**	**100.0%**

Note. Three states/jurisdictions did not respond to this question.

In the comments submitted by some of the respondents, there were two themes worth noting. The first is that several of the states indicated that they were participating with one of the assessment consortia, Smarter Balanced or PARCC, at the time they were establishing their own ALs and ALDs. Since these consortia included consideration of NAEP ALs and ALDs in the standard-setting processes, participating states noted their compliance with these procedures. The second point made clear in the comments is that the NAEP ALs and ALDs were usually one source of several considered in order to establish a context for the setting of the local standards. Other sources mentioned included the TIMSS, PISA, SAT, and ACT.

The next set of survey questions, Q3 through Q5, encouraged respondents to select all choices that applied to each question. Several jurisdictions did indicate that multiple selections applied to them. The body of each table indicates the frequency with which each response choice was selected. Because some jurisdictions selected more than one response, the total number of responses exceeds the number of jurisdictions. The notes below each table report how many jurisdictions made multiple selections.

Question 3 asked about the stability of the local ALs by inquiring whether they had changed over the past 5 years and, if so, why. The results (Table 3) indicate that there were no changes in a slight majority (52.9 percent) of the jurisdictions. When changes were reported, the most common reason given (41.2 percent) was that a new assessment design had been introduced in one or more parts of the local assessments. Existing policy, such as having a mandate to revisit standards on a fixed schedule, accounted for changes in 11.8 percent of the jurisdictions. Other explanations for making changes were not cited very frequently.

Table 3. Have Your State's Achievement Levels Changed in the Past 5 Years? If so, Please Indicate the Rationale(s) for Making the Changes. (Select All That Apply.) [Question 3]

Response Options	Number of Selections	Percentage of Respondents
A) There have been no changes.	27	52.9%
B) Changes were initiated based on existing policy (e.g., revisiting standards on a fixed schedule).	6	11.8%
C) Changes were initiated due to adoption of a new assessment design.	21	41.2%
D) Changes were initiated due to a change in policy as determined by an individual or group (e.g., Chief State School Officer, State Board of Education, legislature).	2	3.9%
E) Changes were initiated for practical or educational reasons, such as the standards were judged to be too high or too low.	1	2.0%

***Note.* 51 jurisdictions provided one or more responses to this item. Of these 51, five made multiple selections: one marked A and C; two marked B and C; one marked C and D; one marked C, D, and E.**

The jurisdictions submitting comments for this question were predominantly attempting to clarify their responses to any changes that were made. The clarifications were generally of two types. The first was to explain that a new assessment design was introduced for part but not all of the local assessments, such as implementing a new design for the science assessments while making no changes to mathematics or English/language arts. The second type of clarification was to explain the nature of the local development that necessitated the change, such as a change in the length of the test.

Question 4 asked about changes to the local ALD that had been made over the past 5 years. As was the case for the previous question, the results (Table 4) indicate that a slight majority of jurisdictions (52.9 percent) reported no changes. About one quarter (25.5 percent) of respondents indicated that the changes were necessitated because of changes to the AL. Only 9.8 percent of the responses cited NAEP as influencing the change, while 17.6 percent specifically ruled out NAEP having any influence on the local changes.

Table 4. Achievement-Level Descriptors are Sometimes Changed Over time in Either Minor or Substantial Ways. Which of the Following Statements Best Describes the Status of the ALDs in Your State Over the Past 5 Years? (Select All That Apply.) [Question 4]

Response Options	Number of Responses	Percentage of Respondents
A) No changes were made.	27	52.9%
B) Changes were made due to changes to the achievement levels.	13	25.5%
C) Changes were partly influenced by NAEP.	5	9.8%
D) Changes were not influenced by NAEP.	9	17.6%

Note. **51 jurisdictions provided one or more responses to this item. Of these 51, two marked both B and D.**

The responses to this question may indicate that there was some misunderstanding on the part of respondents about how to answer. First, 11 states responded that there were changes without selecting changes due to altered AL (option B), suggesting that almost a quarter of the jurisdictions made changes to their ALD for other unspecified reasons. The eight comments submitted in response to this question did not address this issue. Second, there were fewer responses regarding NAEP influence than anticipated: of the 24 respondents who did not select option A, only 13 provided information on the influence of NAEP. There were 9.8 percent of jurisdictions reporting NAEP was an influencing factor and 17.6 percent reporting NAEP had no influence. While it can be fairly inferred that the respondents reporting no change would not address NAEP influence, this still leaves unanswered the status of changes in the other jurisdictions.

Question 5 requested that each jurisdiction identify the empirical methods they used, if any, in setting their ALs and ALDs. The results (Table 5) indicate that over 80 percent used at least one empirical method. The most frequently used procedure (58.8 percent) was the examination of the relationship of their own assessments with other existing measures, such as the NAEP, SAT, or ACT. Also popular was exploring the relationship of their assessments to criterion variables and scale anchoring. Four jurisdictions offered comments regarding one or more local decision(s) that guided this aspect of their standard-setting process.

Table 5. Were Any Empirical Methods, Other Than Sharing Impact Data, Used in the Process of Setting the Current Achievement Levels and/or ALDs? (Select All That Apply.) [Question 5]

Response Options	Number of Responses	Percentage of Respondents
A) No.	10	19.6%
B) Relationships with criterion variables (e.g., benchmarking).	18	35.3%
C) Scale anchoring.	11	21.6%
D) Relationships with existing assessments (e.g., NAEP, ACT, SAT) such as using an equipercentile procedure to link to benchmarks on a comparison assessment.	30	58.8%
E) Other.	4	7.8%

Note. **51 jurisdictions provided one or more responses to this item. Of these 51, 17 made multiple selections: three marked B and C; six marked B and D; two marked C and D; one marked D and E; four marked B, C, and D; one marked B, D, and E.**

Question 6 asked whether additional validation evidence was collected after the AL and ALD were set. The results (Table 6) indicated that two-thirds of the jurisdictions collected at least one type of evidence, with the most common being the examination of student performance on tests other than the state assessment. Additionally, collecting validity evidence from educators and using alternate standard-setting methods were cited frequently, either as the sole validation technique or in combination with another approach (as cited in the comments).

