Committee on Standards, Design and Methodology

August 3, 2021

1:00 – 3:00 pm ET (Virtual)

AGENDA

1:00 – 1:40 pm	**Item Difficulty and Student Ability Distributions (CLOSED)** *Gregory Cizek, Chair* *Enis Dogan, National Center for Education Statistics*	Attachment A
1:45 – 2:35 pm	**Improving Information about Students Scoring Below the *NAEP Basic* Achievement Level** *Gregory Cizek* *Karla Egan, EdMetric* *Jing Chen, National Center for Education Statistics* *Taslima Rahman, National Center for Education Statistics*	Attachment B
2:35 – 2:50 pm	**Update: Review and Revision of Mathematics and Reading Achievement Level Descriptions** *Eric Moyer, Pearson*	Attachment C
2:50 – 3:00 pm	**Next Steps** *Gregory Cizek*	
Information Item	**Strategic Vision 2025 Update**	See Executive Committee Material

IES Institute of Education Sciences

Item Difficulty and Student Ability Distributions in NAEP

Would you include the following item in a fourth-grade assessment?

1+1= …

How about this one?

Solve for x, where $\log_x 81 = 4$

Obviously, the answer is no in both cases. Setting aside the fact that these items would not be measuring skills in a fourth-grade assessment framework, the items would not provide any "information" about a (typical) fourth-grader's mathematics "ability." There is not a good alignment between the student ability and these two (hypothetical) items; you already know how the student would perform on these items. This example is to illustrate that items should not be too difficult, nor too easy for the students—they need to be … "just right"! In fact, the level of "information" an assessment provides is proportional to the degree of alignment between student ability and item difficulty. The most efficient way to achieve such alignment is through adaptive testing, where items are selected for the student in a way that their difficulty match his/her "ability."

But what to do in a linear test, where items are not selected this way? The solution is less efficient, yet quite straightforward—include items in your assessment that are "just right" for all kinds of students, ranging from poor performing to highest performing. **In other words, make sure the item pool varies in difficulty as much as the students vary in "ability" and that it includes items "just right" for every student.** If, say, 20% of the students (typically) perform below a certain score, maybe allocate 20% of the item pool to items that are "just right" for this score range.

In this closed session, we will look at the degree of alignment between student "ability" and difficulty of the items in NAEP assessments. Specifically, we will look at what percent of our students perform below the *NAEP Basic* range and compare that to the percentage of items that are "just right" in this score range across a number of grades and subjects. We will do this by examining item/person maps that show student score distributions and the distributions of difficulty of items side-by-side. We will also touch on past and present efforts in improving the alignment between student "ability" and difficulty of the items in NAEP assessments.

Improving Information about Students Scoring Below the *NAEP Basic* Achievement Level

One of the Governing Board's most important legislated responsibilities is developing the NAEP achievement levels. The Board policy on Developing Student Achievement Levels for NAEP defines three achievement levels: *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced*. The policy specifies that, "The remaining region that falls below the *NAEP Basic* cut score shall be identified as 'below *NAEP Basic'* when a descriptor is necessary." The percentage of students in this lowest category is reported but there is no accompanying achievement level description (ALD).

During recent COSDAM meetings, there has been discussion about the importance of better understanding what the lowest performing students know and can do. Some COSDAM members requested additional discussion about potential options for improving measurement and reporting below the *NAEP Basic* achievement level.

Three background materials are included to inform this August COSDAM discussion:

1. A literature review conducted by Karla Egan of EdMetric that Board staff commissioned at the request of COSDAM Chair Gregory Cizek, to better understand:
    - the number of state assessments that have Below Basic ALDs;
    - the nature of Below Basic ALDs and how they differ from other categories;
    - the pros and cons of including Below Basic ALDs in state assessments; and
    - additional considerations relevant to NAEP;

2. A summary report from a National Center for Education Statistics (NCES) panel convened on December 16-17, 2020, to explore how NAEP can better measure, describe, and report the skills and knowledge of lower-performing students, particularly those below *NAEP Basic;* and

3. A highlights report from a recent NCES special study that was conducted to better understand the knowledge and skills possessed by students who performed below the *NAEP Basic* achievement level on the grade 4 reading assessment. Additional information about the study, along with supplementary materials including audio recordings, can be found on the study website at: https://nces.ed.gov/nationsreportcard/studies/orf/.

During the August COSDAM meeting, there will be very brief presentations (approximately 5 minutes each) on the background materials listed above, followed by Committee discussion.

2021 No. 025

Describing the Lowest Achievement Level

Final Report

Prepared for: National Assessment Governing Board
800 North Capitol Street N.W., Suite 825
Washington DC 20002

Prepared under: ED-NAG-17-C-0002

Authors: Karla Egan, Ph.D.
EdMetric LLC

Date: March 31, 2021

Describing the Lowest Achievement Level

Table of Contents

Defining Achievement Level Descriptors 1

The Lowest Achievement Level 4

What do Other Testing Entities Do? 4

States: Reporting and Policy ALDs 5

States: Range ALDs 19

TIMSS and PISA 19

Other Research 20

Demographic Composition of the Lowest Achievement Level 20

Discussion 22

References 23

List of Tables

Table 1. Types and Uses of ALDs: Proposed Revised Governing Board Policy and Procedures (Governing Board Achievement Levels Procedures Manual, June 2020) 2

Table 2. Grade 8 Mathematics NAEP Achievement Level Descriptors 3

Table 3. Numbers of States Reporting at the Lowest Performance Category 5

Table 4. Descriptors for Lowest Achievement Level 6

List of Figures

Figure 1. Percentage of Students in Below the NAEP Basic achievement level compared to NAEP Basic and Above for Black, Hispanic, and White Students, NAEP Grade 8 Mathematics 21

Describing the Lowest Achievement Level

The use of achievement level descriptors (ALDs) is an accepted part of K–12 assessment. ALDs are used at standard setting, written for reports, developed for teachers to guide learning and instruction, and used by content experts to write items for large-scale assessments. Even though various aspects of ALDs have been explored in literature, one aspect of ALDs remains unexamined—the description of the lowest achievement level. Testing experts disagree on whether descriptors should be written for the lowest achievement level. This paper examines the utility and appropriateness of writing a descriptor for the lowest achievement level. In 1995, the Governing Board's policy on NAEP achievement levels said subject-matter ALDs are "articulated in terms of what students should know and should be able to do" (1995, p. 8) and "they are not written for content below the *Basic* level" (1995, p. 8). In 2018, the Governing Board reaffirmed this approach when the policy was updated to say, "There shall be no content ALDs developed for performance below the *NAEP Basic* level" (2018, p. 6). Therefore, since the inception of ALDs in the 1990s, the Governing Board only provides ALDs for *Basic, Proficient,* and *Advanced.* This paper first defines ALDs, then examines why ALDs are not written for the lowest level and examines the current state of the field regarding descriptors for the lowest achievement level.

Defining Achievement Level Descriptors

Achievement level descriptors (sometimes called performance level descriptors) are now ubiquitous in K–12 assessment programs. These descriptors define the types of knowledge, skills, and abilities of students at different levels of performance. The specificity of the ALD depends on the use of the ALD. For this reason, the Governing Board adopted *content ALDs*, an umbrella term that encompasses framework ALDs, threshold ALDs, and reporting ALDs. Table 1 illustrates this relationship along with the uses of each type of ALD.

The current paper focuses on policy ALDs and reporting ALDs. In a typical state program, policy ALDs are high-level definitions of the types of performance expected in each achievement level. For NAEP, the policy ALDs include:

- **NAEP Advanced**. This level signifies superior performance beyond NAEP Proficient.
- **NAEP Proficient.** This level represents solid academic performance for each NAEP assessment. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real world situations, and analytical skills appropriate to the subject matter.
- **NAEP Basic.** This level denotes partial mastery of prerequisite knowledge and skills that are fundamental for performance at the NAEP Proficient level.
- **Below NAEP Basic.** No descriptor.

Table 1. Types and Uses of ALDs: Proposed Revised Governing Board Policy and Procedures (Governing Board Achievement Levels Procedures Manual, June 2020)

Policy definitions: The policy defines three NAEP achievement levels: NAEP Basic, NAEP Proficient, and NAEP Advanced. These policy definitions apply to all main NAEP assessments.	
Content ALDs	
ALDs in Framework (for item development and achievement-level setting)	Under the revised policy and procedures for framework development, the framework development panel may develop multiple sets of content ALDs for the purposes of informing item development and for use in the achievement-level setting activities. The framework development panel might also determine that one set of ALDs can serve both of these purposes. These ALDs will continue to be written in terms of what students should know and be able to do. If there is a specific need to revise the content ALDs in advance of an achievement-level setting, then a separate activity will be undertaken to do so, but this is not intended to be necessary in most cases.
Threshold/Borderline ALDs (if applicable)	If descriptions of performance right at the cut scores are needed for setting achievement levels (e.g., if a Bookmark or similar procedure is used), then threshold (or borderline) ALDs will be developed by achievement-level setting panelists. Threshold ALDs are for the panelists' own use and are not reported with the NAEP results. The rationale for having the achievement-level setting panelists create threshold ALDs (rather than providing them at the beginning of the process) is that it is an important task to help panelists fully internalize the ALDs. Because the creation of threshold ALDs is an instrumental activity that occurs as part of the achievement-level setting process, panelists are typically discouraged from spending inordinate amounts of time on their development or focusing on minor edits and wordsmithing.
Reporting ALDs	Reporting ALDs are developed following the first operational administration of an assessment and express the empirical findings as to what students have demonstrated they know or can do at each achievement level. The policy calls for conducting a study to derive the reporting ALDs following the first operational administration of an assessment (and again every 3 administrations or 10 years, whichever comes later).

The reporting ALDs describe the knowledge, skills, and abilities that students in each achievement level (e.g., NAEP Basic, NAEP Proficient, NAEP Advanced) demonstrate. Table 2 provides the current reporting ALDs for Grade 8 mathematics NAEP; however, the ALDs in Table 2 are currently being updated to reflect the new requirement in the NAEP policy that reporting ALDs incorporate empirical data on student performance and describe what students *do* know and *can* do rather on what they should know and be able to do.

Table 2. Grade 8 Mathematics NAEP Achievement Level Descriptors

Achievement Level	Description
NAEP Advanced (333)	**Eighth-grade students performing at the *NAEP Advanced* level should be able to reach beyond the recognition, identification, and application of mathematical rules in order to generalize and synthesize concepts and principles in the five NAEP content areas.** Eighth-graders performing at the *NAEP Advanced* level should be able to probe examples and counterexamples in order to shape generalizations from which they can develop models. Eighth-graders performing at the *NAEP Advanced* level should use number sense and geometric awareness to consider the reasonableness of an answer. They are expected to use abstract thinking to create unique problem-solving techniques and explain the reasoning processes underlying their conclusions.
NAEP Proficient (299)	**Eighth-grade students performing at the *NAEP Proficient* level should apply mathematical concepts and procedures consistently to complex problems in the five NAEP content areas.** Eighth-graders performing at the *NAEP Proficient* level should be able to conjecture, defend their ideas, and give supporting examples. They should understand the connections between fractions, percents, decimals, and other mathematical topics such as algebra and functions. Students at this level are expected to have a thorough understanding of *NAEP Basic* level arithmetic operations—an understanding sufficient for problem solving in practical situations. Quantity and spatial relationships in problem solving and reasoning should be familiar to them, and they should be able to convey underlying reasoning skills beyond the level of arithmetic. They should be able to compare and contrast mathematical ideas and generate their own examples. These students should make inferences from data and graphs, apply properties of informal geometry, and accurately use the tools of technology. Students at this level should understand the process of gathering and organizing data and be able to calculate, evaluate, and communicate results within the domain of statistics and probability.
NAEP Basic (262)	**Eighth-grade students performing at the *NAEP Basic* level should exhibit evidence of conceptual and procedural understanding in the five NAEP content areas. This level of performance signifies an understanding of arithmetic operations—including estimation—on whole numbers, decimals, fractions, and percents.** Eighth-graders performing at the *NAEP Basic* level should complete problems correctly with the help of structural prompts such as diagrams, charts, and graphs. They should be able to solve problems in all NAEP content areas through the appropriate selection and use of strategies and technological tools—including calculators, computers, and geometric shapes. Students at this level also should be able to use fundamental algebraic and informal geometric concepts in problem solving. As they approach the *NAEP Proficient* level, students at the *NAEP Basic* level should be able to determine which of the available data are necessary and sufficient for correct solutions and use them in problem solving. However, these eighth-graders show limited skill in communicating mathematically.
Below NAEP Basic	**Not Described**

For a typical state summative assessment, ALDs are found on individual score reports to provide meaning to the scale score the student achieved. Individual results, however, are not reported for the NAEP assessments. There is a rigorous psychometric process being used to create NAEP reporting ALDs where scale scores are extrapolated for individual students, those scale scores are assigned to achievement levels, and NAEP items are linked to the extrapolated scale scores. The group of items assigned to each achievement level represents the knowledge, skills, and abilities (KSAs) an "average student" in each level can demonstrate. Even though it is a misnomer to discuss student performance in terms of NAEP, we use this shorthand when describing the KSAs included in the reporting ALDs or the high-level skills found in policy ALDs.

The Lowest Achievement Level

The Governing Board's 1995 policy and updated 2018 policy to *not* describe the lowest level rests on reasonable psychometric principles. First, NAEP assesses the knowledge and skills of students relative to the NAEP frameworks. These frameworks reflect current educational requirements in the United States. Students in the lowest category may demonstrate knowledge of some entry skills or lower-ability skills, but they are not yet able to demonstrate the bulk of the knowledge and skills measured by the framework. NAEP results are not intended to drive instruction; rather, the NAEP provides a snapshot of student performance in the United States on the subject area tests, and the reporting ALDs provide a snapshot of the framework KSAs found in the NAEP Basic, NAEP Proficient, and NAEP Advanced areas of the scale. To capture the KSAs of the lowest achievement category, the NAEP frameworks may need to be expanded to consider pre-requisite skills the students in the lowest category can demonstrate.

Without expanding the item pool, the conditional standard error of measurement is quite large in the area of the test scale below the NAEP Basic range. There are relatively few items that cover this area of the scale in most NAEP assessments, so it is difficult to identify any KSAs with the same precision as there is in the other performance categories. Currently, the Nation's Report Card website provides sample items and item maps to indicate types of KSAs found in the lowest achievement category. Even with an expanded item pool, it will be difficult to encapsulate the KSAs of the diverse student performance found in the lowest achievement category, below the NAEP Basic level. This area of the scale ranges from the lowest obtainable scale score to the scale score just before the NAEP Basic cut score. Student performance at the lowest obtainable scale score is unknown. Student test performance is assigned to the lowest obtainable scale score by default or because students have performed poorly on the test. We cannot say anything about the KSAs associated with the lowest obtainable scale score; however, much could be summarized regarding the KSAs of the students just below the NAEP Basic cut score.

What do Other Testing Entities Do?

Even though it is difficult, some testing organizations describe student performance in the lowest category. To understand the current state of the field, we gathered information from:

- State websites
- TIMSS and PIRLS
- Academic literature

We examined all 50 states' websites and Washington, DC. Within the state websites, we searched technical reports, score interpretation guides, and standard-setting information for

evidence regarding how each state addressed the lowest achievement level for grade 4 English language arts (ELA) and grade 8 mathematics. This information was not easily located for most states. In some cases, links were broken or outdated. In other cases, information was located for one grade/content area but not another. In Utah, we could only locate the reporting ALDs for grade 6 math. For Alabama, we found the grade 8 math descriptors but not the grade 4 ELA descriptors.

