

An Exploration of Opportunity to Learn and Implications for NAEP

Sheila W. Valencia

University of Washington

James W. Pellegrino

University of Illinois at Chicago

Richard P. Durán

University of California, Santa Barbara

October 2024

Commissioned by the NAEP Validity Studies (NVS) Panel

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

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

Panel Members:

Keena Arbuthnot
Louisiana State University

Peter Behuniak
Criterion Consulting, LLC

Derek Briggs
University of Colorado Boulder

Jack Buckley
American Institutes for Research

Phil Daro
*Strategic Education Research Partnership (SERP)
Institute*

Richard P. Durán
University of California, Santa Barbara

David Grissmer
University of Virginia

Gerunda Hughes
Howard University

Akisha Osei Sarfo
Council of the Great City Schools

James Pellegrino
University of Illinois at Chicago

Gary Phillips
Cambium Assessment

Jennifer Randall
University of Michigan

Lorrie Shepard
University of Colorado Boulder

David Thissen
University of North Carolina, Chapel Hill

Gerald Tindal
University of Oregon

Sheila Valencia
University of Washington

Denny Way
Consultant

Project Director:

Sami Kitmitto
American Institutes for Research

Project Officer:

Grady Wilburn
National Center for Education Statistics

For Information:

NAEP Validity Studies (NVS) Panel
American Institutes for Research
1400 Crystal Drive, 10th Floor
Arlington, VA 22202-3289
Email: naepvaliditystudies@air.org

CONTENTS

INTRODUCTION.....	1
OTL—A BRIEF HISTORICAL PERSPECTIVE	5
Pre-NCLB.....	5
Post-NCLB.....	6
OTL and Large-Scale Assessment.....	7
RECONSIDERATION OF OTL: IMPLICATIONS OF CONTEMPORARY UNDERSTANDINGS ABOUT INSTRUCTION AND LEARNING.....	10
Implications of a Sociocultural Framing of Learning and Instruction for OTL.....	13
NAEP AND PISA CONTEXTUAL FRAMEWORKS AND SURVEY QUESTIONNAIRES	15
NAEP	15
PISA.....	17
NAEP DISCIPLINARY FRAMEWORKS AS GUIDES FOR FUTURE OTL INDICATORS.....	20
CONSIDERATIONS RELATED TO GENERATING AND USING OTL DATA.....	24
Beyond Item Data: Construction of OTL-Related Scales.....	24
Multivariate Frameworks for the Analysis of OTL and Academic Achievement	25
FINAL THOUGHTS AND CONSIDERATIONS	28
Concluding Thoughts	30
REFERENCES.....	31

ACKNOWLEDGMENTS

We want to thank the members of the NVS Panel for their feedback on earlier drafts of this paper. We extend special appreciation to Lorrie Shepard for her insights and suggestions. Thanks, too, to Brittany Boyd and members of the AIR staff for their support.

INTRODUCTION

The aim of this white paper is twofold: (1) to explore the importance of opportunities to learn (OTL) for interpreting National Assessment of Educational Progress (NAEP) results and (b) to expand our thinking about OTL to include indicators of the social contexts of classroom teaching and learning that have emerged from research in the last 20 years. Our ultimate goal is to help the National Center for Education Statistics (NCES) begin to explore how OTL data tied to instructional practices and their outcomes can inform the design, reporting, and interpretation of NAEP subject-matter achievement results and related actions of policymakers and researchers.

NAEP began collecting background information associated with its assessments since its origination (1969–70). At that time, background information was limited to gender, race/ethnicity, and literacy materials in the home. Since then, the scope of background questions has expanded greatly to include additional demographic factors, teacher qualifications, home and school resources, student motivation, and subject-specific instructional content and practices collected from teacher, student, and school questionnaires.

Although current NAEP documents do not directly define OTL or provide reports specifically dealing with OTL, the background questionnaires target several OTL factors identified in research (e.g., classroom resources, teacher quality, content covered, home factors, classroom organization). The most recent NAEP Contextual Framework (National Assessment Governing Board [NAGB], 2013), still used today, calls for collecting a range of “information beyond the academic assessment that NAEP uses to make its academic results more meaningful to the public” (p. 15). The framework uses the terms “non-cognitive,” “background,” and “contextual” interchangeably, connecting one of the goals of data collection to OTL:

[I]nform educational policy by describing the contexts for learning, sometimes called opportunity-to-learn (Mullis, 2002). Broadly, this involves the content specified in the curriculum, whether and how that content actually is taught, students’ propensity to learn, as well as home and school factors that can enhance learning. (p. 16)

Users of NAEP data have always had online access to the survey questions and results through the NAEP Report Card releases. (We will look more closely at the questionnaires later in this paper.) In June 2023, a new online tool, the NAEP Achievement Gap Tool, debuted, which includes selected data from the surveys so that users can gain a deeper understanding of achievement gaps and the “differential learning opportunities” of diverse students (Goodman & Walton, 2023). Thus, context indicators have been part of the NAEP data collection for some time. However, other than reporting descriptive response rate estimators to questionnaire items at the population or subpopulation level with occasional statistical significance tests, these contextual indicators are not in wide use (Floden, 2002; NAGB, 2013).