Table 6. After the Achievement Levels and/or ALDs Were Established, Was Any Additional Validity Evidence Collected About Them? [Question 6]

Response Options	Number of Responses	Percentage of Responses
No.	17	33.3%
Yes, one or more alternate standard setting procedures were employed to validate the achievement levels and/or ALDs.	7	13.7%
Yes, performance of students on tests other than the state tests was examined.	13	25.5%
Yes, validity evidence was collected from educators in the state.	7	13.7%
Other (please specify below).	7	13.7%
Total	**51**	**100.0%**

Respondents used the comment section to explain their choices of validation procedures. A number of jurisdictions pointed out that they followed the processes implemented by the assessment consortium to which they belonged. A few jurisdictions cited the SAT and NAEP as the external assessments used for validation purposes.

Question 7 inquired about the amount of influence the state's ALs and/or ALDs have on the release of assessment results to stakeholders. Not surprisingly, the results (Table 7) indicate that the ALs and/or ALDs exert at least some influence in all jurisdictions. Two thirds of the jurisdictions report that the influence of the ALs and ALDs is considerable.

Table 7. How Much Influence Do Your State's Current Achievement Levels and/or ALDs Have When Your State's Assessment Results Are Released to Stakeholders? [Question 7]

Response Options	Number of Responses	Percentage of Responses
None.	0	0.0%
Some. The release of results occurs in a variety of ways, some of which include the percentage of students meeting each achievement level.	16	33.3%
Considerable. The major focus of each release is on the performance of students in relation to the achievement levels, but some of the discussion does not relate to the achievement levels or ALDs.	32	66.7%
Exclusive. The entire focus of each release is on the achievement levels and ALDs.	0	0.0%
Total	**48**	**100.0%**

Note. Three states/jurisdictions did not respond to this question.

Question 8 asked about whether the ALs and/or ALDs were judged to be effective for communicating the results statewide. The results (Table 8) indicate that almost two-thirds of the jurisdictions do feel they are effective and about one-third believe they are somewhat effective. The two jurisdictions responding negatively explained that their response was due to the fact that the process of setting the ALs or ALDs was not yet completed.

Table 8. Have the Current Achievement Levels and/or ALDs Been Judged Effective by Your State Department for Communicating Statewide Assessment Results? [Question 8]

Response Options	Number of Responses	Percentage of Responses
No.	2	4.0%
Somewhat.	17	34.0%
Yes.	31	62.0%
Total	**50**	**100.0%**

Note. One state/jurisdiction did not respond to this question.

Question 9 asked about how jurisdictions treat the release of assessment data in years when results are available for both their own assessments and NAEP. The results (Table 9) reveal that almost three-quarters of the jurisdictions coordinate the release of assessment information. Differences in timing appeared to be the factor that limited the degree of coordination possible.

Table 9. In Years When NAEP Releases State Results, How Are They and the State Assessment Results Treated When Discussed With Stakeholders? [Question 9]

Response Options	Number of Responses	Percentage of Responses
There is a great deal of consideration given to the results of both assessments, as they are viewed as providing two valuable perspectives on student achievement.	9	17.6%
There is no attempt to compare the results from the two assessments as they are viewed as entirely different entities.	14	27.5%
There is some consideration given to the results of both assessments, but differences between them (e.g., timing, content) present significant limitations to interpretation.	28	54.9%
Total	**51**	**100.0%**

Two strategies for using NAEP data were identified. The first was the use of the NAEP state-by-state results to examine the performance of a given jurisdiction relative to other states and the nation. The second was to compare trends on both NAEP and the local assessment. Several jurisdictions pointed out that the way in which these types of analyses are used varies for different local stakeholder groups.

The final question, Question 10, was open ended, and asked whether there were any reporting strategies used locally that should be considered by NAEP. Most jurisdictions (60.8 percent) either skipped answering or answered negatively. The comments from the jurisdictions that did respond fell into two categories: 1) those that described a local practice that was successful, and 2) suggestions for specific additional NAEP reporting options.

Several jurisdictions indicated that their use of reporting vehicles such as dashboards, report cards, or electronic portals was helpful. Local accountability systems were cited as well. One jurisdiction indicated that they used a longitudinal tracking approach to follow high school and college completion rates. Another interesting strategy involved holding workshops and webinars aimed at increasing stakeholders' ability to interpret score reports.

Suggestions to expand or modify NAEP reporting were idiosyncratic. One idea was to expand NAEP subgroup reporting, including following trends. Another request was for NAEP to provide access to a website that would allow a comparison among states judged to have similar demographics on such variables as socioeconomic status and population diversity. Finally, some jurisdictions simply commented that they were satisfied with the current NAEP reporting procedures.

Discussion

It was pointed out in the introduction to this paper that NAEP operates in an educational environment that includes the use and dissemination of information from many different educational assessments. The results of the survey revealed some of the ways in which NAEP interacts with these other assessments across different jurisdictions. Many aspects of local assessment systems are influenced by NAEP practices, starting with the establishment of local ALs and ALDs. There was certainly a great variety of ALs and ALDs implemented by the states and other jurisdictions. However, even jurisdictions that ended up using ALs and/or ALDs different from those in use by NAEP often reported having reviewed the NAEP approach in the process of making their own decisions. It was evident from the survey results that a majority of jurisdictions were influenced by NAEP while conducting their procedures for setting their local ALs.

NAEP's influence also extended to areas other than the establishment of local ALs and ALDs. A number of jurisdictions identified NAEP as one of the assessments considered in the collection of validation evidence. Some cited NAEP as a factor in deciding to make changes to their own ALs or ALDs. One strong area of NAEP influence is in regard to each jurisdiction's process for interpreting and releasing the results of their own assessments. Many jurisdictions reported using NAEP data, such as state-by-state comparisons, to help interpret the results from their own assessments and to supplement their releases of assessment results to specific stakeholders.