We focused on collecting information related to policy and/or reporting ALDs, as these were the ALDs most often reported to stakeholders. The following states are not included in the list or counts of state ALDs:

- Arkansas, Kentucky, Maryland: Information could not be located
- New Jersey: 2008 descriptors were located and determined to be out of date
- Arizona: an example ALD was located in a guide for score reporting on a graphic of an individual student report but a full list could not be located

For the most part, state departments of education followed the same patterns across their ALDs—if they reported a descriptor for the lowest category in mathematics, then they also reported it in ELA.

States: Reporting and Policy ALDs

Table 3 shows the number of states where policy or reporting ALDs could be located. It shows we located information for 46 states and could not locate information for five states. Table 3 shows that 43 of the 46 states reported something at the lowest category. The states belonging to PARCC (Colorado, New Mexico, and Washington, DC) do not report at the lowest performance category. Of the 43 states reporting at the lowest level, we located policy descriptors for 37 states and reporting descriptors for eight states (Alaska, Maine, Minnesota, Ohio, Pennsylvania, Texas, Utah, and Virginia).

Table 3. Numbers of States Reporting at the Lowest Performance Category

ALD Status	Total Number of States	Number of States Reporting at Lowest AL
Located	46	43
Not Located	5	

Table 4 lists the ALDs for grade 8 mathematics and grade 4 ELA. States that belong to the Smarter Balanced Assessment Consortium are grouped together. The policy ALDs are listed first, followed by the reporting ALDs. The policy descriptors appear to split between the use of negative language or positive language. Slightly more than half of the states with policy descriptors used negative language to describe student performance in this category. In other words, the descriptors stated the student “has not met” the achievement standard. For example, see Illinois and Indiana descriptors in Table 4. Slightly fewer than half use positive language to come to a similar conclusion. These states tend to use the phrase: “demonstrates minimal understanding.” Many states also assert the students in the lowest achievement level will need academic support. For example, see New York or North Dakota descriptors for an example of positive language.

Table 4. Descriptors for Lowest Achievement Level

State	Lowest AL Name	Grade 8 MA	Grade 4 ELA	Link
Policy				
Smarter Balanced: California, Hawaii, Idaho, Michigan, Oregon, Nevada, Montana, South Dakota	Standard Not Met	The student has not met the achievement standard and needs substantial improvement to demonstrate the knowledge and skills in mathematics needed for likely success in future coursework.	The student has not met the achievement standard and needs substantial improvement to demonstrate the knowledge and skills in English language arts/literacy needed for likely success in future coursework.	https://www.cde.ca.gov/ta/tg/ca/sbachievedescript.asp
Smarter Balanced: Connecticut	Does Not Meet the Achievement Standard	The student has not yet met the achievement standard for mathematics expected for this grade. Students performing at this standard require substantial improvement toward mastery of mathematics knowledge and skills. Students performing at this standard will likely need substantial support to get on track for success in high school and college coursework or career training.	The student has not yet met the achievement standard for English language arts and literacy expected for this grade. Students performing at this standard require substantial improvement toward mastery of English language arts and literacy knowledge and skills. Students performing at this standard will likely need substantial support to get on track for success in the next grade.	https://portal.ct.gov/-/media/SDE/Student-Assessment/Smarter-Results-Resources/Interpretive_Guide_8_23_17FINAL.pdf?la=en
Smarter Balanced: Delaware	Minimal Understanding	The Level 1 student demonstrates minimal understanding of and ability to apply the English language arts and literacy (mathematics) knowledge and skills needed for success in college and career, as specified in the Common Core State Standards.	The Level 1 student demonstrates minimal understanding of and ability to apply the English language arts and literacy (mathematics) knowledge and skills needed for success in college and career, as specified in the Common Core State Standards.	https://www.doe.k12.de.us/cms/lib/DE01922744/Centricity/Domain/111/DeSSA%20Executive%20State%20Summary%202017.pdf

Table 4. Descriptors for Lowest Achievement Level (Continued)

State	Lowest AL Name	Grade 8 MA	Grade 4 ELA	Link
Smarter Balanced: Vermont, Washington	Level 1	Student demonstrates minimal understanding of and ability to apply the knowledge and skills associated with college content readiness.	Student demonstrates minimal understanding of and ability to apply the knowledge and skills associated with college content readiness.	https://vt.portal.cambiumast.com/resources/test-blueprints/
Alabama	Level 1	The student has a minimal understanding of grade-level standards and is likely to need additional support at this level of learning as described in the Alabama Course of Study.		https://www.alsde.edu/sec/sa/Pages/relatedinfo-all.aspx?navtext=Resources
Florida	Inadequate	Highly likely to need substantial support for the next grade	Highly likely to need substantial support for the next grade	http://www.fldoe.org/core/fileparse.php/5663/urlt/Grade-LevelFS2021.pdf
Georgia	Beginning Learners	do not yet demonstrate proficiency in the knowledge and skills necessary at this grade level/course of learning, as specified in Georgia's content standards. The students need substantial academic support to be prepared for the next grade level or course and to be on track for college and career readiness.	do not yet demonstrate proficiency in the knowledge and skills necessary at this grade level/course of learning, as specified in Georgia's content standards. The students need substantial academic support to be prepared for the next grade level or course and to be on track for college and career readiness.	https://www.gadoe.org/Curriculum-Instruction-and-Assessment/Assessment/Documents/Milestones/ALD/ALDS_for_Grade_8_Milestones_EOG_Mathematics.pdf
Illinois	Did not yet meet expectations	Students performing at this level do not yet meet academic expectations for the knowledge, skills, and practices contained in the standards for ELA/L or mathematics assessed at their grade level. They will need academic support to engage successfully in further studies in this content area.	Students performing at this level do not yet meet academic expectations for the knowledge, skills, and practices contained in the standards for ELA/L or mathematics assessed at their grade level. They will need academic support to engage successfully in further studies in this content area.	https://www.isbe.net/Documents/New-Meridian-Tech-Rpt-2019.pdf

Table 4. Descriptors for Lowest Achievement Level (Continued)

State	Lowest AL Name	Grade 8 MA	Grade 4 ELA	Link
Indiana	Below Proficiency	Indiana students below proficiency have not met current grade level standards. Students may require significant support to develop the knowledge, application, and analytical skills needed to be on track for college and career readiness.	Indiana students below proficiency have not met current grade level standards. Students may require significant support to develop the knowledge, application, and analytical skills needed to be on track for college and career readiness.	https://www.doe.in.gov/assessment/ilearn
Iowa	Not-Yet Proficient	Students performing at the not-yet-proficient level have not yet demonstrated the knowledge and skills to be classified as Proficient.	Students performing at the not-yet-proficient level have not yet demonstrated the knowledge and skills to be classified as Proficient.	https://itp.education.uiowa.edu/ia/documents/Research-Guide-Form-E-F.pdf
Kansas	Level 1	A student at Level 1 shows a limited ability to understand and use the skills and knowledge needed for post-secondary readiness.	A student at Level 1 shows a limited ability to understand and use the skills and knowledge needed for post-secondary readiness.	https://ksassessments.org/resources-and-training
Louisiana	Unsatisfactory	Students performing at this level have not yet met the college and career readiness expectations and will need extensive support to be prepared for the next level of studies in this content area.	Students performing at this level have not yet met the college and career readiness expectations and will need extensive support to be prepared for the next level of studies in this content area.	https://www.louisianabelieves.com/resources/library/assessment
Massachusetts	Not Meeting Expectations	A student who performed at this level did not meet grade-level expectations in this subject. The school, in consultation with the student's parent/guardian, should determine the coordinated academic assistance and/or additional instruction the student needs to succeed in this subject.	A student who performed at this level did not meet grade-level expectations in this subject. The school, in consultation with the student's parent/guardian, should determine the coordinated academic assistance and/or additional instruction the student needs to succeed in this subject.	https://www.doe.mass.edu/mcas/tech/

Table 4. Descriptors for Lowest Achievement Level (Continued)

State	Lowest AL Name	Grade 8 MA	Grade 4 ELA	Link
Mississippi	Minimal Understanding	Students performing below the Basic level inconsistently demonstrate the knowledge or skills that define basic level performance.	Students performing below the Basic level inconsistently demonstrate the knowledge or skills that define basic level performance.	https://districtaccess.mde.k12.ms.us/studentassessment/Public%20Access/Forms/AllItems.aspx?RootFolder=%2Fstudentassessment%2FPublic%20Access%2FStatewide_Assessment_Programs%2FMAAP-Mississippi%20Academic%20Assessment%20Program%2FMAAP%20Report%20Interpretation%20Guides&FolderCTID=0x0120008C41041A907A304BA89A4587F88962BC&View=%7B5FB78E06-9076-48F4-9A3B-C02F433B41D0%7D
Missouri	Below Basic	Students do not yet demonstrate proficiency in the knowledge and skills necessary at this grade level/course of learning, as specified in content expectations. These students need substantial academic support to be prepared for the next grade level or course and to be on track for college and career readiness.	Students performing at the Below Basic level on the Missouri Assessment Program demonstrate a minimal command of the skills and processes identified in the Missouri Learning Standards. They demonstrate these skills inconsistently and/ or incorrectly in reading processes responding to literary and informational texts and in writing, listening, and speaking forms. Students performing at the Below Basic level use few strategies to comprehend and interpret texts, demonstrate little understanding of literary forms, and apply few strategies for accessing information. They demonstrate little or no ability to organize and/ or develop writing or exhibit little command of the conventions of standard English.	https://dese.mo.gov/sites/default/files/asmt-gl-gir-spring-2019.pdf

Table 4. Descriptors for Lowest Achievement Level (Continued)

State	Lowest AL Name	Grade 8 MA	Grade 4 ELA	Link
Nebraska	Developing	Developing learners do not yet demonstrate proficiency in the knowledge and skills necessary at this grade level, as specified in the assessed Nebraska College and Career Ready Standards.	With a range of texts with text complexity commonly found in Grade 4, a student performing in Developing can likely [Insert text from specific standard here]	https://www.education.ne.gov/assessment/nscas-general-summative-assessment/nscas-mathematics/
New Hampshire	Level 1 Below Proficient	The Level 1 student is below proficient in applying mathematics knowledge/skills as specified in the standards. The student generally performs significantly below the standard for the grade level/course, is likely able to partially access grade-level content, and engages with higher order thinking skills with extensive support.	The Level 1 student is below proficient in applying English language arts knowledge/skills as specified in the standards. The student generally performs significantly below the standard for the grade level/course, is likely able to partially access grade-level content, and engages with higher order thinking skills with extensive support.	https://www.education.nh.gov/sites/g/files/ehbemt326/files/inline-documents/sonh/nhsas-v1-tech-report-2018-19.pdf
New York	NYS Level 1	Students performing at this level are well below proficient in standards for their grade. They demonstrate limited knowledge, skills, and practices embodied by the New York State P-12 Common Core Learning Standards for Mathematics that are considered insufficient for the expectations at this grade.	Students performing at this level are well below proficient in standards for their grade. They demonstrate limited knowledge, skills, and practices embodied by the New York State P-12 Common Core Learning Standards for English Language Arts/Literacy that are considered insufficient for the expectations at this grade.	https://www.engageny.org/resource/performance-level-descriptions-for-ela-and-mathematics

Table 4. Descriptors for Lowest Achievement Level (Continued)

State	Lowest AL Name	Grade 8 MA	Grade 4 ELA	Link
North Carolina	Not Proficient	Students who are Not Proficient demonstrate inconsistent understanding of grade level content standards and will need support.	Students performing at this level have limited command of the knowledge and skills contained in the Common Core State Standards (CCSS) Reading Standards for Literature as assessed by referring to the text when drawing inferences, as well as when explaining what the text directly says; summarizing the text and determining the theme from details; using specific details to describe a character, setting, or event in a story; and determining the meaning of words and phrases as they are used in a text, including those words referring to mythological characters. Students will need academic support to engage successfully in this content area.	https://www.dpi.nc.gov/documents?field_document_type_tid=388&field_document_type_tid_op=or
North Dakota	Novice	The Level 1 student is below proficient in applying mathematics knowledge/skills as specified in the standards. The student generally performs significantly below the standard for the grade level/course, is likely able to partially access grade-level content, and engages with higher order thinking skills with extensive support	The Level 1 student is below proficient in applying English language arts knowledge/skills as specified in the standards. The student generally performs significantly below the standard for the grade level/course, is likely able to partially access grade-level content, and engages with higher order thinking skills with extensive support	https://www.nd.gov/dpi/districtsschools/assessment/ndsa
Oklahoma	Below Basic	Students have not performed at least at the Basic level.	Students have not performed at least at the Basic level.	https://sde.ok.gov/assessment-material

Table 4. Descriptors for Lowest Achievement Level (Continued)

State	Lowest AL Name	Grade 8 MA	Grade 4 ELA	Link
Pennsylvania	Below Basic	The Below Basic Level reflects inadequate academic performance, and work at this level demonstrates a minimal command of and ability to apply the knowledge, skills, and practices represented in the Pennsylvania standards. Consistent performance at this level indicates extensive additional academic support may be needed for engaging successfully in further studies in this content area.	The Below Basic Level reflects inadequate academic performance, and work at this level demonstrates a minimal command of and ability to apply the knowledge, skills, and practices represented in the Pennsylvania standards. Consistent performance at this level indicates extensive additional academic support may be needed for engaging successfully in further studies in this content area.	https://www.education.pa.gov/K-12/Assessment%20and%20Accountability/PSSA/Pages/DescriptorsCutScores.aspx
Rhode Island	Not Meeting Expectations	A student who performed at this level did not meet grade-level expectations in this subject. The school, in consultation with the student's parent/guardian, should determine the coordinated academic assistance and/or additional instruction the student needs to succeed in this subject.	A student who performed at this level did not meet grade-level expectations in this subject. The school, in consultation with the student's parent/guardian, should determine the coordinated academic assistance and/or additional instruction the student needs to succeed in this subject.	https://www.ride.ri.gov/InstructionAssessment/Assessment/RICASAssessments.aspx
South Carolina	Does Not Meet Expectations	A student who does not meet expectations in the knowledge and skills necessary at this grade level of learning, as defined by the grade-level content standards, needs substantial academic support to be prepared for the next grade level and to be on track for college and career readiness.	A student who does not meet expectations in the knowledge and skills necessary at this grade level of learning, as defined by the grade-level content standards, needs substantial academic support to be prepared for the next grade level and to be on track for college and career readiness.	https://ed.sc.gov/tests/middle/sc-ready/

Table 4. Descriptors for Lowest Achievement Level (Continued)