The main purpose for conducting this survey of assessment practices across states and other jurisdictions was to provide some contextual evidence that might be useful as NAGB moves forward in its process of studying NAEP ALs and ALDs. It is appropriate that NAGB is devoting the time and effort necessary to ensure that this process is conducted in a thorough and thoughtful manner, particularly given the influence that NAEP procedures exert on local assessment decisions.

One of the decisions made by NAGB concerns the interpretation of NAEP reporting ALDs. This was stated most recently in the *Update to Achievement Levels Work Plan* (National Assessment Governing Board, 2020c):

> Reporting ALDs, as described in the Board's revised policy statement, will be created following administration of an assessment to communicate about what performance at each NAEP achievement level indicates about what students do know and can do. (p. 3)

The decision to adopt a descriptive approach to reporting what students know and can do appears to be relevant and well advised. While not addressed specifically in the survey, this interpretation is in widespread use in many jurisdictions. By adopting this approach, NAGB likely enhances the positive influence that future administrations of NAEP will have on the assessment practices in use throughout the country.

INTERNATIONAL

Background

This section focuses on the practices used by two international assessments to establish achievement levels (ALs) and achievement level descriptors (ALDs), and why these practices may now be more relevant to NAEP than in the past. The Organisation for Economic Co-operation and Development (OECD) began administering the *Programme for International Student Assessment* (PISA) in 2000. PISA is a collaborative effort among OECD Member countries to assess how well 15-year-olds approaching the end of compulsory schooling are prepared to meet "the challenges of today's knowledge societies." The major domain of the PISA survey rotates between reading, mathematics, and science in each 3-year cycle. PISA also measures general or cross-curricular competencies that vary across assessment cycles. The *Trends in International Mathematics and Science Study* (TIMSS) is an international assessment of student achievement in mathematics and science at the fourth and eighth grades. It began as a series of studies conducted by the International Association for the Evaluation of Educational Achievement (IEA), and evolved into the current assessment, which is administered every 4 years. Note that there is a third prominent international assessment, The *Progress in International Reading Literacy Study* (PIRLS), administered every 5 years; it focuses on reading literacy achievement, as well as home and school contexts for learning to read. Because PIRLS is also sponsored by IEA and utilizes the same approach to ALs and ALDs as TIMSS, our comparisons to NAEP will be made with reference to TIMSS but apply equally to both assessments.

These international assessments were established after NAEP and clearly borrow primary aspects of the modern NAEP program, including complex matrix sampling and reporting of proficiency at the group rather than individual level. It is also worth noting that the techniques utilized by PISA and TIMSS for interpreting and communicating about ALs involve approaches that have all been considered by NAEP at some point. In fact, one of the most comprehensive treatments on NAEP interpretation was published shortly after NCES began reporting NAEP results by ALs (Phillips et al., 1993). This report identified seven methods that NAEP either used or "could use" to interpret its scales. These included 1) percentage correct for each item, 2) average percentage correct for groups of items, 3) item mapping, 4) scale anchoring, 5) achievement levels, 6) use of scoring rubrics, and 7) benchmarking. Phillips et al. (1993) also made the important distinction between *scale anchoring* and *achievement levels*. This distinction forms the basis for the major difference between NAEP and these international assessments with respect to ALs and ALDs.

In her paper addressing the history of NAEP achievement levels, Bourque (2009) describes the formation of the National Assessment Governing Board and how, in response to directives to set appropriate achievement goals for NAEP performance, the iconic ALs of "Basic," "Proficient," and "Advanced" were established. A central tenet of the ALs was that they were to describe the content students *should* know and be able to do if they reached a given level. Thus, beginning in 1990, NAEP described levels of student performance in terms of ALs rather than anchor levels. Phillips et al. (1993) distinguished between the two terms as follows:

> In contrast to anchor levels which describe actual student performance on NAEP, achievement levels are performance standards on the NAEP assessment that identify what students should know and be able to do at various points along the proficiency scale. In developing threshold values (cut scores) for the levels, a broadly constituted panel of judges rated each grade-specific NAEP item pool using operationalized policy definitions developed by the Board for "Basic," "Proficient," and "Advanced" student performance. In contrast, the numerical values for anchor levels represent selected points at regular intervals on the scale that have a statistical meaning in describing the distribution of scores. (p. 35)

International ALs and ALDs

For the international assessments, a centralized entity in charge of policy considerations is organizationally infeasible, and for this reason, neither PISA nor TIMSS has established policy definitions regarding what students should be able to do on their assessments. Rather, both adopted anchoring approaches for establishing ALs and ALDs that were similar to those initially used for NAEP. Although there are important distinctions in the approaches for the two assessments, they both broadly involved the following steps (Olsen & Nilsen, 2017):

1. Definition of frameworks for the construct to be measured, including a generic articulation of the range of performance
2. Item development guided by the frameworks
3. Development of *item descriptors*, i.e., short statements describing the knowledge and skills needed to solve each item
4. Assessment data analysis and establishment of a score scale
5. Decision about number and locations of ALs is arbitrarily made
6. Items from the assessment are identified to represent each of the levels established in step 5
7. ALDs are established based on the item descriptors for the items mapped to each AL and the frameworks

For both assessments, these steps were followed in inaugural administrations to establish score scales, ALs, and initial ALDs. However, unlike for NAEP, in subsequent administration cycles of both PISA and TIMSS, minor revisions of the ALDs routinely occurred based on revisions to the frameworks and new item types introduced into the assessments. The score scales and locations of the ALs on the scales were assumed to remain the same by virtue of statistical linking utilizing sets of common items across assessment cycles.

In the case of PISA, interpretive materials emphasize their ability to report item difficulty and examinee proficiency on the same continuous scale: "By showing the difficulty of each question on this scale, it is possible to locate the level of proficiency in the subject that the

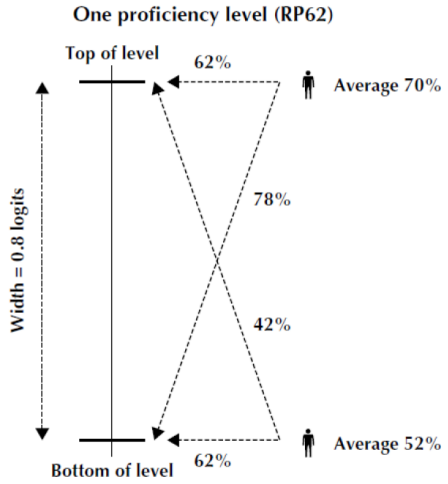
question demands. By showing the proficiency of test-takers on the same scale, it is possible to describe each test-taker's level of skill or literacy by the type of tasks that he or she can perform correctly most of the time" (OECD, 2019, p. 43). In practice, mapping items to ALs ("proficiency levels" in PISA nomenclature) is more complicated. PISA chooses to map all items to proficiency levels and to use the information from all of the items in developing ALDs. This is in contrast to more traditional anchor studies, in which only subsets of items that are similar in difficulty and that discriminate well at a particular AL are assigned to represent the AL.

As stated in the most recent PISA Technical Report, this item mapping approach is characterized by three variables (OECD, 2020, chapter 15):

- The expected success of a student at a particular level on a test containing items at that level (proposed to be set at a minimum that is near 50% for the student at the bottom of the level and greater for students who are higher in the level).
- The width of the levels in that scale (determined largely by substantive considerations of the cognitive demands of items at the level and data related to student performance on the items).
- The probability that a student in the middle of a level would correctly answer an item of average difficulty for that level (in fact, the probability that a student at any particular level would get an item at the same level correct), sometimes referred to as the "RP value" for the scale, where "RP" indicates "response probability."

To accommodate this approach, PISA developed a solution with three features: 1) proficiency levels are defined to have equal intervals (0.8 logits on the underlying IRT scale); 2) items are mapped to levels such that students at the bottom of the level will answer about 50 percent of items spread across the level; and 3) a model-based response probability of 62 percent (RP62) applies to both students at the bottom of the level taking items mapped at the bottom of the level and students at the top of the level taking items mapped at the top of the level. Figure 2 below depicts this approach graphically (see OECD, 2020, chapter 15 for more details). Initially, PISA defined six levels ranging from Level 1 to Level 6, but in more recent administrations has broken Level 1 into sub-levels (i.e., Level 1a, 1b, 1c). One consequence of this approach that complicates score interpretation is that the thresholds for ALs are located at different scale score values for the reading, math, and science score scales.

Figure 2: Calculating the RP Values Used to Define PISA Proficiency Levels



Note. Figure is reprinted with permission from OECD, 2020, Chapter 15, Figure 15.2.

In contrast to PISA, TIMSS utilizes a scale anchoring approach much more similar to anchoring studies used for NAEP (e.g., Donohue, Pitoniak, & Beaulieu, 2010; Pitoniak, Dion, & Garber, 2010). For both math and science, TIMSS establishes "International Benchmarks" that describe what students know and can do at four points along their achievement scale: Advanced International Benchmark (625), High International Benchmark (550), Intermediate International Benchmark (475), and Low International Benchmark (400). The scale anchoring process involves calculating the mean percent correct for students scoring within 5 scale-score points of each benchmark and assigning rules for classifying the items based on these four conditional percent correct values. In general, multiple-choice items will be mapped to a particular Benchmark level if they have a percent correct value of at least 65 percent at that level and a percent correct value of less than 50 at the next level below. Items are mapped to the Low International Benchmark based only on having a percent correct value of at least 65 percent (since it is the lowest level) and constructed response items are mapped at the level where the percent correct is at least 50 percent (because correct answers from guessing are unlikely). A detailed description of the TIMSS scale anchoring procedures for the most recent TIMSS assessment can be found in Mullis and Fishbein (2020).

One consequence of the TIMSS scale anchoring process is that not all items included in the assessment meet the scale anchoring criteria. Thus, to expand the items available to content experts in developing ALDs, additional items are included based on slightly relaxed criteria; these are designated "almost anchored" items. For the 2019 TIMSS benchmarks, between 174 and 274 items anchored and almost anchored items were available to committees for developing the ALDs (Mullis & Fishbein, 2020).

In addition to differences in the details of how PISA and TIMSS develop ALDs, the ALDs for the two assessments are also different. Olsen and Nilsen (2017) note that TIMSS assessments include many more items than PISA assessments, with large numbers of items meeting anchoring or almost anchoring criteria, resulting in ALDs that reflect clear

progressions from one Benchmark level to the next. With fewer items mapping to six or more ALs, the progression in proficiency from the low to the high levels in the PISA ALDs may be less apparent to the reader. Olsen and Nilsen also state that TIMSS ALDs are relatively long and reveal a closer reference to the content of the items. This is supported by the fact that TIMSS technical appendices include the short, item-level descriptions for every item that is included in their anchoring studies. In contrast, PISA has developed ALDs with shorter generic statements that more closely resemble a theory of what constitutes progress in their defined constructs and that are more stable to changes in the items that appear across assessment cycles.

Increased Relevance to NAEP

In response to the *NAS* report, the Governing Board commissioned a HumRRO report addressing reporting achievement level descriptors for NAEP (Michels, Egan, Thacker & Schultz, 2018). According to the authors, reporting ALDs play a critical role in communicating assessment results with relevant stakeholders: "Reporting ALDs are developed once cut scores have been established such that the KSAs articulated in reporting ALDs are based on student test performance." (p. 2).