State	Lowest AL Name	Grade 8 MA	Grade 4 ELA	Link
Tennessee	Below	Performance at this level demonstrates that the student has a minimal understanding and has a partial ability to apply [Insert course here] knowledge and skills as defined by the Tennessee Academic Standards.	Performance at this level demonstrates that the student has a minimal understanding and has a partial ability to apply [Insert course here] knowledge and skills as defined by the Tennessee Academic Standards.	https://www.tn.gov/education/assessment/tnready.html
Texas	Did Not Meet Grade Level	Performance in this category indicates that students are unlikely to succeed in the next grade or course without significant, ongoing academic intervention. Students in this category do not demonstrate a sufficient understanding of the assessed knowledge and skills.	Performance in this category indicates that students are unlikely to succeed in the next grade or course without significant, ongoing academic intervention. Students in this category do not demonstrate a sufficient understanding of the assessed knowledge and skills.	https://tea.texas.gov/student-assessment/testing/staar/staar-performance-level-descriptors
West Virginia	Does Not Meet Standards	The student generally demonstrates a minimal understanding of, and ability to, apply grade-level math knowledge, skills, and abilities relative to the West Virginia College- and Career-Readiness Standards for Mathematics.	The student generally demonstrates a minimal understanding of, and ability to, apply grade-level English language arts (ELA) knowledge, skills, and abilities relative to the West Virginia College- and Career-Readiness Standards for ELA.	https://wvde.us/assessment/scaled-score-information/wvgsa-in-grades-3-8/
Wisconsin	Below Basic	Student demonstrates minimal understanding of and ability to apply the knowledge and skills for their grade level that are associated with college content-readiness.	Student demonstrates minimal understanding of and ability to apply the knowledge and skills for their grade level that are associated with college content-readiness.	https://dpi.wi.gov/sites/default/files/imce/assessment/pdf/WI_Math_PerformanceLevelDescriptors.pdf
Wyoming	Below Basic	Students performing at below basic level in Mathematics have minimal or no academic performance indicating understanding and little display of the knowledge and skills included in the Wyoming Content and Performance Standards.	Students performing at the below basic level in English Language Arts have minimal academic performance indicating minimal understanding and little display of the knowledge and skills included in the Wyoming Content and Performance Standards.	https://edu.wyoming.gov/educators/state-assessment/plds/

Table 4. Descriptors for Lowest Achievement Level (Continued)

State	Lowest AL Name	Grade 8 MA	Grade 4 ELA	Link
Reporting				
Alaska	Far Below Proficient	Students who score at this level may be able to recognize that irrational numbers are different from rational numbers, understand exponents as repeated multiplication, find the slope of a line using a graph, represent whole numbers in scientific notation, identify whether a relation is a function, recognize congruence and similarity, recognize single transformations of geometric figures, find the hypotenuse in a right triangle with sides whose lengths are whole numbers that are Pythagorean triples, and recognize associations in data that represent two quantities.	Students who score at this level attempt to read and minimally comprehend grade 4 text to identify main ideas and explicit details, determine meanings of basic words and phrases while identifying literal and figurative language, identify text features and structures used to organize a text, and identify relationships between parts of a text. When writing or revising, students attempt to use appropriate language and conventions, use strategies particular to a type of text, and structure a text to support a purpose or opinion.	https://education.alaska.gov/tls/Assessments/Peaks/EducatorGuide_Assessments_Reports.pdf
Maine	Well Below State Expectations	The student's work demonstrates a minimal understanding of, and ability to apply the mathematics knowledge and skills needed for achievement relative to the grade level Math Content and Practice Standards. The student solves some problems that require applying simple strategies to basic areas of mathematics without an understanding of the reasoning behind the strategies.	The student's work demonstrates a minimal understanding of the knowledge and skills needed to meet Maine's ELA/Literacy Content Standards with texts of appropriate complexity for the grade level.	https://www.maine.gov/doe/sites/maine.gov.doe/files/inline-files/MEA_2018_ALDs%2BCut%20Scores_10-09-18.pdf

Table 4. Descriptors for Lowest Achievement Level (Continued)

State	Lowest AL Name	Grade 8 MA	Grade 4 ELA	Link
Minnesota	Does Not Meet the Standards	A student at this level of mathematics succeeds at few of the most fundamental mathematics skills of the Minnesota Academic Standards. Some of the skills demonstrated may include: **Number & Operation** • Recognizes fractions and terminating decimals as rational numbers **Algebra** • Recognizes linear functions in graphic presentations • Translates linear representations from a table to a graph • Identifies slope by counting whole number units on a graph Identifies patterns in a table of a linear function (e.g., recognizes patterns for x or y-values but not the relationship between x and y) • Substitutes "easy" numbers and evaluates simple expressions **Geometry & Measurement** • Recognizes the equation for the Pythagorean Theorem • Recognizes parallel or perpendicular lines on a graph **Data Analysis** • Generalizes the properties of the line of best fit of a graphed data set • Displays data using scatterplots	When interacting with literature and informational text, students at this achievement level demonstrate the following skills inconsistently and with minimal accuracy. **Key Ideas and Details (Standards 1, 2, 3)** • Recall details from text • Make simple predictions based on explicit text • Identify a cause or an effect • Identify obvious fact and opinion in explicit text • Make general comparisons based on explicit text • Locate explicit main idea and central message • Identify basic sequence of events **Craft and Structure (Standards 4, 5, 6)** • Recognize simple figures of speech • Locate obvious context clues to understand word meanings • Identify key words and phrases • Recognize the features, format, and function of basic text structures (e.g., listing) and their impact on meaning • State author's obvious purpose in explicit text **Integration of Knowledge and Ideas (Standard 8, Informational Text sub-strand only)** • Identify obvious evidence in text (e.g., logical connections between sentences and paragraphs)	https://education.mn.gov/MDE/dse/test/ald/

Table 4. Descriptors for Lowest Achievement Level (Continued)

State	Lowest AL Name	Grade 8 MA	Grade 4 ELA	Link
Ohio	Limited	A student performing at the Limited Level demonstrates a minimal command of Ohio's Learning Standards for Grade 8 Mathematics. A student at this level has an emerging ability to formulate and reason about expressions and equations, use functions to describe quantitative relationships, and analyze two- and three-dimensional space and figures using distance, angle, similarity, and congruence, and to understand and apply the Pythagorean Theorem.	A student performing at the Limited Level demonstrates a minimal command of Ohio's Learning Standards for Grade 4 English Language Arts. A student at this level has an emerging ability to determine the main idea or theme of a text and explain how it is supported by key details while providing a summary of the text, explain how an author uses evidence to support particular points in a text, and write an opinion or explanatory piece that introduces a topic, develops the topic with facts, details, and supported opinions, and links ideas with categories of information.	https://oh.portal.cambiumast.com/resources/reporting-resources/
Pennsylvania	Emerging/Below Basic	Students performing at this level identify rational numbers and locate approximate positions on a number line. They evaluate radical notation for perfect squares. Students solve one- and selected two-step equations in one variable with one solution. Students identify a function using a graph or table. They use the Pythagorean theorem or volume formulas to solve simple or routine problems. Students identify line of best fit or determine patterns of association in bivariate data.	A student performing at the below basic level demonstrates inadequate understanding of literary and informational texts. The student demonstrates minimal or no understanding of vocabulary, word meaning, and conventions of language. The student demonstrates minimal or no understanding of writing skills.	https://www.education.pa.gov/K-12/Assessment%20and%20Accountability/PSSA/Pages/DescriptorsCutScores.aspx

Table 4. Descriptors for Lowest Achievement Level (Continued)

State	Lowest AL Name	Grade 8 MA	Grade 4 ELA	Link
Texas	Did Not Meet Grade Level Performance	Students achieving Did Not Meet Grade Level Performance can • Solve problems using direct variation • Solve application problems involving the Pythagorean theorem • Use proportional and non-proportional relationships to develop foundational concepts of functions • Approximate the value of irrational numbers	When reading texts of increasing complexity, students achieving Did Not Meet Grade Level Performance can • Determine the meaning of unfamiliar words using explicit context • Demonstrate a literal understanding of literary and informational texts • Make plausible inferences about literary and informational texts	https://tea.texas.gov/student-assessment/testing/staar/staar-performance-level-descriptors
Utah *Grade 6 ma	Below Proficient	Understands and represents situations with rates and ratios. Understands and uses negative numbers. Uses equations and expressions to solve problems, including equations and expressions that contain variables in place of numbers. Works with statistical data to find a measure of center, including mean, median, and mode.		https://www.schools.utah.gov/assessment/assessments?mid=1173&tid=7

Table 4. Descriptors for Lowest Achievement Level (Continued)

State	Lowest AL Name	Grade 8 MA	Grade 4 ELA	Link
Virginia	Fail/Below Basic	A student performing at this level should be able to *Reporting Category 1: Number, Number Sense, Computation, and Estimation* • compare fractions and decimals • identify natural numbers, whole numbers, and integers • use manipulatives to recognize perfect squares • solve practical problems involving fractions, decimals, and integers *Reporting Category 2: Measurement and Geometry* • define and recognize acute, obtuse, right, and straight angles • determine the areas of circles, triangles, and rectangles • determine the volume/surface area of a rectangular prism given a labeled figure • identify the image of a polygon resulting from a single transformation • use manipulatives to describe the views (top/front/side) of a three-dimensional figure • define the Pythagorean Theorem *Reporting Category 3: Probability, Statistics, Patterns, Functions, and Algebra* • determine the probability of a simple event • name the dependent and independent variables represented in a scatterplot • apply the order of operations to numerical expressions • simplify algebraic expressions	A student performing at this level should be able to • Define context clues. • Recognize that some words have prefixes and/or suffixes. • Define synonym or antonym. • Recall the purpose of word-reference materials. • Recall literary terms. • Recall plot events. • Explain the characteristics of fiction and nonfiction. • Define sensory words. • Locate information in texts to answer literal questions. • Define cause or effect. • Locate text features. • Recognize that an author has a purpose for writing. • Define main idea or topic. • Define fact or opinion.	https://www.doe.virginia.gov/testing/scoring/performance_level_descriptors/index.shtml

Each state uses a different strategy when writing reporting ALDs. Alaska uses “may be able” when describing the types of knowledge and skills students in the lowest achievement level possess, while Ohio discusses “emerging” knowledge and skills. Texas describes what students in the lowest level “can do.” Pennsylvania uses action verbs to describe math skills for the students in the lowest level, while using “minimal or no understanding” to discuss reading skills.

The use of the phrases “may be able” or “can” is important when writing ALDs. The term “can” implies these are KSAs most students in the lowest category would be able to demonstrate. The term “may be able” does not carry the same meaning. “May be able” implies some students might have the KSAs while others do not.

States: Range ALDs

Range ALDs are detailed definitions of student performance that encompass the entirety of student performance in a particular achievement level. These are sometimes written at the beginning of the test development cycle and describe the types of KSAs students should be able to do. In other instances, these ALDs are based on items and describe the KSAs students can do. This ALD type was not the primary focus of our work; however, we encountered this ALD type as we searched for policy and reporting ALDs. These ALDs are developed for each content strand within a content standard. We have pulled examples from New York, Alaska, and Alabama for standards related to expressions and equations.

New York uses the following text to describe what a Level 1 student *should* do in relationship to cluster 8.EE.1,3,4 (Students work with integers): “Write simple numerical expressions involving whole number exponents and evaluate expressions with exponents of between 1 and 10.”

Alaska describes “what a typical student scoring at each level can do” (ADEED, 2017). For Alaska Standards 8.EE.1-8.EE.8, the student in the lowest category (Far Below Proficient) is summarized as:

- A student at this level understands exponents as representing repeated multiplication.
- A student at this level finds the slope of a line using a graph.
- A student at this level represents whole-number multiples of ten in scientific notation.

Alabama also describes “what a typical student in each performance level can do... A student would not necessarily demonstrate all the skills listed a particular performance level.” For Alabama Standards 8.EE.3-8.EE.10, a student in the lowest performance category is summarized as:

- estimates very large or very small quantities as a single digit times an integer power of 10,
- interprets the unit rate as the slope of a proportional relationship represented in a graph, and
- finds the slope of a line using a graph.

TIMSS and PISA

Achievement level descriptors are also used for the TIMSS and PISA assessments. Like NAEP, these assessments only produce scores at the group level and not for individual students. Unlike NAEP, neither assessment system uses a judgmental standard setting; instead, using a “pragmatic and empirically-based approach” (Olson & Nilson, 2017) normative cut scores are established. Both assessments use item maps to create item-level descriptors that are aggregated into ALDs. PISA splits the lowest level into two performance categories: 1a and 1b. The summary descriptors for the lowest performance levels for scientific literacy read:

> At **Level 1a**, students are able to use every day content and procedural knowledge to recognise or identify explanations of simple scientific phenomenon. With support, they can undertake structured scientific enquiries with no more than two variables. They are able to identify simple causal or correlational relationships and interpret graphical and visual data that require a low level of cognitive demand. Level 1a students can select the best scientific explanation for given data in familiar personal, local and global contexts.
>
> At **Level 1b**, students can use every day content knowledge to recognise aspects of simple scientific phenomenon. They are able to identify simple patterns in data, recognise basic scientific terms and follow explicit instructions to carry out a scientific procedure.

PISA uses "can" statements to describe the knowledge and skills demonstrated by students in the lowest proficiency level.

TIMSS describes performance at four benchmarks, including the lowest benchmark. For the 8th grade mathematics "Low Benchmark," the descriptor reads, "Have some knowledge of whole numbers and basic graphs." Consequently, TIMSS is describing, at a very high level, the skills possessed by students classified as "Low".

Other Research

Outside of psychometrics, other researchers have explored which students are in the lowest performance level. For example, Valencia and Buly (2004) examined 108 grade 5 students who scored in the lowest performance level of a grade 4 statewide reading assessment in a northwestern United States school district. The students were administered assessments that targeted expert-identified aspects of reading, including word identification, meaning (comprehension and vocabulary), and fluency (rate and expression). All students were measured individually on a battery of tests. Word identification was measured using the 1989 Woodcock-Johnson Psycho-Educational Battery-Revised and the 1995 Qualitative Reading Inventory-II (QRI-II). Comprehension was measured with the QRI-II and the 1981 Peabody Picture Vocabulary Test-Revised. Fluency was evaluated by measuring the reading rate of all passages from the QRI-II and state test. They also scored reading expression using the NAEP Oral Study rubric. All data were analyzed through a cluster analysis to look for similar patterns on word identification, meaning, and fluency.

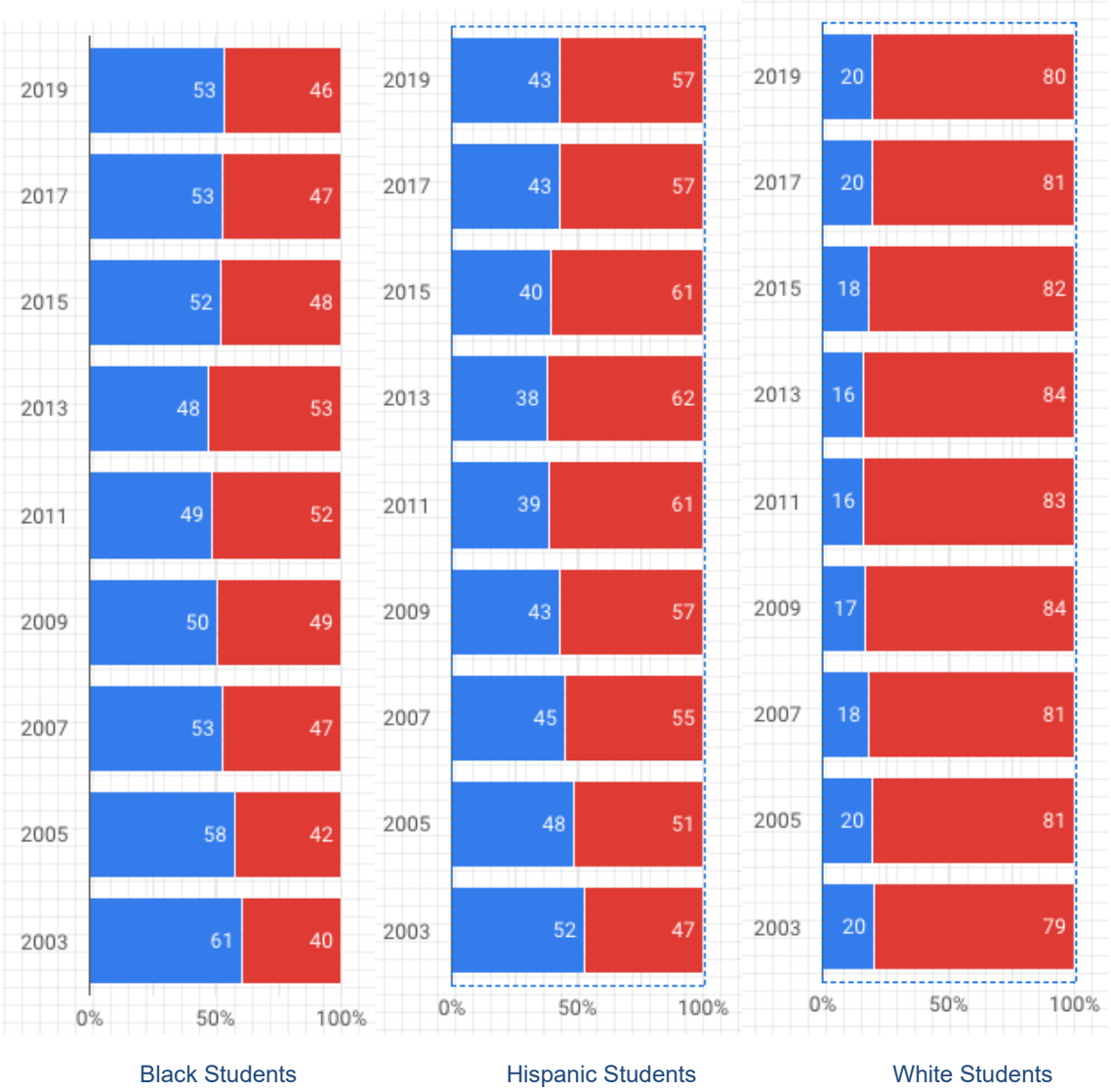
The researchers created six profiles of prototypical students who fall in the lowest achievement level. These profiles include: Automatic Word Callers, Struggling Word Callers, Word Stumblers, Slow Comprehenders, Slow Word Callers, and Disabled Readers. For each prototypical student, they identified level of word identification, comprehension, vocabulary, expression, reading rate, and writing.

Demographic Composition of the Lowest Achievement Level

The main purpose of this paper is to explore how the lowest achievement level is described (or not described) by entities outside of the Governing Board. Even so, it is important to understand which student groups are classified in the lowest category of achievement on NAEP, for it is their performance that goes undescribed. Here, we examine 2019 NAEP Grade 8 Mathematics data for patterns of performance. Similar patterns are seen in other grade/content area combinations.

In 2019, the assessment performance of approximately 31% of all students was classified as below the NAEP Basic achievement level. When disaggregated by racial group, the 2019

assessment performance of 53% of Black students[1], 43% of Hispanic students, 49% of American Indian/Alaska Native, 14% of Asian/Pacific Islander students, and 20% of white students falls into the lowest category. When disaggregated by students participating in the free and reduced lunch (FRL) program (an indicator often used as a proxy for socioeconomic status), the assessment performance of 46% of FRL students falls in the lowest category while the performance of 18% of non-FRL students and 20% of unknown FRL status falls in the lowest category. In short, the test performance of students of color or students receiving FRL is more likely to fall into the lowest level than is the test performance of white students or non-FRL students. The graphics in Figure 1 show the relative stability of the percentages of Black students, Hispanic students, and white students in each category on the 2019 NAEP Grade 8 Mathematics assessment.



Note. *Data were extracted from NAEP Data Explorer

Figure 1. Percentage of Students in Below the NAEP Basic achievement level compared to NAEP Basic and Above for Black, Hispanic, and White Students, NAEP Grade 8 Mathematics

[1] The achievement gap between black and white students on NAEP is well-known (see USDOE, 2015).

The students comprising the lowest achievement level are disproportionately students of color or students living in poverty. Performance on NAEP, like any large-scale summative assessment, reflects societal inequities. Students of color and/or students in poverty often do not have access to the same educational resources as white, Asian, and/or wealthy students. Students' performance on NAEP may reflect systemic racism that exists in U.S. society.

Discussion

Like all well-designed assessments, the NAEP measures the content frameworks that underlie it. The NAEP is not intended to measure students' performance outside of the NAEP frameworks, and the Governing Board cannot report what is *not* measured. If the Governing Board decides to report the KSAs of the students in the lowest performance category, it may be necessary to add pre-requisite KSAs to the content frameworks to capture what students in this level can demonstrate on an assessment. This may culminate in a special test form specifically designed to capture the KSAs of the students in the lowest category. Or, NAEP could perform special studies to understand and report more about the knowledge and skills of students below the NAEP Basic level. If the Governing Board decides to create a descriptor for the lowest performance category, then a policy descriptor will also be needed for that level.

If the Governing Board decides to create descriptors, the range of performance in the lowest category will need to be addressed. This category covers a range from students who are unable to demonstrate skills to students who are almost Basic. There is not a clear best method for creating Below Basic ALDs. These are sometimes written to refer to what half of the students in the lowest level can do. In other cases, they are written to refer to the prerequisite skills students should have to enter the Basic category. If an ALD is created for this area, language that describes for whom the ALDs refer will need to be carefully crafted.

The reversal of current policy to create an ALD for the lowest achievement category would provide information on the KSAs of the nearly 50% of Black students and approximately 40% of Hispanic students represented in the lowest achievement level. This may become even more critical as the United States enters post-pandemic life. If U.S. students have experienced learning loss during the pandemic, then we can expect the percentages of students in the lowest category to increase.

Almost all states provide some sort of description of the lowest achievement category. In many cases, this is just a policy descriptor; however, a handful of states, as well as TIMSS and PISA, provide a reporting descriptor that lists skills students in the lowest category either can or should be able to do. States are creating the more detailed range ALDs that also list skills students in the lowest performance category either can or should be able to do. In short, there is precedence in the K–12 assessment space to create descriptors for the lowest performance category should the Governing Board decide to provide additional information about the knowledge and skills of students below the NAEP Basic level.

References

ACT (2017). Performance Level Descriptors. https://actinc.my.salesforce.com/sfc/p/#300000000Wu5/a/1B000000QhZj/KQPOSfgRjBbvCP.9biXbH_v8btPvm8QgCjd5zj.7AqE

Alaska Department of Education and Early Development (2017). Achievement Level Descriptors (ALDS) Mathematics Grades 6-8. https://docs.google.com/document/d/1GPbaGD7xqAt9H05eTB1oND3zdExEz5aI1GtTRcPBHS0/edit#

Olsen, R.V., & Nilsen, T. (2017). Standard setting in PISA and TIMSS and How these procedures can be used nationally. In S. Blomeke & J.-E. Gustafsson (Eds), *Standard Setting in Education* (pp. 69-84). Springer International Publishing.

PISA (2015). Proficiency scale construction. https://www.oecd.org/pisa/sitedocument/PISA-2015-Technical-Report-Chapter-15-Proficiency-Scale-Construction.pdf

TIMSS (2015). http://timssandpirls.bc.edu/timss2015/international-results/wp-content/uploads/filebase/full%20pdfs/T15-International-Results-in-Mathematics.pdf

Valencia, S., & Buly, M. (2004). Behind Test Scores: What Struggling Readers Really Need. *The Reading Teacher*, 57(6), 520-531. Retrieved February 26, 2021, from http://www.jstor.org/stable/20205395

The Below *NAEP Basic* Panel

December 2020

2019 Grade 4 Reading

≈ 1,245K or **34%**

Performed below the *NAEP Basic* level in 2019

2019 ↑ higher than that of 2017, 2015, or 2013

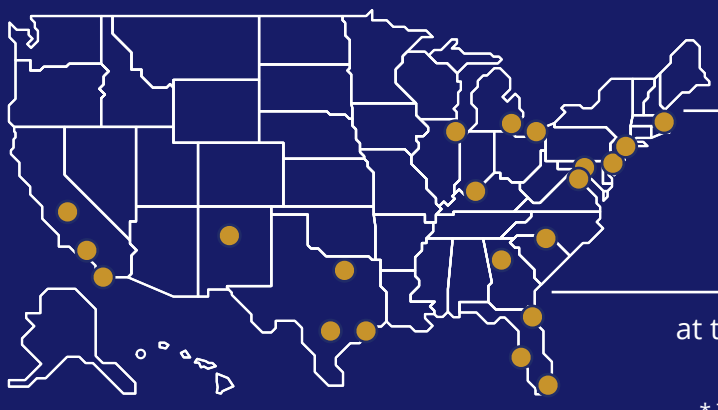
2019 Grade 12 Mathematics

≈ 1,454K or **40%**

Performed below the *NAEP Basic* level in 2019

2019 ↑ higher than that of 2015, 2013, or 2009

Highest percentage of below *NAEP Basic* students in **2019 Grade 8** Mathematics



45%

at the state level

73%

at the TUDA* district level

* Trial Urban District Assessment

IES Institute of Education Sciences

NAEP NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

National Center for Education Statistics, Assessment Division

An Expert Panel Meeting on Performance Below the *NAEP Basic* Achievement Level:

A highlight of the meeting summary

May 2021

Background

As the responsible party for reporting National Assessment of Educational Progress (NAEP) results, the Assessment Division of the National Center for Education Statistics (NCES) has become increasingly concerned that it is overlooking students who perform below the *NAEP Basic* achievement level. The current NAEP reporting approach renders these students largely invisible and impedes efforts to provide evidence-based, targeted support to those most in need. To better inform policies to serve this population of students, NCES convened an expert panel on December 16–17, 2020, and charged its members with addressing questions in the three following areas:

- **Data:** What do we need to know about students who receive scores below *NAEP Basic*—who are they, what do they know, and what can they do in the various NAEP subject areas?
- **Measurement:** Are changes in NAEP procedures (e.g., design, administration, and analysis) needed to improve measurement and get a more accurate description of students who receive scores below *NAEP Basic*?
- **Reporting:** How can NAEP better describe the performance of students who score below *NAEP Basic* when reporting results?

Summary of Meeting

The panel was composed of nine experts, selected because of their strong background either in educational measurement, research, policy, or expertise in curriculum and teaching in mathematics or reading (see **pages 4–6** of this document for a brief bio of the panel members and the moderator). NCES hosted the expert panel meeting with support from the American Institutes for Research (AIR). The meeting, which consisted of two half-day sessions, opened with remarks from NCES Associate Commissioner Dr. Peggy Carr, followed by presentations of data on students at grades 4, 8, and 12 who performed below the *NAEP Basic* achievement level. Dr. George Bohrnstedt, AIR, moderated the panel discussions. In her opening remarks, Dr. Carr stated:

> ...high percentages of our students are performing below NAEP's basic level. We do not know enough about these students. Knowing more requires improving NAEP's processes for gathering data about them and describing them more completely. The American public, especially stakeholders in the education enterprise, should be alarmed about this growing group of underperforming students. NAEP needs to play a leadership role in better identifying who they are and what their educational needs are.

In addition, Dr. Carr differentiated the roles of NCES and the National Assessment Governing Board in NAEP assessments and procedures. She asked the panel "to be expansive and creative in thinking about the issues before us today."

NCES staff member Taslima Rahman, who organized the meeting, and two team members from AIR, Sakiko Ikoma and Markus Broer, presented data organized as three observations to give a comprehensive picture of those students who do not score, at a minimum, at the *NAEP Basic* achievement level. The first presentation, or Observation I, briefly described how NAEP results are reported to introduce the concept of "below *NAEP Basic*" and then showed results from the national, state, and district levels on students performing below *NAEP Basic.* The second presentation, or Observation II, focused on subscale and item-level data, and the third presentation, or Observation III,

focused on auxiliary data; that is, process data on what students did during the assessment and data from other NCES assessment surveys that could be linked to NAEP. Below are the takeaways from these three presentations.

- Results presented in Observation I showed that in 2019, the percentage of students scoring below *NAEP Basic* ranged from 20 to 40 percent across the three NAEP grades (4, 8, and 12) at the national level in reading and mathematics; the percentages were alarmingly high in some states (i.e., over 45 percent) and districts (i.e., over 60 percent) and they varied among student groups. Although the 2019 percentages were lower than the percentages seen in 1990, the percentage of students performing below *NAEP Basic* did not change much since 2003. This presentation also showed that the average scores of grade 8 and 12 students at the lower percentiles (i.e., 10th and 25th) were within the below *NAEP Basic* score range in both subjects and declined in recent years.
- The data shown in Observation II indicated that the percentages of students who received full credit for multiple-choice and constructed-response items were much lower for below *NAEP Basic* students than for students at any other achievement level; nevertheless, many below *NAEP Basic* students could answer some of these items correctly. The data in this presentation also showed that the score differences between below *NAEP Basic* and at or above *NAEP Basic* students across subcontent areas of mathematics and reading were similar when considered in standard deviation units.
- The results shown in Observation III indicated that although the not reached (NR) item percentage increased through the last few items for all students, students performing below *NAEP Basic* had the highest NR rate. This presentation also showed that more students among those performing below *NAEP Basic* compared to others had higher instances of not responding or responding incorrectly to practice prompts in the tutorial section.

Major Recommendations

Both the three guiding questions and the three presentations described above fostered dialogue throughout the two-day meeting. The generation of the recommendations came from overall panel discussion as well as a group process in which the panel was divided into two subgroups where each developed a set of recommendations that were then reviewed by the entire panel. The recommendations based on the panel discussion are as follows.

A Special Note:

It should be noted that, with respect to any specific recommendation, panel members did not vote individually on each recommendation; no consensus or priority ratings were sought from the panel. Therefore, it should not be implied that every panel member agreed with every recommendation and suggestion described in this document.

First, the panel recommended the development of achievement-level descriptors for students who perform in the score range below the *NAEP Basic* cut point by outlining what students at this level know and can do. (The panel did note that achievement-level descriptions and cut points are set by NAGB.) The panel believes that the NAEP framework needs to carefully describe the construct of measurement and skill progressions required across all of the achievement levels, including what is now described as below *NAEP Basic*. This recommendation also underscores the need to name the level that is below

Questions related to this content should be directed to Taslima Rahman at Taslima.Rahman@ed.gov.

the *NAEP Basic* achievement level. Given the large range of scores below *NAEP Basic*, the panel also suggested giving consideration to including multiple levels below *NAEP Basic*, as is done in other large-scale assessment programs, such as the Program for International Student Assessment (PISA). Naming the below *NAEP Basic* score range and providing descriptions of what students who perform at this level know and can do would enrich the reporting of NAEP.

Second, the panel recommended that the distribution of items included in NAEP assessments correspond to the distribution of student ability, especially at the lower range. The current distribution of NAEP item difficulty is right-skewed and, therefore, lower performing students may become discouraged by what they see as inaccessible items. The panel suggested adding more items measuring the lower part of the NAEP scale so that the distribution of item difficulty more closely mirrors the entire distribution of student performance. The items of more appropriate difficulty will allow more precise measures of what students performing below *NAEP Basic* know and can do and add more insight into the performance of these students.

Third, the panel recommended that the NAEP reporting emphasis on students who perform below the *NAEP Basic* achievement level should, at a minimum, match the reporting emphasis for the three current achievement levels (*NAEP Basic*, *NAEP Proficient,* and *NAEP Advanced*). In addition, the panel suggested that further contextual information about students who perform below *NAEP Basic* be collected from teachers and schools so that policymakers, researchers, and the general public have a more robust set of variables from which to gain an understanding of these students' educational performance.