Soon after the HumRRO report was completed, the Governing Board released a revised policy statement on developing student achievement levels for NAEP (National Assessment Governing Board, 2018). In this statement, reporting ALDs were established as an element of NAEP achievement levels to provide validity evidence for the intended uses and interpretations and help make NAEP results more informative to policy makers, educators, and the public:

> To develop ALDs for reporting, following the achievement level setting the Board shall revisit and may revise content ALDs to ensure that they are consistent with empirical evidence of student performance. In particular, these "Reporting ALDs" chosen to illustrate the knowledge and skills demonstrated at different achievement levels shall be written to incorporate empirical data from student performance. Reporting ALDs shall describe what students at each level do know and can do rather than what they should know and should be able to do (p. 9).

In many ways, the adoption of reporting ALDs for NAEP signals a possible shift towards the approaches taken by the international assessments with respect to ALs and ALDs. As described in the recent *NAGB Achievement Levels Work Plan* (National Assessment Governing Board, 2020b), a "model-based anchoring approach" will focus on the current reporting ALDs for mathematics and reading at Grades 4, 8, and 12 using methods similar to prior scale anchoring studies (Donohue, Pitoniak, & Beaulieu, 2010; Pitoniak, Dion, & Garber, 2010). The study will be used to revise the current ALDs as needed to create reporting ALDs that indicate what students at each achievement level do know and can do. As reporting ALDs become more routinely established for NAEP assessments, the international assessments provide examples of how item mapping and the scale anchoring results can be used more coherently to make NAEP ALs and ALDs more understandable to the public.

For example, the TIMSS technical documentation lists the item descriptors for all items used in each scale anchoring exercise as well as the ALDs and selected released items that serve as exemplars for the different ALs. The *NAEP Achievement Levels Procedures Manual* (National

Assessment Governing Board, 2020a) calls for exactly this type of documentation as part of anchor studies: "Finally, the draft Reporting ALDs should be evaluated relative to the exemplar items to represent each achievement level. It is important that the exemplar items serve to illustrate the performance described in the reporting ALDs" (p. 33).

As NAEP anchor studies become more common and expand to other content areas, it is likely that consistent statistical criteria (i.e., RP values) will be adopted. Both PISA and TIMSS have applied consistent statistical approaches for mapping items to their scale (although their approaches differ from each other). Loomis (2018) reviewed a variety of anchoring studies done for NAEP up until that time, including studies done by NCES as well as studies conducted by the Governing Board. Although she noted that different criteria had been used over time, she recommended standardizing anchor study criteria and pointed to criteria used in the Donohue, Pitoniak, and Beaulieu (2010) and Pitoniak, Dion, and Garber (2010) studies. Judging from the proposal for the model-based anchoring study for math and reading at Grades 4, 8, and 12 (Pearson, 2020), standardized criteria for ongoing NAEP anchor studies now appear to be in place.

A final advantage that can result from the Governing Board's decision to establish reporting ALDs through the use of anchor studies is that they provide a mechanism for periodic review of achievement levels or even more substantial changes, such as a new or revised framework. For example, the revised NAEP achievement level policy statement (National Assessment Governing Board, 2018) includes the following comment:

> If a framework is replaced or revised for a major update, a new achievement level setting process may be implemented, except in circumstances where scale score trends are maintained. In this latter instance, COSDAM shall determine how to revise the ALDs and review the cut scores to ensure that they remain reasonable and meaningful (p. 10).

The international assessments frequently revise their content frameworks, and anchor studies are the vehicle through which the changes are accommodated and incorporated into ALDs while scale score trends are maintained. The Governing Board has now incorporated that option into their achievement levels toolkit. In fact, even if content and construct changes are judged to be too great to maintain scale score trends through statistical linking, the inclusion of reporting ALDs could offer an option for bridging between old and new versions of NAEP assessments more seamlessly. At this point in NAEP's history, the ALs have a well-established normative frame of reference. Thus, when new reading and math frameworks are implemented, there will still be a public expectation based on the stable trends for the percentages of students classified at *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced* levels on these assessments over so many years. The challenge in setting achievement levels is that recommendations resulting from the process could be inconsistent with those trends, in which case the Governing Board will be in a situation where they either have to overrule the panel recommendations to achieve consistency with past trends (which has proven controversial in the past) or accept results and face potential controversy and interpretive confusion because of changes to the trends.

NAEP's achievement level setting process is unparalleled in terms of its maturity and thoroughness, and because of this it is likely that new achievement level setting activities will

end in a reasonable place. But this will come at a significant investment of time, effort, and cost. As an alternative to achievement level setting with the introduction of a new content framework, it is possible to break trend but still align scales (e.g., through equipercentile linking) to preserve a well-established interpretive frame of reference. In such an instance, scale anchoring could be used without a new achievement level setting study to establish the reporting ALDs and to revise the new content ALDs. In effect, the messaging would be something like, yes, the test is now different and results for the new test should not be compared to the old test, but we are starting in a similar place with respect to the percentages of students falling in each NAEP ALs, and the reporting ALDs give you information about what these levels mean in terms of the content on the new test. Resources saved from this alternative approach could be reallocated to other studies designed to achieve the goals established in the *Achievement Levels Work Plan*. Although this would be a radical break with tradition, the examples provided by the international assessments suggest that it would be defensible—and possibly better—practice.

CONCLUDING REMARKS

This paper consisted of two very different efforts intended to provide NAGB with contextual data and perspective that might be considered in building a case to support removal of the NAEP achievement levels' "trial status." The first section of the paper described a survey of state assessment coordinators and directors that documented the influence of NAEP on how states have set and maintained achievement levels for their own assessments, as well as how NAEP results are interpreted in conjunction with state assessment results. In the second section of the paper, we compared and contrasted the approaches and practices of international assessments related to setting and maintaining achievement levels with those of NAEP. Through this exercise, we noted the relevance of certain international practices to NAGB's response to the recent NAS evaluation.