Other Recommendations

Other recommendations were offered during this two-day meeting to improve the NAEP program more generally. For example, the panel recommended that membership should overlap across the different committees that build the NAEP framework, develop the test specifications, write the items, and review the items both before and after reporting results. The panel believes such an overlap would promote better coordination among the various steps inherent in designing and administering the NAEP assessments, which in turn would ensure that the assessment clearly reflect the requirements laid out for the performance of students at all levels of achievement. This is an approach applied in some other large-scale assessments, such as PISA and the Program for the International Assessment of Adult Competencies (PIAAC). Other recommendations were to:

a) consider adaptive testing to assess more precisely the performance of students at various score ranges on the NAEP scale,

b) collect more information on instruction provided to students performing below *NAEP Basic* compared with students performing at or above *NAEP Basic*,

c) take into consideration that the lack of reading skills and general background knowledge may hinder demonstrating the ability that is actually measured in certain items, especially for students performing below *NAEP Basic*, and

d) conduct studies using process data to better understand the differences between those students performing below *NAEP Basic* and those performing at or above *NAEP Basic*, including how they approach items of varying type and difficulty and how motivated they are in taking the test.

Questions related to this content should be directed to Taslima Rahman at Taslima.Rahman@ed.gov.

Moderator

George Bohrnstedt is a senior vice president and Institute Fellow at the American Institutes for Research (AIR). His earliest work at AIR was as Project Director of the congressionally mandated evaluation of Trial State NAEP done in collaboration with the National Academy of Education. He also chaired the National Center for Education Statistics' NAEP Validity Studies Panel (NVS) for 23 years before stepping down in October 2018. While chairing NVS, he authored or co-authored numerous studies undertaken by NVS, and provided oversight for the panel's research agenda and projects.

Panel members

Henry Braun is the Boisi Professor of Education and Public Policy and Director of the Center for the Study of Testing, Evaluation and Education Policy in the Lynch School of Education & Human Development at Boston College. After serving as an assistant professor of statistics at Princeton University, he joined the Educational Testing Service in 1979, where he served as vice-president for research management from 1990 to 1999. He held the title of distinguished presidential appointee from 1999 until his retirement in 2006, when he moved to Boston College. A fellow of the American Statistical Association and AERA, he is an elected member of the National Academy of Education.

Ray Hart has more than 30 years of experience in research and evaluation. His work has spanned policy areas such as post-secondary success and college readiness, school improvement, teacher effectiveness, early childhood education, and adult and workforce literacy. He has worked with clients from the U.S. Department of Education, the U.S. Department of Housing and Urban Development, the U.S. Department of State, the National Science Foundation, the National Academies of Sciences, Engineering, and Medicine, and many state and local departments of education. Dr. Hart currently serves as the Director of Research for the Council of the Great City Schools.

Panel members

Hanseul Kang began her role as assistant dean and the first Anita and Joshua Bekenstein '80 B.A. Executive Director of The Broad Center at Yale School of Management in November 2020. Most recently, Kang served as the State Superintendent of Education for the District of Columbia from 2015 to 2020, providing strategic vision, clear direction, and steady leadership to the District's state education agency, the Office of the State Superintendent of Education. Under Kang's leadership, the District of Columbia continued to make major strides in student achievement outcomes on NAEP, on its annual state assessments in math and English language arts, and on other key measures.

Irwin Kirsch is the Ralph Tyler Chair in Large-Scale Assessment and Director of the Center for Global Assessment at ETS in Princeton, NJ. Over the course of his career, Dr. Kirsch has worked in close collaboration with a number of state, national and international organizations, including the World Bank, UNESCO, the International Association for the Evaluation of Educational Achievement (IEA), and the Organization for Economic Co-operation and Development (OECD), where he currently oversees the development and conduct of the two largest international assessments that provide policy-makers and key stakeholders with national and international comparative data on literacy and workforce preparedness: PIAAC and PISA.

Michele Mailhot joined the Maine Department of Education in July 2009 as the mathematics specialist. In this role, she works with a wide audience of stakeholders in education across the state providing professional learning to the field supporting research-based effective instructional practices in mathematics education. Before accepting the position of mathematics specialist, Michele taught a combination of middle and high school mathematics for 11 years. In addition to her work in the classroom and at the state level, she also has worked with NASA developing curriculum for teachers across the nation with a focus on the integration of mathematics, science, and literacy.

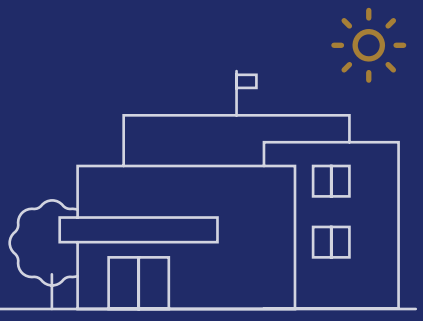
Pamela A. Mason is a senior lecturer on education and the director of the Language and Literacy Master's program and the Jeanne Chall Reading Lab at the Harvard Graduate School of Education. Her professional and research interests encompass the role of culturally sustaining pedagogy in promoting literacy achievement, the interaction of text complexity and background knowledge, and the intersection of literacy learning, culture, and multilingualism. She collaborates with colleagues nationally and globally on preparing reading specialist teachers and literacy coaches, developing the capacity of school leaders as literacy advocates, and evaluating schoolwide literacy programs.

Panel members

Gary W. Phillips is Vice President and Chief Psychometrician at Cambium Assessment, Inc. Dr. Phillips provides senior-level guidance on all statistical and psychometric techniques and procedures used in assessment activities. This involves item response theory calibrating and equating, classical statistical analysis, differential item functions, standard setting, reliability and validity studies, generalizability studies, item banking, distributional projections, and value-added studies involving hierarchical linear models and structural equation models. Dr. Phillips has published or presented over 200 papers, taught dozens of advanced graduate level statistics courses, and presented hundreds of workshops on advanced statistical and psychometric topics.

Jennifer Randall is an associate professor in the Research, Educational Measurement, and Psychometrics program, Director of Evaluation in the Center for Educational Assessment, and Associate Dean of Academic Affairs in the College of Education at the University of Massachusetts. Currently, she is particularly interested in the differential negative impact of both large and small-scale assessments on historically marginalized populations in the U.S. and abroad; and the ways in which a culturally sustaining, anti-racist approach to instruction and assessment can mitigate these negative outcomes.

Lorrie A. Shepard is University Distinguished Professor in the School of Education at the University of Colorado Boulder. Dr. Shepard is past president of the American Educational Research Association and past president of the National Council on Measurement in Education. She was elected to the National Academy of Education in 1992 and served as president of the NAEd from 2005-2009. She has served on the NAEP Validity Studies Panel from 1995 to the present. Dr. Shepard has received distinguished career awards recognizing her contributions in measurement, research, and teacher education respectively from NCME and ETS, AERA, and AACTE.



IES Institute of Education Sciences

NAEP
NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Highlights of the 2018 NAEP Oral Reading Fluency Study

NCES 2021-026
U.S. DEPARTMENT OF EDUCATION

IES Institute of Education Sciences

NAEP NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Highlights of the 2018 NAEP Oral Reading Fluency Study

APRIL 2021

Sheida White
American Institutes for Research

John Sabatini
University of Memphis

Bitnara Jasmine Park
American Institutes for Research

Jing Chen
National Center for Education Statistics

Jared Bernstein
Analytic Measures, Inc.

Mengyi Li
American Institutes for Research

NCES 2021-026
U.S. DEPARTMENT OF EDUCATION

U.S. Department of Education
Miguel Cardona
Secretary

Institute of Education Sciences
Mark Schneider
Director

National Center for Education Statistics
James L. Woodworth
Commissioner

Assessment Division
Peggy G. Carr
Associate Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high-priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high-quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public. Unless specifically noted, all information contained herein is in the public domain.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

NCES, IES, U.S. Department of Education
Potomac Center Plaza (PCP)
550 12th Street SW
Washington, DC 20202

April 2021

The NCES Home Page address is https://nces.ed.gov.
The NCES Publications and Products address is https://nces.ed.gov/pubsearch.

This publication is only available online. To download, view, and print the report as a PDF file, go to the NCES Publications and Products address shown above.

This report was prepared for the National Center for Education Statistics under Contract No. ED-IES-12-D-0002 with American Institutes for Research. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. government.

Suggested Citation
White, S., Sabatini, J., Park, B. J., Chen, J., Bernstein, J., and Li, M. (2021). *Highlights of the 2018 NAEP Oral Reading Fluency Study* (NCES 2021-026). U.S. Department of Education. Washington, DC: Institute of Education Sciences, National Center for Education Statistics. Retrieved [date] from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2021026.

Content Contact
Jing Chen
(202) 245-8324
jing.chen@ed.gov

Contents

Page

List of Tables iv

List of Figures iv

Introduction 1

Purpose of the ORF Study 1

Importance of Measuring Oral Reading Fluency 1

Foundational Skills for Fluency 2

The Role of Language Comprehension in Reading Comprehension and Oral Reading 2

Method 3

Measures of Oral Reading Fluency and Foundational Skills 3

Operationalization of the Measures 3

Tasks 4

Scoring 4

Findings 5

New Data on the Reading Skills of Fourth-Graders Performing Below *NAEP Basic* 5

Characteristics of Students in the Below *NAEP Basic* Subgroups 5

Oral Reading Fluency and Foundational Skills for the Below *NAEP Basic* Subgroups 7

1: Oral reading fluency (passage reading) 7

ORF passage reading WCPM 7

ORF passage reading accuracy 8

ORF passage reading expression 10

Summary of findings on ORF passage reading and NAEP reading performance 10

2: Foundational skills (word and pseudoword list reading) 12

Word reading WCPM 12

Pseudoword reading WCPM 13

Conclusion and Implications 14

Conclusion 14

Implications 14

The NAEP reading framework and future assessments 14

Policy and research 15

References 16

List of Tables

Table **Page**

1. Percentage of fourth-graders performing below *NAEP Basic,* by below *NAEP Basic* subgroup and selected student characteristics: 2018 6

List of Figures

Figure **Page**

1. Average passage reading WCPM, by NAEP reading achievement level and below *NAEP Basic* subgroup: 2018 8
2. Average passage reading accuracy, by NAEP reading achievement level and below *NAEP Basic* subgroup: 2018 9
3. Average passage reading expression, by NAEP reading achievement level and below *NAEP Basic* subgroup: 2018 11
4. Average word reading WCPM, by NAEP reading achievement level and below *NAEP Basic* subgroup: 2018 12
5. Average pseudoword reading WCPM, by NAEP reading achievement level and below *NAEP Basic* subgroup: 2018 13

Introduction

This publication highlights the key concepts and findings of the 2018 National Assessment of Educational Progress (NAEP) Oral Reading Fluency (ORF) study. For additional details, see its companion publication, *The 2018 NAEP Oral Reading Fluency Study* (White et al. 2021), which is available on the NAEP ORF website at https://nces.ed.gov/nationsreportcard/studies/orf/.

Oral reading fluency is defined as the ability to read text aloud with speed, accuracy, and proper expression. The 2018 NAEP ORF study was the first such NAEP study since 2002. It was administered to a nationally representative sample of over 1,800 fourth-graders from 180 public schools.

The students in the ORF study first completed the NAEP fourth-grade reading assessment. Then they moved on to the tasks that were administered for this study: reading out loud four short passages to assess oral reading fluency and two word lists to assess skills that provide the foundation for fluency. Both the NAEP reading tasks and the study tasks were administered to students on tablets, and students' responses were recorded on the tablets.

Purpose of the ORF Study

The purpose of the 2018 NAEP ORF study is to add new, policy-relevant information to the NAEP reading assessment. It includes, for the first time, a close examination of the oral reading fluency and foundational skills of fourth-grade public school students who perform below *NAEP Basic* on the NAEP reading assessment. "Foundational skills," word reading and phonological decoding, are defined under the heading, *Foundational Skills for Fluency.*

Importance of Measuring Oral Reading Fluency

Students who read aloud with appropriate speed, accuracy, and expression (i.e., students who have oral reading fluency) are more likely to comprehend connected text (Sabatini, Wang, and O'Reilly 2019) because they are able to conserve cognitive resources that can be applied to the comprehension of meaning (Perfetti 2007). Thus, oral reading fluency is a reliable and easily accessible indicator of overall reading competence—and a strong marker of progress in learning to read (Fuchs et al. 2001)—and its assessment has become one of the primary means of determining which elementary school students are on track toward meeting state reading standards and which students would benefit from additional services and intervention (McGlinchey and Hixson 2004; Reschly et al. 2009).

Foundational Skills for Fluency

The term "foundational skills" refers to two skills that are assessed in this study with two different word lists: (1) *word reading* (also known as word recognition)—the ability to read familiar words with accuracy and speed—and (2) *phonological decoding*—the ability to pronounce unfamiliar words based on knowledge of spelling-sound correspondences.[1] Research has established that fast and accurate word reading is a major driver of oral reading fluency (e.g., Eason et al. 2013; Metsala and David 2017; Silverman et al. 2013). In addition, phonological decoding is regarded by almost all reading researchers as a critical prerequisite for the development of skilled, fluent reading and reading comprehension. In essence, as children apply phonological decoding skills to the unfamiliar words that they encounter in text, they make a transition from being "novices" to being "experts" who read familiar words rapidly and automatically (Castles, Rastle, and Nation 2018; Share 1995).

The Role of Language Comprehension in Reading Comprehension and Oral Reading

Although the ORF study focuses on oral reading fluency, word reading, and phonological decoding, they are not the only factors that may affect performance on the fourth-grade NAEP reading assessment. One of the most important factors is language comprehension. "Language comprehension" is the ability to understand language based on knowledge of the meaning of words, sentence structure, and other aspects of language.[2] An extensive body of research and theory supports the view that language comprehension is necessary for reading comprehension (see, e.g., Foorman, Petscher, and Herrera 2018; Hoover and Gough 1990).

It is important to recognize that oral reading also involves language comprehension, just as silent reading and reading comprehension do. First, when students read a passage out loud with appropriate expression, they are using their ability to comprehend language as well as read the words in the passage. Second, when students read a passage out loud, they use their knowledge of word meaning and sentence structure to anticipate and recognize (read) the words in the text. This process is called "contextual facilitation of word recognition."[3] Therefore, oral passage reading (fluency) tasks are measuring language comprehension in addition to fast and accurate word reading. This implicit measurement of language comprehension is one of the reasons why oral reading fluency assessments are valued by educators and widely used in elementary schools (Reschly et al. 2009).

---

[1] Many researchers consider phonological awareness to be another critically important foundational skill. It was not measured in this study because it is rapidly and fully acquired by normally developing readers in preschool, kindergarten, and first grade.

[2] Language comprehension is measured by tests that require no reading, such as orally administered vocabulary tests and listening comprehension tests.

[3] Contextual facilitation has been extensively studied by researchers. Evidence comes from (1) experimental studies showing, for example, that coherent passages are read more rapidly than text containing the same words in random order; and (2) correlational studies showing that, for example, vocabulary and listening comprehension affect oral reading fluency when word reading skills are controlled statistically.

Method

Measures of Oral Reading Fluency and Foundational Skills

The 2018 NAEP ORF study included measures of oral reading fluency, word reading, and phonological decoding. The last two are regarded as foundational skills for fluency.

- *Oral reading fluency* (passage reading) refers to the ability to read connected text such as paragraphs and passages with appropriate rate, accuracy, and expression, which is an indicator of comprehension.
- *Word reading* (also known as *word recognition*) refers to the ability to recognize familiar written words with appropriate speed and accuracy, relying primarily on orthographic memory (memory of how the words are pronounced).
- *Phonological decoding* refers to the ability to pronounce unfamiliar words based on knowledge of spelling-sound correspondences.

As noted previously, many words that students initially pronounce by "sounding them out" eventually become automatically recognized as chunks of letters or whole words in a process that requires minimal conscious effort. This is why it is important to measure both the ability to phonologically decode unfamiliar words and the ability to recognize familiar words.

Operationalization of the Measures

Each of the above measures was operationalized in terms of two aspects of performance—rate and accuracy—as well as a combination of the two, words correct per minute.

- *Words correct per minute (WCPM)* refers to the total number of words correctly read divided by the amount of time taken to read the passages or word-level lists. This is the WCPM score.
- *Accuracy* refers to the percentage of words that was read accurately. For passages, the total number of attempted words[4] in the passage was the denominator, and for word lists, the total number of words presented to students was the denominator.

[4] Attempted words included words read correctly or incorrectly as well as those that were skipped.

Passage reading was operationalized in terms of one additional measure—expression—defined below:

- *Expression* refers to appropriate intonation, rhythm, emphasis, and pausing that groups words into phrasal and larger units in ways that express the meaning and structure of the text and enhance understanding and enjoyment in a listener.

Tasks

The following text materials were given to students to be read aloud:

- *Text passages,* consisting of 152–162 words, providing a measure of fourth-graders' ability to read words and sentences in connected text.
- *Word lists,* consisting of 24 English words arranged in increasing order of complexity, providing a measure of individual students' ability to recognize familiar words.
- *Pseudoword lists,* consisting of 18 made-up but pronounceable words (e.g., *jad*), providing a measure of students' ability to decode words they are unfamiliar with.

The word and pseudoword lists used in this study were developed based on principles derived from clinically valid measures of children's acquisition of word recognition and phonological decoding. Moreover, these word-level tasks along with the text passages were tested in cognitive laboratory studies administered by NAEP ORF team researchers to ensure they were within typical fourth-graders' ability to perform.

Scoring

In this study, NCES used a new automatic speech analysis/scoring system that calculated accuracy, rate, and WCPM variables to score recordings of students' reading. In preparation for scoring the tasks administered for this study, extensive work was done to ensure that correct word pronunciation would be scored reliably and that speakers of nonstandard varieties of English would not be unfairly penalized. The scoring system considered nonstandard pronunciations acceptable as long as they were consistent with the participants' general speaking pattern.

Scoring of the *Expression* variable, which is based on a detailed rubric,[5] involved thorough training of human scorers and multiple levels of quality checks. To ensure reliability of scoring, supervisors spot-checked scores and provided feedback to scorers. In addition, a second scorer rescored 25 percent of all passage reading recordings to monitor interrater reliability (i.e., agreement between scorers on the scores assigned).

[5] The scoring rubric for the Expression variable can be found in the companion publication, *The 2018 NAEP Oral Reading Fluency Study* (White et al. 2021) on the NAEP ORF website, https://nces.ed.gov/nationsreportcard/studies/orf/.

Findings

The findings of the 2018 NAEP ORF study are uniquely useful for exploring the question of how NAEP reading performance is related to oral reading fluency, word reading, and phonological decoding skills. As with all NAEP findings, it is important to remember that cause-and-effect relationships cannot be inferred from descriptive and correlational results. NAEP reading performance, oral reading fluency, word reading, and phonological decoding may be affected by a complex mixture of factors beyond the scope of the study.

New Data on the Reading Skills of Fourth-Graders Performing Below *NAEP Basic*

A major objective of the 2018 ORF study was to provide a nuanced picture of the reading performance of low-performing fourth-grade readers. To accomplish this, students performing below *NAEP Basic* were evenly divided into three groups based on the NAEP reading score distribution. The three groups were labeled below *NAEP Basic* **Low** (i.e., the bottom one-third of the students performing below *NAEP Basic*), below *NAEP Basic* **Medium** (i.e., the middle one-third of the students performing below *NAEP Basic*), and below *NAEP Basic* **High** (i.e., the top one-third of the students performing below *NAEP Basic*). Students' characteristics and oral reading performance were then compared across these subgroups.

Characteristics of Students in the Below *NAEP Basic* Subgroups

Overall, 36 percent of fourth-grade public school students performed below *NAEP Basic,* but 51 percent of Black fourth-grade students and 46 percent of Hispanic fourth-grade students performed below *NAEP Basic.*[6] We found that Black students were also overrepresented in the lowest below *NAEP Basic* subgroup—i.e., below *NAEP Basic* **Low**. As shown in table 1, while 26 percent of the White students performing below *NAEP Basic* were at the lowest level of below *NAEP Basic,* 40 percent of the Black fourth-graders and 37 percent of the Hispanic fourth-graders who performed below *NAEP Basic* fell into this subgroup. Because 51 percent of Black students were in the below *NAEP Basic* group, this finding means that 20 percent of Black fourth-grade students (or one out of every five Black fourth-graders) performed at the lowest end of below *NAEP Basic* (51 percent × 40 percent = 20 percent). Similarly, 17 percent (or one out of six) of Hispanic fourth-graders were in the lowest below *NAEP Basic* group, below *NAEP Basic* **Low** (46 percent × 37 percent = 17 percent).

[6] Here we have reported the observed percentages for the ORF study sample, which are very close to the percentages for the operational NAEP sample. For Black students in the operational NAEP sample, the percentages of students performing below *NAEP Basic* were 50 and 53 for 2017 and 2019, respectively. For Hispanic students in the operational NAEP sample, the percentage of students performing below *NAEP Basic* was 46 in both 2017 and 2019.

Overall, 36 percent of fourth-grade public school students performed below *NAEP Basic,* but 50 percent of National School Lunch Program (NSLP)-eligible fourth-grade students performed below *NAEP Basic.* As shown in table 1, among students who performed below *NAEP Basic,* NSLP-eligible students were nearly equally divided among the three below *NAEP Basic* subgroups. About 35 percent of the NSLP-eligible students performed at the lowest below *NAEP Basic* level.

Table 1. Percentage of fourth-graders performing below *NAEP Basic*, by below *NAEP Basic* subgroup and selected student characteristics: 2018

Student characteristics	below *NAEP Basic* Low	below *NAEP Basic* Medium	below *NAEP Basic* High	Total
All students	**33**	**33**	**33**	**100**
Race/ethnicity				
White	26	35	39	**100**
Black	40	31	28	**100**
Hispanic	37	33	30	**100**
NSLP eligibility				
Eligible	35	34	31	**100**
Not eligible	27	32	41	**100**

NOTE: Rows may not sum to totals because of rounding. For National School Lunch Program (NSLP) eligibility, about 2 percent of the students lacked valid eligibility information. These students were also excluded because of small sample size. Learn more about the NAEP achievement levels here.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2018 Oral Reading Fluency study.

Oral Reading Fluency and Foundational Skills for the Below *NAEP Basic* Subgroups

As mentioned earlier, what this study adds to the previous studies of NAEP reading is a closer examination of the difficulties faced by fourth-grade students performing below *NAEP Basic* on the NAEP reading assessment. Because the 2018 NAEP ORF study participants had completed the NAEP reading assessment, it was possible to examine the relationship between reading achievement and each of the measures.

1: Oral reading fluency (passage reading)

ORF passage reading WCPM

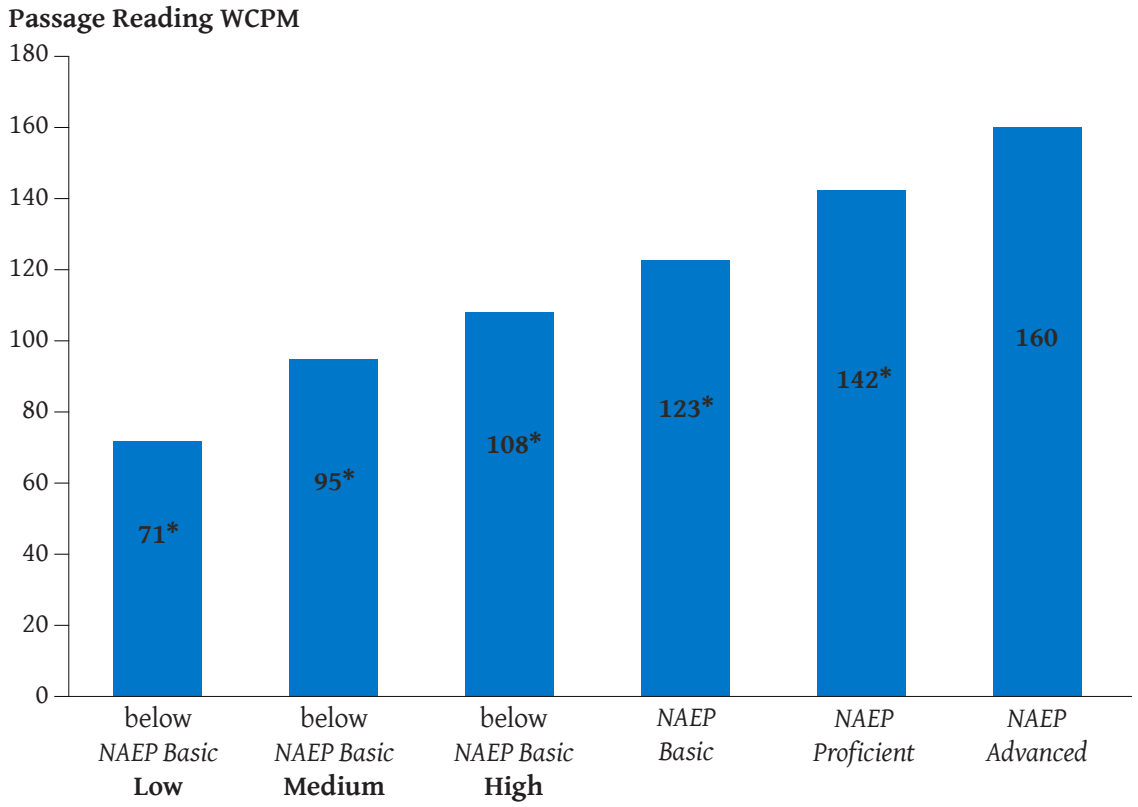
As shown in figure 1, passage reading words correct per minute (WCPM) decreased significantly in moving down from the *NAEP Advanced* group to the *NAEP Proficient* group and *NAEP Basic* group.[7] Also, and importantly, passage reading WCPM decreased significantly within the below *NAEP Basic* group. In moving down the subgroups, the average for students in the below *NAEP Basic* **High** subgroup was 108 WCPM, the average for students in the below *NAEP Basic* **Medium** subgroup was 95 WCPM, and the average for students in the below *NAEP Basic* **Low** subgroup was 71.

It is noteworthy that the passage reading WCPM difference between the lowest below *NAEP Basic* subgroup (below *NAEP Basic* **Low**) and the highest below *NAEP Basic* subgroup (below *NAEP Basic* **High**) is as large at 38 WCPM[8] as the difference between the *NAEP Basic* and *NAEP Advanced* groups (37 WCPM).

The average passage reading WCPM across all levels was 120. To help put all of these numbers in perspective, based on the 2003 National Assessment of Adult Literacy (NAAL) (Baer et al. 2009), adult readers performing at the Intermediate and Proficient levels read orally at an average of 166 and 178 words correctly per minute (WCPM), respectively. This indicates that there is room for improvement even for fourth-grade students performing at the *NAEP Proficient* level (142 WCPM) and considerable room for improvement for fourth-grade students performing at the *NAEP Basic* level (123 WCPM).

[7] All comparisons were conducted with an alpha level of 0.05, with multiple pairwise comparison adjustments applied when needed using the False Discovery Rate (FDR) procedure.
[8] Unrounded numbers were used for calculating the differences between the estimates.

Figure 1. Average passage reading WCPM, by NAEP reading achievement level and below *NAEP Basic* subgroup: 2018



* Statistically significant difference compared to the next higher NAEP reading achievement level category, $p < .05$. All comparisons were conducted with an alpha level of 0.05, with multiple pairwise comparison adjustments applied using the False Discovery Rate (FDR) procedure.
NOTE: WCPM is an abbreviation for words correct per minute. The positions of the data points in the graphics are based on the unrounded numbers.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2018 Oral Reading Fluency study.

ORF passage reading accuracy

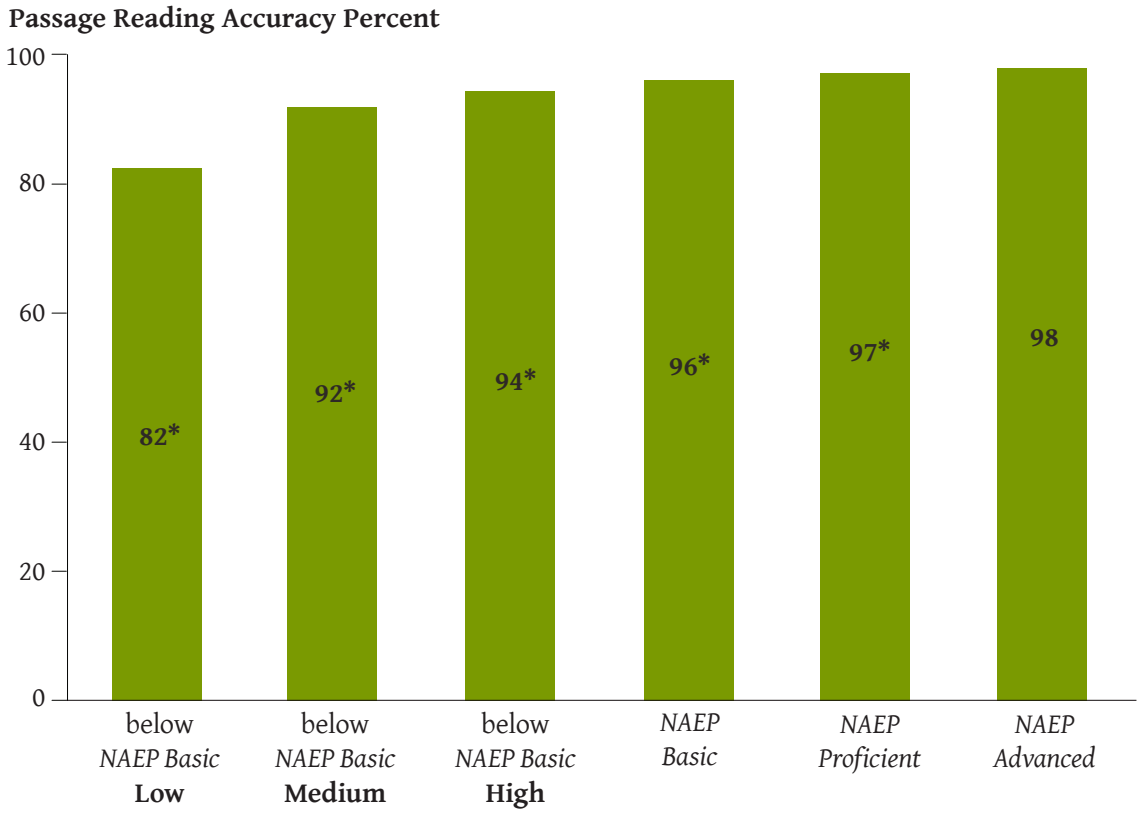
Perhaps the most noticeable difference among students performing below *NAEP Basic* is in passage reading accuracy. As shown in figure 2, the passage reading accuracy of students performing below *NAEP Basic* **Low** was 82 percent, about 9 percentage points[9] and 12 percentage points lower than the below *NAEP Basic* **Medium** and **High** subgroups, respectively.