In different ways, each effort helps to illustrate the maturity and influence of the NAEP achievement levels. Moreover, recent changes to the NAEP achievement levels policy statement and actions taken by NAGB to more explicitly align the NAEP content frameworks, item pools, achievement-level descriptors, and cut scores are directly responsive to NAS evaluation recommendations. Although there will likely remain some challenges stemming from potential disconnects between the aspirational genesis of the NAEP achievement levels and performance on actual NAEP items, their status seems reasonably permanent at this point.

REFERENCES

Bourque, M. L. (2009). *A history of NAEP achievement levels: Issues, implementation, and impact 1989–2009*. Paper commissioned for the 20th anniversary of the National Assessment Governing Board. https://www.nagb.gov/content/dam/nagb/en/documents/who-we-are/20-anniversary/bourque-achievement-levels-formatted.pdf

Donahue, P., Pitoniak, M., & Beaulieu, N. (2010). *Final report on the study to draft achievement-level descriptions for reporting results of the 2009 National Assessment of Educational Progress in reading for Grades 4, 8, and 12*. Educational Testing Service.

Loomis, S. (2018). *Anchor studies for analysis of NAEP achievement levels*. Paper prepared for an Expert Panel Meeting on NAEP Achievement Levels, July 12–13, 2018. https://www.nagb.gov/content/dam/nagb/en/documents/publications/achievement/Anchor-Studies-for-Analysis-of-NAEP-Achievement-Levels.pdf

Michaels, H., Egan, K., Thacker, A., & Schultz, S. (2018). *Reporting achievement level descriptors for the National Assessment of Educational Progress*. HumRRO. https://www.nagb.gov/content/dam/nagb/en/documents/publications/achievement/Reporting-Achievement-Level-Descriptors.pdf

Mullis, I. V. S, & Fishbein, B. (2020). Using scale anchoring to interpret the TIMSS 2019 achievement scales. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and procedures: TIMSS 2019 technical report*. Boston College, TIMSS & PIRLS International Study Center: https://timssandpirls.bc.edu/timss2019/methods

National Academies of Sciences, Engineering, and Medicine. (2017). *Evaluation of the achievement levels for mathematics and reading on the National Assessment of Educational Progress*. The National Academies Press. https://doi.org/10.17226/23409.

National Assessment Governing Board. (2018). *Developing student achievement levels for the National Assessment of Educational Progress: Policy statement*. https://www.nagb.gov/content/nagb/assets/documents/policies/ALS-revised-policy-statement-11-17-18.pdf

National Assessment Governing Board. (2020a). *NAEP achievement levels procedures manual.*

National Assessment Governing Board. (2020b). *NAGB achievement levels work plan.*

National Assessment Governing Board. (2020c). *Update to achievement levels work plan.*

OECD. (2019). *PISA 2018 results (volume I): What students know and can do*. OECD Publishing. https://doi.org/10.1787/5f07c754-en

OECD. (2020). *PISA 2018 technical report*. https://www.oecd.org/pisa/data/pisa2018technicalreport/

Olsen, R. V., & Nilsen, T. (2017). Standard setting in PISA and TIMSS and how these procedures can be used nationally. In S. Blömeke & J.-E. Gustafsson (Eds.), *Standard Setting in Education*, Methodology of Educational Measurement and Assessment. https://doi.org/10.1007/978-3-319-50856-6_5

Pearson. (2020*). Developing achievement level descriptions for mathematics and reading* (Design Document). National Assessment Governing Board.

Phillips, G. W., Mullis, I. V. S., Bourque, M. L., Williams, P. L., Hambleton, R. K., Owen, E. H., & Barton, P. E. (1993). *Interpreting NAEP scales*. Office of Educational Research and Improvement, U.S. Department of Education.

Pitoniak, M., Dion, G., & Garber, D. (2010). *Final report on the study to draft achievement-level descriptions for reporting results of the 2009 National Assessment of Educational Progress in mathematics for grade 12*. Educational Testing Service.

APPENDIX A. NAEP ACHIEVEMENT LEVEL DESCRIPTORS

NAEP Achievement Level Descriptors for Reading Grade 4

NAEP Basic

Fourth-grade students performing at the *NAEP Basic* level should be able to locate relevant information, make simple inferences, and use their understanding of the text to identify details that support a given interpretation or conclusion. Students should be able to interpret the meaning of a word as it is used in the text.

When reading literary texts such as fiction, poetry, and literary nonfiction, fourth-grade students performing at the *NAEP Basic* level should be able to make simple inferences about characters, events, plot, and setting. They should be able to identify a problem in a story and relevant information that supports an interpretation of a text.

When reading informational texts such as articles and excerpts from books, fourth-grade students performing at the *NAEP Basic* level should be able to identify the main purpose and an explicitly stated main idea, as well as gather information from various parts of a text to provide supporting information.

NAEP Proficient

Fourth-grade students performing at the *NAEP Proficient* level should be able to integrate and interpret texts and apply their understanding of the text to draw conclusions and make evaluations.

When reading literary texts such as fiction, poetry, and literary nonfiction, fourth-grade students performing at the *NAEP Proficient* level should be able to identify implicit main ideas and recognize relevant information that supports them. Students should be able to judge elements of author's craft and provide some support for their judgment. They should be able to analyze character roles, actions, feelings, and motives.

When reading informational texts such as articles and excerpts from books, fourth-grade students performing at the *NAEP Proficient* level should be able to locate relevant information, integrate information across texts, and evaluate the way an author presents information. Student performance at this level should demonstrate an understanding of the purpose for text features and an ability to integrate information from headings, text boxes, graphics and their captions. They should be able to explain a simple cause-and-effect relationship and draw conclusions.

NAEP Advanced

Fourth-grade students performing at the *NAEP Advanced* level should be able to make complex inferences and construct and support their inferential understanding of the text. Students should be able to apply their under-standing of a text to make and support a judgment.

When reading literary texts such as fiction, poetry, and literary nonfiction, fourth-grade students performing at the *NAEP Advanced* level should be able to identify the theme in stories and poems and make complex inferences about characters' traits, feelings,

motivations, and actions. They should be able to recognize characters' perspectives and evaluate character motivation. Students should be able to interpret characteristics of poems and evaluate aspects of text organization.

When reading informational texts such as articles and excerpts from books, fourth-grade students performing at the *NAEP Advanced* level should be able to make complex inferences about main ideas and supporting ideas. They should be able to express a judgment about the text and about text features and support the judgment with evidence. They should be able to identify the most likely cause given an effect, explain an author's point of view, and compare ideas across two texts.

Links to Other NAEP Achievement Level Descriptors

Reading achievement level descriptors (all grades):
https://nces.ed.gov/nationsreportcard/reading/achieve.aspx

Mathematics achievement level descriptors (all grades):
https://nces.ed.gov/nationsreportcard/mathematics/achieve.aspx

Links to achievement levels for all subjects can be found here:
https://nces.ed.gov/nationsreportcard/guides/scores_achv.aspx

APPENDIX B. SURVEY INSTRUMENT

NAEP Validity Studies Panel Survey About State Achievement Levels and Descriptors

Please answer the following questions about the state assessments in Mathematics and ELA that are used in your state for accountability purposes. Please consult with Assessment Directors and other knowledgeable staff in your state assessment office as appropriate.

1. A) Are the achievement levels and labels provided below the ones currently used on your state's assessments?

In Survey Monkey, this question will be pre-populated with the achievement levels that we found for your state. A table with this information for all states is provided in Appendix A. Please refer to this table in Appendix A to check information for your state.

1. B) Is this the correct source for information about the achievement levels and achievement-level descriptors (ALDs) in your state?

In Survey Monkey, this question will be pre-populated with a link to the location where we found information for your state. A table with these links for all states is provided in Appendix B. Please refer to this table in Appendix B to check the link for your state.

Comment?

2. Was the setting of your state's current achievement levels (cut scores) and descriptors influenced at all by the NAEP achievement levels and/or ALDs? (Select one per row)

	No	**Yes, a little**	**Yes, to a considerable degree**
Achievement levels	☐	☐	☐
ALDs	☐	☐	☐

Comment?

3. Have your state's achievement levels changed in the past 5 years? If so, please indicate the rationale(s) for making the changes. (Select all that apply.)

☐	There have been no changes.
☐	Changes were initiated based on existing policy (e.g. revisiting standards on a fixed schedule).
☐	Changes were initiated due to the adoption of a new assessment design.
☐	Changes were initiated due to a change in policy as determined by an individual or group (e.g., Chief State School Officer, State Board of Education, legislature).
☐	Changes were initiated for practical or educational reasons, such as the standards were judged to be too high or too low.

Comment?

4. Achievement-level descriptors are sometimes changed over time in either minor or substantial ways. Which of the following statements best describes the status of the ALDs in your state over the past 5 years? (Select all that apply.)

☐	No changes were made.
☐	Changes were made due to changes to the achievement levels.
☐	Changes were partly influenced by NAEP.
☐	Changes were not influenced by NAEP.

Comment?

5. Were any empirical methods, other than sharing impact data, used in the process of setting the current achievement levels and/or ALDs? (Select all that apply.)

☐	No.
☐	Relationships with criterion variables (e.g., benchmarking).
☐	Scale anchoring.
☐	Relationships with existing assessments (e.g., NAEP, ACT, SAT) such as using an equipercentile procedure to link to benchmarks in a comparison assessment.
☐	Other (specify).

Comment?

6. After the achievement levels and/or ALDs were established, was any additional validity evidence collected about them?

☐	No.
☐	Yes, student performance on tests other than the state tests was examined.
☐	Yes, validity evidence was collected from educators in the state.
☐	Yes, one or more alternate standard-setting procedures were employed to validate the achievement levels and/or ALDs.
☐	Other (specify).

Comment?

7. How much influence do your state's current achievement levels and/or ALDs have when your state's assessment results are released to stakeholders?

☐	None.
☐	Some. The release of results occurs in a variety of ways, some of which include the percentage of students meeting each achievement level.
☐	Considerable. The major focus of each release is on the performance of students in relation to the achievement levels, but some of the discussion does not relate to the achievement levels or ALDs.
☐	Exclusive. The entire focus of each release is on the achievement levels and ALDs.

Comment?

8. Have the current achievement levels and/or ALDs been judged effective by your state department for communicating statewide assessment results?

☐	No.
☐	Somewhat.
☐	Yes.

Comment?

9. In years when NAEP releases state results, how are they and the state assessment results treated when discussed with stakeholders?

☐	There is no attempt to compare the results from the two assessments as they are viewed as entirely different entities.
☐	There is some consideration given to the results of both assessments, but differences between them (e.g., timing, content) present significant limitations to interpretation.
☐	There is a great deal of consideration given to the results of both assessments, as they are viewed as providing two valuable perspectives on student achievement.

Comment?

10. Does the state use any especially effective reporting strategy for the state assessment that might be considered by NAEP?

Please describe.