Eighty-two percent accuracy in practical terms means that students misread 1 out of every 6 words. Students who frequently misread words are likely to have difficulty understanding the text because the words are apt to be content words that are important for comprehension, not

[9] Unrounded numbers were used for calculating the differences between the estimates.

function words[10] (e.g., the, and, on). Also, at 92 percent correct, the below *NAEP Basic* **Medium** group was missing 1 out of every 11 words, which is 1 word in nearly every sentence. The average percentage of words read correctly across all levels was 94 percent.

Figure 2. Average passage reading accuracy, by NAEP reading achievement level and below *NAEP Basic* subgroup: 2018



* Statistically significant difference compared to the next higher NAEP reading achievement level category, $p < .05$. All comparisons were conducted with an alpha level of 0.05, with multiple pairwise comparison adjustments applied using the False Discovery Rate (FDR) procedure.
NOTE: Accuracy refers to the percentage of words that was read accurately. The positions of the data points in the graphics are based on the unrounded numbers.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2018 Oral Reading Fluency study.

[10] To hear an audio recording that illustrates the kind of words that were misread or read with difficulty, see https://nces.ed.gov/nationsreportcard/studies/orf/illustrative_audio.aspx.

ORF passage reading expression

Passage reading expression scores (figure 3) showed the same pattern as the passage reading WCPM. Performance declined steadily from *NAEP Advanced* to *NAEP Proficient* to *NAEP Basic* and continued to decline from below *NAEP Basic* **High** to below *NAEP Basic* **Medium** and below *NAEP Basic* **Low**. The average passage reading expression score for all fourth-grade students was at Level 4 on a scale of 0–5. That indicated that their oral reading expressed sentence structure and meaning, and that more than three-quarters of the words in the passage were read with appropriate expression.

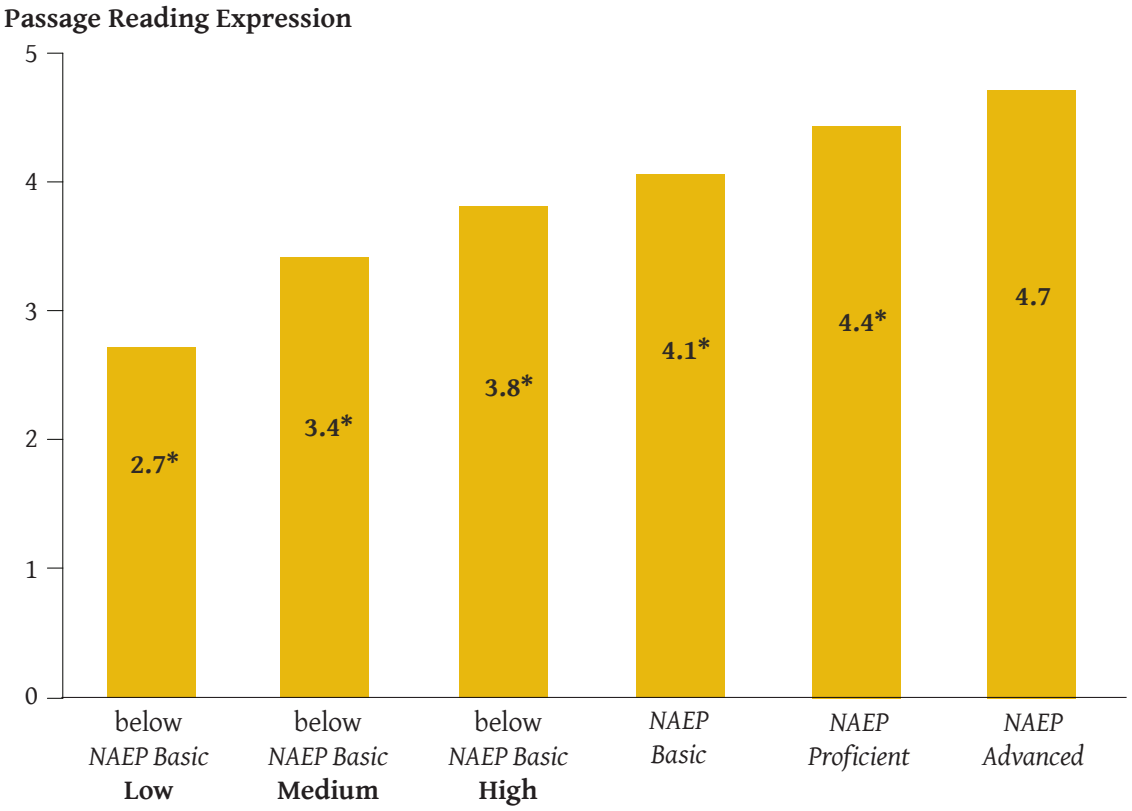
For all readers performing below *NAEP Basic,* the average score was in the Level 3 range. That meant that their oral reading expressed the meaning of words, phrases, clauses, and a few sentences, and that they read more than half of the words in the passage with appropriate expression. For the lowest below *NAEP Basic* subgroup, below *NAEP Basic* **Low**, the average expression score fell below Level 3. That indicated that these students tended to focus on local word groupings, which means that they often paused in the middle of a phrase. For example, the sentence "Hawaii is a warm place, but parts of it are cold" would be read as *[Hawaii] [is a] [warm place], [but parts of] [it are] [cold].*[11]

Summary of findings on ORF passage reading and NAEP reading performance

Overall, across all of the passage reading data described above (WCPM, accuracy, and expression), there is a strong and consistent relationship between the NAEP reading assessment performance and passage reading. The above figures also show that there is noticeable variation among the below *NAEP Basic* subgroups for every passage reading measure.

[11] Passage reading expression by a student in the below *NAEP Basic* **Low** subgroup is illustrated in an audio recording that can be found here: https://nces.ed.gov/nationsreportcard/studies/orf/illustrative_audio.aspx.

Figure 3. Average passage reading expression, by NAEP reading achievement level and below *NAEP Basic* subgroup: 2018



* Statistically significant difference compared to the next higher NAEP reading achievement level category, $p < .05$. All comparisons were conducted with an alpha level of 0.05, with multiple pairwise comparison adjustments applied using the False Discovery Rate (FDR) procedure.

NOTE: The positions of the data points in the graphics are based on the unrounded numbers. Expression score 0 = Insufficient passage reading sample for accurate rating; 1 = Reading is word by word; less than a quarter of the words are read with appropriate expression; 2 = Reading focuses on local grouping; less than half of the words are read with appropriate expression; 3 = Reading expresses the meaning of words, phrases, clauses, and a few sentences; more than half of the words are read with appropriate expression; 4 = Reading expresses sentence structure and meaning; more than three-quarters of the words are read with appropriate expression; 5 = Passage is read as if for a listener and is expressive throughout. For detailed passage reading expression score description, see *The 2018 NAEP Oral Reading Fluency Study* (White et al. 2021) on the NAEP ORF website, https://nces.ed.gov/nationsreportcard/studies/orf/.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2018 Oral Reading Fluency study.

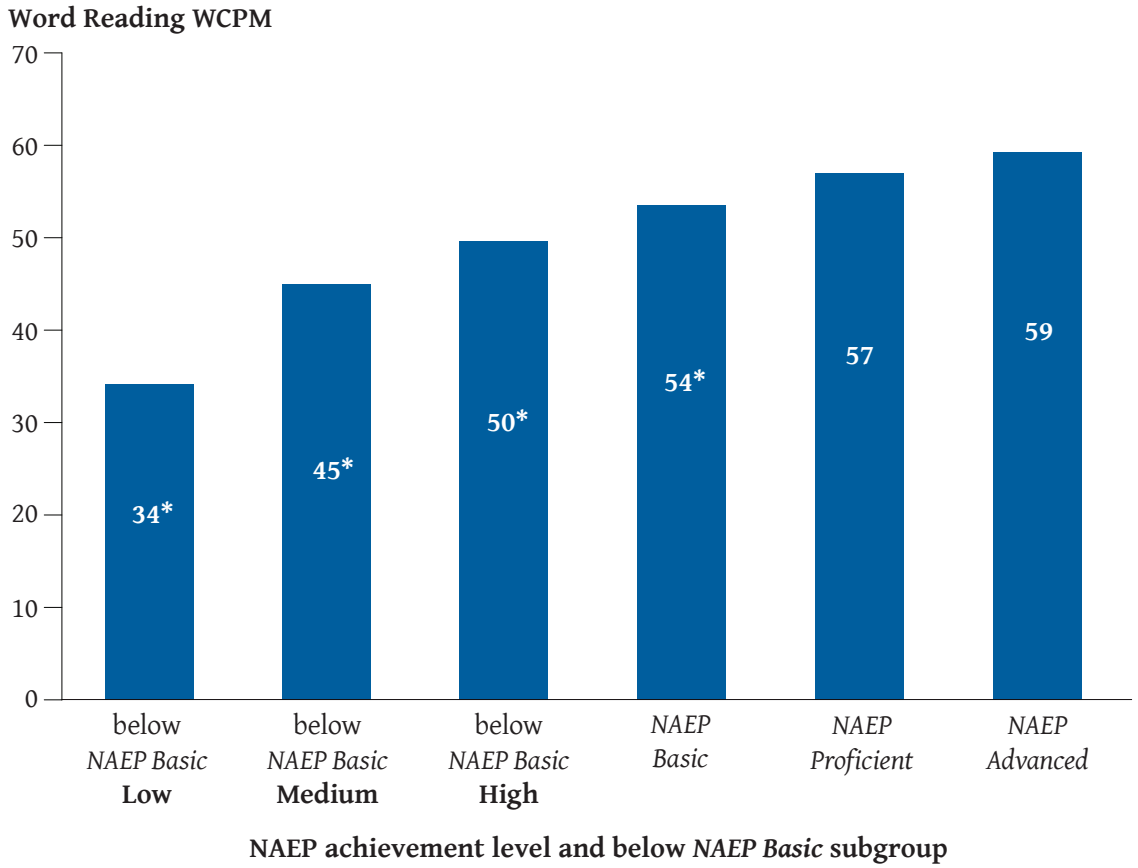
2: Foundational skills (word and pseudoword list reading)

Word reading WCPM

By examining word-level reading apart from a passage, we could better understand the word-level processes that underlie fluency and passage comprehension. What we learned is that the foundational skills—word reading and phonological decoding—also varied widely within the below *NAEP Basic* subgroups (figures 4 and 5).

In word list reading, students read high-frequency words that have known meanings to most students in fourth grade. Performance on this task was regarded an indicator of accumulating knowledge of printed words and an increasing ability to read words rapidly and automatically without effortful decoding. As shown in figure 4, word reading declined across the *NAEP Basic* level through all the below *NAEP Basic* subgroups. The sharpest decline was between the below *NAEP Basic* **Medium** and below *NAEP Basic* **Low** subgroups.

Figure 4. Average word reading WCPM, by NAEP reading achievement level and below *NAEP Basic* subgroup: 2018



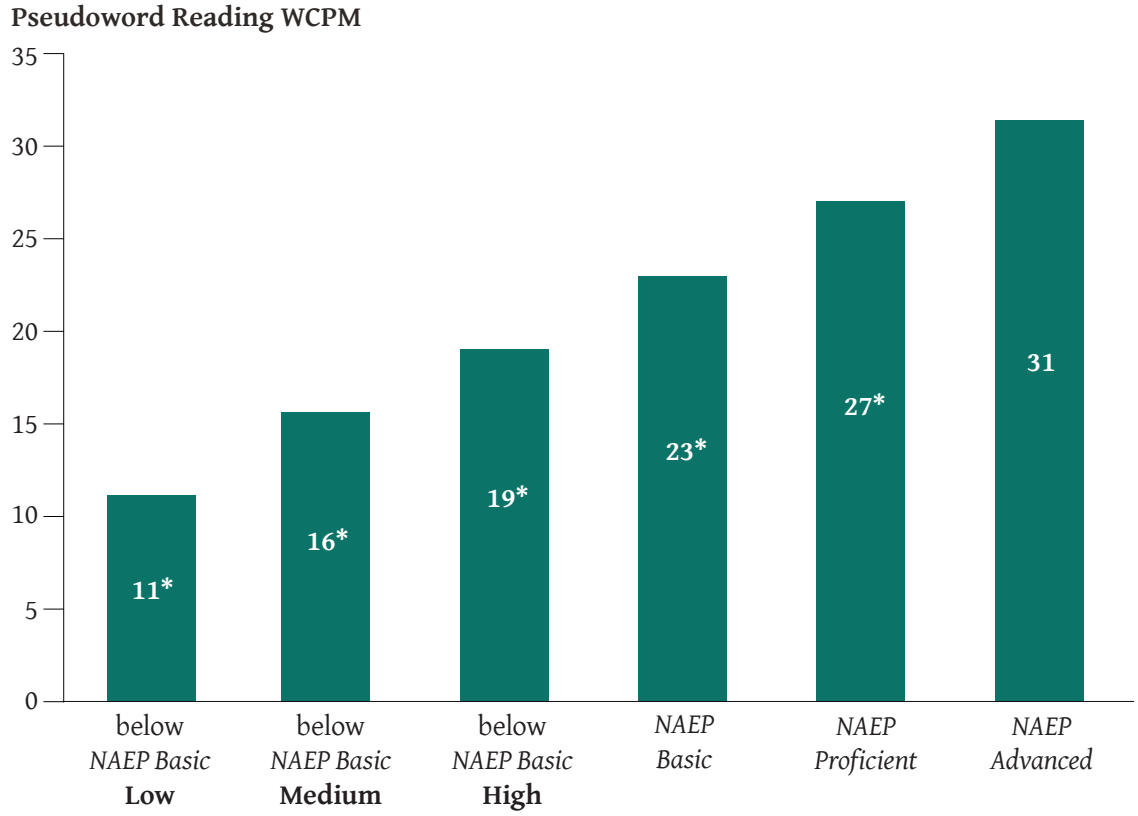
* Statistically significant difference compared to the next higher NAEP reading achievement level category, $p < .05$. All comparisons were conducted with an alpha level of 0.05, with multiple pairwise comparison adjustments applied using the False Discovery Rate (FDR) procedure.
NOTE: WCPM is an abbreviation for words correct per minute. The positions of the data points in the graphics are based on the unrounded numbers.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2018 Oral Reading Fluency study.

Pseudoword reading WCPM

In pseudoword list reading (i.e., reading lists of made-up, but pronounceable words), students read made-up words that required them to use phonological decoding skills, the skills that enable a reader to pronounce sequences of letters based on knowledge of spelling-sound correspondences and orthographic patterns.

Like word reading skills, pseudoword reading skills declined across the NAEP reading achievement levels, including the below *NAEP Basic* subgroups (figure 5). The decline in mean performance was especially sharp between the below *NAEP Basic* **Medium** and below *NAEP Basic* **Low** subgroups. Moreover, there was a wide range within the below *NAEP Basic* subgroups. Fourth-graders in the below *NAEP Basic* **High** group read almost twice as many words correctly per minute (19) as those in the below *NAEP Basic* **Low** group (11), as shown in figure 5. The average number of pseudowords read correctly per minute was 22 for all fourth-grade students.

Figure 5. Average pseudoword reading WCPM, by NAEP reading achievement level and below *NAEP Basic* subgroup: 2018



* Statistically significant difference compared to the next higher NAEP reading achievement level category, $p < .05$. All comparisons were conducted with an alpha level of 0.05, with multiple pairwise comparison adjustments applied using the False Discovery Rate (FDR) procedure.

NOTE: Pseudoword is a made-up but pronounceable word. WCPM is an abbreviation for words correct per minute. The positions of the data points in the graphics are based on the unrounded numbers.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2018 Oral Reading Fluency study.

Conclusion and Implications

Conclusion

The 2018 ORF study reveals that for an estimated 1.27 million[12] fourth-grade public school students performing below *NAEP Basic,* and particularly for an estimated 0.42 million[13] fourth-grade students in the below *NAEP Basic* **Low** subgroup, fluent reading of connected text—sufficiently fast and accurate reading of sentences and passages—can be a major challenge. The study also shows that word reading and phonological decoding skills are underdeveloped in students performing below *NAEP Basic,* particularly for students in the below *NAEP Basic* **Low** subgroup.

Students in the below *NAEP Basic* **Low** subgroup not only have difficulty reading the words in the text quickly and accurately but also show a lack of appropriate expression in reading out loud, which is an indicator of poor comprehension. This makes it difficult for them to engage in the cognitive processes described in the 2017 NAEP reading framework. For an illustrative audio recording, see https://nces.ed.gov/nationsreportcard/studies/orf/illustrative_audio.aspx.

Implications

The NAEP reading framework and future assessments

First, the current reading framework does not describe any specific reading behaviors that characterize fourth-grade students performing below *NAEP Basic.* It states only that "These students are not necessarily nonreaders; many can complete some tasks on the assessment but are not able to attain the minimum score required for *Basic*" (National Assessment Governing Board 2017, p. 44). Based on the findings of this study, the new framework should incorporate a description of readers performing below *NAEP Basic.* It should acknowledge the fact that, compared to students performing at the *NAEP Basic* level or higher, students performing below *NAEP Basic* are more likely to have underdeveloped fluency, word reading, and phonological decoding skills. There should also be additional testing of fourth-grade students' oral reading fluency and foundational skills with a subsample of the students who take the main NAEP reading assessment. Such testing would provide much-needed information about the students who are performing below *NAEP Basic.*

[12] This number refers to 36 percent of 3.54 million (the number of public school, fourth-graders represented in the 2018 ORF study sample) = 1.27 million.
[13] This number refers to a third of 1.27 million fourth-grade students who performed below *NAEP Basic.* Recall that students performing below *NAEP Basic* were evenly divided into three groups based on the NAEP reading score distribution.

Second, the framework (p. 4) notes that text comprehension is influenced by phonics knowledge and fluency; and, importantly, it recognizes that "without these foundational skills, comprehension will not occur." It goes on to state a goal or aspiration for fourth-grade students that is universally accepted by reading experts and reading educators: "By grade 4, when the NAEP Reading Assessment is first administered, students should have a well-developed understanding of how sounds are represented alphabetically and should have had sufficient practice in reading to achieve fluency with different kinds of texts" (p. 4). But what if this goal has not been met?

In the future, the framework should acknowledge that: "Although the majority of fourth-grade students do not have problems with fluency, word reading, and phonological decoding, these skills are not adequately developed for a significant percentage of readers performing below *NAEP Basic,*" as shown by the findings of the 2018 NAEP ORF report (White et al. 2021).

Policy and research

First, the problems of fourth-grade students performing below *NAEP Basic* highlighted by this report call for a solution-oriented discussion among education policymakers. The discussion may begin with recognition of the large income-based gaps in prereading skills that exist at kindergarten entry (Quinn 2015; Reardon and Portilla 2016) and proceed to a fresh and intensive look at programs of instruction in preschools and the early elementary grades, especially programs that enroll large numbers of Black and Hispanic children. Second, research is needed to determine the extent to which elementary schools teach accurate and efficient word reading skills, in systematic ways, as supported by existing research (e.g., Castles, Rastle, and Nation 2018). This is a topic that is being vigorously debated in policy circles at the present time.

References

Baer, J., Kutner, M., Sabatini, J., and White, S. (2009). *Basic Reading Skills and the Literacy of America's Least Literate Adults: Results From the 2003 National Assessment of Adult Literacy (NAAL) Supplemental Studies* (NCES 2009-481). U.S. Department of Education. Washington, DC: National Center for Education Statistics, Institute of Education Sciences.

Castles, A., Rastle, K., and Nation, K. (2018). Ending the Reading Wars: Reading Acquisition From Novice to Expert. *Psychological Science in the Public Interest, 19:* 5-51.

Eason, S.H., Sabatini, J., Goldberg, L., Bruce, K., and Cutting, L.E. (2013). Examining the Relationship Between Word Reading Efficiency and Oral Reading Rate in Predicting Comprehension Among Different Types of Readers. *Scientific Studies of Reading, 17:* 199-223.

Foorman, B.R., Petscher, Y., and Herrera, S. (2018). Unique and Common Effects of Decoding and Language Factors in Predicting Reading Comprehension in Grades 1–10. *Learning and Individual Differences, 63:* 12-23.

Fuchs, L., Fuchs, D., Hosp, M.K., and Jenkins, J.R. (2001). Oral Reading Fluency as an Indicator of Reading Competence: A Theoretical, Empirical, and Historical Analysis. *Scientific Studies of Reading, 5:* 239-256.

Hoover, W.A., and Gough, P.B. (1990). The Simple View of Reading. *Reading and Writing: An Interdisciplinary Journal, 2:* 127-160.

McGlinchey, M.T., and Hixson, M.D. (2004). Using Curriculum-based Measurement to Predict Performance on State Assessments in Reading. *School Psychology Review, 33:* 193-203.

Metsala, J.L., and David, M.D. (2017). The Effects of Age and Sublexical Automaticity on Reading Outcomes for Students With Reading Disabilities. *Journal of Research in Reading, 40:* S209-S227.

National Assessment Governing Board. (2017). *Reading Framework for the 2017 National Assessment of Educational Progress.* Washington, DC: Author.

Perfetti, C. (2007). Reading Ability: Lexical Quality to Comprehension. *Scientific Studies of Reading, 11:* 357-38.

Quinn, D.M. (2015). Kindergarten Black–White Test Score Gaps: Re-examining the Roles of Socioeconomic Status and School Quality With New Data. *Sociology of Education, 88:* 120-139.

Reardon, S.F., and Portilla, X.A. (2016). Recent Trends in Income, Racial, and Ethnic School Readiness Gaps at Kindergarten Entry. *AERA Open, 2:* 1-18.

Reschly, A.L., Busch, T.W., Betts, J., Deno, S.L., and Long, J.D. (2009). Curriculum-based Measurement of Oral Reading as an Indicator of Reading Achievement: A Meta-analysis of the Correlational Evidence. *Journal of School Psychology, 47:* 427-469.

Sabatini, J., Wang, Z., and O'Reilly, T. (2019). Relating Reading Comprehension to Oral Reading Performance in the NAEP Fourth-Grade Special Study of Oral Reading. *Reading Research Quarterly, 42:* 253-271.

Share, D.L. (1995). Phonological Recoding and Self-teaching: Sine Qua Non of Reading Acquisition. *Cognition, 55:* 151-218.

Silverman, R.D., Speece, D.L., Harring, J.R., and Ritchey, K.D. (2013). Fluency has a Role in the Simple View of Reading. *Scientific Studies of Reading, 17:* 108-133.

White, S., Sabatini, J., Park, B.J., Chen, J., Bernstein, J., and Li, M. (2021). *The 2018 NAEP Oral Reading Fluency Study* (NCES 2021-025). U.S. Department of Education. Washington, DC: Institute of Education Sciences, National Center for Education Statistics.

Studies to Review and Revise NAEP Achievement Level Descriptions (ALDs) for Mathematics, Reading, and Other Subjects

Background

On September 24, 2020, the National Assessment Governing Board (Governing Board) awarded contract# 91995920C0004 to Pearson (as a result of a competitive bidding process) for conducting studies to review and revise NAEP achievement level descriptions (ALDs) in mathematics and reading using the 2019 NAEP assessments at grades 4, 8, and 12[1]. This work is intended to address the first recommendation of the evaluation of NAEP achievement levels that was conducted by the National Academies of Sciences, Engineering, and Medicine:

Recommendation #1: Alignment among the frameworks, the item pools, the achievement-level descriptors, and the cut scores is fundamental to the validity of inferences about student achievement. In 2009, alignment was evaluated for all grades in reading and for grade 12 in mathematics, and changes were made to the achievement-level descriptors, as needed. Similar research is needed to evaluate alignment for the grade 4 and grade 8 mathematics assessments and to revise them as needed to ensure that they represent the knowledge and skills of students at each achievement level. Moreover, additional work to verify alignment for grade 4 reading and grade 12 mathematics is needed.

The Board committed to conducting studies to review and revise the NAEP ALDs in its initial response to the evaluation that was formally adopted and sent to the Secretary of Education and Congress in December 2016. The Board's Achievement Levels Work Plan, adopted in March 2020, further describes the intention for this work: "Addressing Recommendation #1 should focus on the current reporting ALDs for mathematics and reading at grades 4, 8, and 12. The methodology will be similar to what was done to evaluate the alignment and revise the 2009 NAEP Reading ALDs for grades 4, 8, and 12 (Donohue, Pitoniak, & Beaulieu, 2010) and the 2009 NAEP Mathematics ALDs for grade 12 (Pitoniak, Dion, & Garber, 2010). This process will generate new reporting ALDs that comply with the revised Board policy statement" (p. 3).

According to Principle 1a of the Board policy on Developing Student Achievement Levels for NAEP, "Content achievement level descriptions translate the policy definitions into specific

[1] The base period of this contract includes the review and revision of ALDs in mathematics and reading at grades 4, 8, and 12; in addition, an option may be exercised for a second phase of the contract focusing on review and revision of ALDs in U.S. history, civics, science, technology and engineering literacy (TEL) at grade 8 based on data from the most recent administrations of those assessments in 2018 and 2019.

expectations about student knowledge and skills in a particular content area, at each achievement level, for each subject and grade. Content ALDs provide descriptions of specific expected knowledge, skills, or abilities of students performing at each achievement level. They reflect the range of performance that items and tasks should measure. When setting achievement levels, the content ALDs provide consistency and specificity for panelist interpretations of policy definitions for a given assessment. During reporting, content ALDs communicate the specific knowledge and skills represented by *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced* for a given assessment" (p. 5).

Principles 3g and 4a of the Board policy apply specifically to this project of reviewing and revising the current ALDs and creating reporting ALDs (based on empirical data) that indicate what students at each achievement level *do* know and *can* do rather than what they *should* know and *should* be able to do[2]. Additional details for carrying out the work described by principles 3g and 4a are included in the Achievement Levels Procedures Manual.

The basis for the evaluation of NAEP achievement levels (and subsequently for this project) is the *existing* NAEP frameworks and item pools, not the new NAEP Mathematics Framework currently scheduled for implementation in 2026 or the NAEP Reading Framework that is currently under development and consideration by the Board. In accordance with principle 4b of the Board policy, the achievement levels and/or ALDs will need to be reviewed again once the new frameworks are implemented. Such work is beyond the scope of this project.

Project Overview

Dr. Eric Moyer is the project director at Pearson and Dr. Jennifer Galindo is the assistant project director at Pearson. Pearson will conduct a pilot study and an operational meeting using scale anchoring studies where panels of content experts judge the alignment of the current mathematics and reading ALDs and produce a set of recommended reporting ALDs for the Governing Board to consider in reporting the results from the next regular administration of the NAEP reading and mathematics assessments at grades 4, 8, and 12. The Governing Board is expected to take action on the reporting ALDs for mathematics and reading at grades 4, 8, and 12 in advance of the next release of these results.

Based on careful review of the history of ALD development, review, and revisions for NAEP mathematics and reading, a model-based anchored approach for reviewing the alignment of the ALDs for NAEP mathematics and reading will be used. The methodology for this alignment review study is based on that of previous studies, including the ALD development and review meeting held in 2009. The methodology was specified by the Board's Achievement Levels Work Plan and was selected to reduce the potential for possible inconsistencies from the use of different methods. The process of the model-based anchored approach will result in organizing

[2] According to the Board policy, ALDs will continue to describe what students *should* know and *should* be able to do for the purposes of item development and standard setting; only the reporting ALDs will be written in terms of what students *do* know and *can* do.

specific NAEP items by achievement level, which will serve as a key referent for panelists in reviewing and revising the current ALDs.

The model-based anchored approach includes three stages. The first stage will involve conducting statistical analyses to determine the items from the subject and grade that are anchored to a level corresponding to the score range within cut scores set to represent the achievement level descriptors (ALDs). The second stage relies on panels of content experts for each individual assessment. The panelists individually review the items that are anchored to each performance level and create summary descriptions of what students in each level are expected to know and be able to demonstrate based on the knowledge and skills measured by the items. In the final stage, the panelists compare the current ALDs for the respective assessment with their summary descriptions. The panelists note the similarities and differences, to make a recommendation regarding whether the current ALDs accurately describe what students in each level are expected to know and be able to demonstrate or if revisions to the current ALDs are needed to improve alignment. The final alignment judgment will be used to report whether the panels determined that there exists alignment between the current ALDs and student expectations. The final panel summary descriptions will be used to revise the current ALDs to create reporting ALDs that indicate what students at each achievement level do know and can do.

There is a technical advisory committee (TAC) consisting of the following experts in ALDs:

Dr. Karla Egan (Principal, EdMetric)

Dr. Ellen Forte (CEO and Chief Scientist, edCount)

Dr. Susan Loomis (Independent Consultant)

Dr. Marianne Perie (President, Measurement in Practice)

Dr. Mark Reckase (University Distinguished Professor Emeritus, Michigan State University)

Dr. Lauress Wise (Principal Scientist, Human Resources Research Organization)

The TAC is scheduled to meet for more than 100 hours (approximately 4 hours per month, with additional meeting time following the pilot and operational meetings) to provide technical advice on all aspects of the project to review and revise the mathematics and reading ALDs; this is intended to help ensure that all procedures, materials, and reports are carried out in accordance with current best practices, providing additional validity evidence for the process and results. In addition to frequent meetings and reviews of materials, two TAC members will attend the pilot and operational meetings to observe and provide feedback on the process.

In response to previous COSDAM discussions, the project schedule was modified to account for conducting the panel meetings in person in late 2021 and early 2022. The pilot meeting will take place in Atlanta on October 25-28, 2021, and the operational meeting will take place in the same location on February 22-25, 2022. The resulting ALDs will be presented for Board discussion at the May 2022 Board meeting and Board action at the August 2022 Board meeting. The intention is for the ALDs from this project to be used in the reporting of NAEP results in fall 2022.

Project Update (August 2021)

During the May 2021 COSDAM meeting, the final Design Document for the NAEP ALD Review study was discussed by the Committee members and there were no recommended changes. Based on this approval of the study design, project staff continued to work with the TAC in reviewing procedures and materials for the study. During the TAC meetings over the past couple of months, the TAC has provided valuable feedback on the role of text complexity as part of the panelist item review process and the inferences that can be made from the ALD alignment review process. These discussions with the TAC were useful in evaluating and revising the materials for the study, to ensure that we are utilizing the most appropriate procedures to fulfill the purposes of the study.

An important part of the study process is the recruitment of panelists. The multi-phase recruitment process for this study will begin with a panelist nomination process, where individuals from different organizations and state departments of education will have the opportunity to nominate outstanding classroom and non-classroom educators to participate in the process. Materials and procedures for this nomination process were reviewed by the TAC to ensure that the recruitment process results in set of representative panelists. The nomination materials have been finalized and the nomination process will begin in July.

Next Steps

During the next few months, the project staff will finalize the meeting presentation materials and the study website and will start training facilitators in September in preparation for the October pilot meeting. The recruitment process will continue the nomination phase while also collecting information on nominated panelists, which will be used to select the set of representative panelists for the pilot study in October.

During the November COSDAM meeting, project staff will provide a brief preliminary update on the pilot study, which will be held in late October. After a more thorough review of the pilot study results, a more complete briefing will be provided during a COSDAM webinar to be scheduled for December or January.