APPENDIX C. SUPPLEMENTAL TABLES

Table C1. Achievement Level Descriptors by State

State	ALD List				
Alabama	Exceeds Standards: Level IV	Meets Standards: Level III	Partially Meets Standards: Level II	Does Not Meet Standards: Level I	
Alaska	Advanced	Proficient	Below Proficient	Far Below Proficient	
Arizona	Exceeds the Standard	Meets the Standard	Falls Far Below the Standard	Approaches the Standard	
Arkansas	Independent	Functional Independence	Supported Independence	Emergent	
California	Standard Exceeded	Standard Met	Standard Nearly Met	Standard Not Met	
Colorado	Advanced	Proficient	Partially Proficient	Unsatisfactory	
Connecticut	Exceeds the Achievement Standard	Meets the Achievement Standard	Approaching the Achievement Standard	Does Not Meet the Achievement Standard	
Delaware	Advanced	Meets the Standard	Below the Standard	Well Below the Standard	
District of Columbia	Advanced	Proficient	Basic	Below Basic	
Florida	Meets the Achievement Standard	Level 4	Level 3	Level 2	
Georgia	Level 4	Level 3	Level 2	Level 1	
Hawaii	Level 4 (Exceeded)	Level 3 (Met)	Level 2 (Nearly Met)	Level 1 (Not Met)	
Idaho	Advanced	Proficient	Basic	Below Basic	
Illinois	Exceeded Expectations	Met Expectations	Approached Expectations	Partially Met Expectations	Did Not Yet Meet Expectations
Indiana	Above Proficiency	At Proficiency	Approaching Proficiency	Below Proficiency	
Iowa	Advanced	Early Advanced	Intermediate	Early Intermediate	Beginning
Kansas	Excellent Ability	Effective Ability	Basic Ability	Limited Ability	
Kentucky	Distinguished	Proficient	Apprentice	Novice	
Louisiana	Advanced	Early Advanced	Intermediate	Early Intermediate	Beginning
Maine	Above State Expectations	At State Expectations	Below State Expectations	Well Below State Expectations	
Maryland	Advanced	Proficient	Basic		
Massachusetts	Exceeding Expectations	Meeting Expectations	Partially Meeting Expectations	Not Meeting Expectations	
Michigan	Surpassed the Performance Standard	Attained the Performance Standard	Emerging toward the Performance Standard		
Minnesota	Exceeds Expectations	Meets Expectations	Partially Meets Expectations		
Mississippi	Advanced	Proficient	Basic		
Missouri	Advanced	Proficient	Basic	Below Basic	

State	ALD List				
Montana	Advanced	Proficient	Nearing Proficiency	Novice	
Nebraska	CCR Benchmark	On Track	Developing		
Nevada	Level 4	Level 3	Level 2	Level 1	
New Hampshire	Highly Proficient	Proficient	Approaching Proficient	Below Proficient	
New Jersey	Advanced Proficient	Proficient	Partially Proficient		
New Mexico	Exceeds Expectations	Meets Expectations	Approaches Expectations	Partially or Does Not Yet Meet Expectations	
New York	Level 4	Level 3	Level 2	Level 1	
North Carolina	Level 5	Level 4	Level 3	Not Proficient	
North Dakota	Advanced	Proficient	Partially Proficient	Novice	
Ohio	Advanced	Accelerated	Proficient	Basic	Limited
Oklahoma	Advanced	Proficient	Basic	Below Basic	
Oregon	Exceeds	Meets	Nearly Meets	Does Not Yet Meet	
Pennsylvania	Advanced	Proficient	Basic	Below Basic	
Rhode Island	Exceeding Expectations	Meeting Expectations	Partially Meeting Expectations	Not Meeting Expectations	
South Carolina	Exceeds Expectations	Meets Expectations	Approaches Expectations	Does Not Meet Expectations	
South Dakota	Advanced	Proficient	Basic		
Tennessee	Level 4	Level 3	Level 2	Level 1	
Texas	Masters Grade Level	Meets Grade Level	Approaches Grade Level	Did Not Meet Grade Level	
Utah	Advanced	At Target	Approaching the Target	Emerging	
Vermont	Exceeds	Meets	Approaching	Beginning	
Virginia	Pass/Advanced	Pass/Proficient	Fail/Basic	Fail/Below Basic	
Washington	Level 4	Level 3	Level 2	Level 1	
West Virginia	Exceeds Standard	Meets Standard	Partially Meets Standard	Does Not Meet Standard	
Wisconsin	Advanced	Proficient	Basic	Below Basic	
Wyoming	Advanced	Proficient	Basic	Below Basic	

Table C2. Achievement Level Descriptor Themes – Detailed

Advanced/Proficient Schemes	**N**
Advanced, Proficient, Basic, Below Basic/Pre-Basic	8
Advanced, Proficient, Below Proficient, Far Below Proficient	1
Advanced, Proficient, Partially Proficient, Not Proficient/Novice	3
Advanced, Proficient, Partially/Not Yet Proficient	2
Advanced, Proficient, Passing, Basic, Minimal	1
Total "Advanced/Proficient" scheme	**15**
Exceed/Meets Schemes	**N**
Exceeding Expectations, Meeting Expectations, Partially Meeting Expectations, Not Meeting Expectations	4
Exceeds, Meets, Approaches, Beginning/Partially/Does Not Meet	7
Total "Exceed/Meets" scheme	**11**
Level *N* Schemes	**N**
Level 4, Level 3, Level 2, Level 1	6

Level 4 (Exceeded), Level 3 (Met), Level 2 (Nearly Met), Level 1 (Not Met)	2
Level 5, Level 4, Level 3, Level 2, Level 1	1
Level 5 [description], Level 4 [description], Level 3 [description], Level 2 [description], Level 1 [description]	2
Total "Level *N*" scheme	**11**
Other Schemes	**N**
Highly Proficient/Above Proficiency, Proficient/At Proficiency, Approaching, Below	3
Above State Expectations, At State Expectations, Below State Expectations, Well Below State Expectations	1
Advanced, Accelerated, Proficient, Basic, Limited	1
Advanced, Mastery, Basic, Approaching Basic, Unsatisfactory	1
Attaining, Progressing, Emerging, Attempting	1
CCR Benchmark, On Track, Developing	1
Distinguished, Proficient, Apprentice, Novice	1
Exceeding, Ready, Close, Needs Support, Did Not Yet Meet Expectations	1
Excellent Ability, Effective Ability, Basic Ability, Limited Ability	1
Mastered, On Track, Approaching, Below	1
Masters Grade Level, Meets Grade Level, Approaches Grade Level, Did Not Meet Grade Level	1
Meets the Standards, Partially Meets the Standards, Does Not Meet the Standards	1
Pass/Advanced, Pass/Proficient, Fail/Basic, Fail/Below Basic	1
Total other schemes	**15**