Our goal in this paper is to build on NAEP’s current approach to defining and assessing contextual variables—to add new thinking and suggest possible indicators that capture contemporary research understandings of instruction and learning at the classroom level.

Our position is that understanding the conditions known to influence the learning of students from diverse backgrounds and those who underperform on NAEP requires expanding traditional definitions and measurement of OTL beyond accountings of content/topic coverage, indicators of poverty, teacher characteristics, and general school behaviors, for example. In findings from the National Academy of Sciences, Engineering, and Medicine (NASEM; 2019); Education First (2022); and many researchers, indicators of educational equity are essential for understanding student performance and informing more effective supports and interventions. More complex and nuanced understandings of OTL have emerged from research in the learning sciences and social sciences that require us to broaden the definitions and indicators of OTL to include contextual factors that mediate and moderate learning at the classroom level.

It is beyond the scope of this paper to characterize and explore the expansive range of indicators of the many variables—and the interaction of variables—associated with OTL, such as the distribution of resources, teacher professional development, and disciplinary content covered. Our aim, instead, is to open up thinking about one OTL variable typically called “instructional quality.” We will call attention to research and theory on the sociocultural contexts of instruction and learning as experienced by classroom participants. This focus does not disregard OTL indicators at the school, home, or community levels, which are important to consider. We elected to concentrate on classroom indicators of instructional quality because they often have been defined by proxy variables, such as teacher experience and credentials, topic coverage, and particular procedures. However, thanks to research on the complex social, cognitive, and cultural nature of instruction and learning, we know a good deal about factors that contribute to equitable learning and the importance of gathering indicators that are closer to actual instruction-learning interactions. Doing so, as part of the large-scale context of NAEP, should increase the likelihood of understanding variation in student achievement across time and jurisdictions (Bennett, 2023; Floden, 2002; Marion, 2020; Moss et al., 2008; Organisation for Economic Co-operation and Development [OECD], 2023a). As a large-scale assessment, NAEP has limited ability to directly assess classroom interactions. However, we believe it can get closer to critical features of those interactions through updated survey questions, data-gathering strategies, and special studies. In fact, the Programme for International Student Assessment (PISA) 2022 Context Questionnaire Framework (OECD, 2023a) acknowledges that how classroom teachers facilitate students’ thinking and understanding of disciplinary content and concepts has the strongest “direct school-based influence” on student achievement (p. 200). As a result, in addition to OTL questions that examine classroom organization and content exposure, the PISA questionnaire examines instructional quality (i.e., Teaching Practices and Learning Opportunities) through multiple indicators of teachers’ classroom instructional practices and behaviors aligned with domain-specific content and more general achievement.

Three factors motivated and shaped this white paper. First, the 2013 Contextual Framework identified one of its goals as providing descriptive information that would allow NAEP to track changes in contextual and informational factors related to achievement measures and the distribution of educational resources (NAGB, 2013). Since the adoption of the Contextual Framework in 2013, and even before, a great deal of research illuminated the influence of social contexts on schooling, instruction, and learning in today’s complex and diverse classrooms. As NAEP programs and educators continue to try to understand persistent disparities in academic achievement, they also must consider OTL gaps implicated

from this research. Valid interpretations and uses of test (NAEP) results must be contextualized (National Academy of Education [NAE], 2021), which requires a closer examination of OTL—one that is more inclusive of social and culturally diverse ways of instruction and learning, most especially those that are proximal to the classroom and how instruction and learning occur (NAE, 2021; OECD, 2019). Differences among subgroups reported on NAEP and other large-scale assessments often are interpreted as academic learning differences residing in the group rather than differences attributable to the group's contexts of teaching and learning. In short, groups of students are often referred to as “falling short” rather than labeling the context of teaching and learning as “falling short” or failing to provide students with access to high-quality instruction related to the tested content and practices. As our field was reminded in 1993, “Failure to try to capture opportunity to learn as comprehensively as possible is *prima facie* evidence that an achievement study is flawed” (Burstein, cited in Rothman, 1993). It's time to update the NAEP Contextual Framework and reexamine definitions and indicators of OTL.

Second, issues of equity and quality of instruction in schools, not simply outcomes, drive the work of the NAEP Validity Studies Panel. In addition to providing context for understanding and interpreting NAEP results, the collection of OTL data is essential to identifying inequalities of access among students from different social, ethnic, racial, gender, socioeconomic, language, and disability groups (National Education Policy Center, 2023; Pullin, 2008). President Biden's Executive Order 15 of January 2021, “Advancing Racial Equity and Support for Underserved Communities Through the Federal Government,” sets the stage for NCES and NAEP to help identify inequities that impact school learning and participation. Specific to our goal here, the Executive Order states as follows:

The Federal Government's goal in advancing equity is to provide everyone with the opportunity to reach their full potential. Consistent with these aims, each agency must assess whether, and to what extent, its programs and policies perpetuate systemic barriers to opportunities and benefits for people of color and other underserved groups. Such assessments will better equip agencies to develop policies and programs that deliver resources and benefits equitably to all. (White House, 2021)

The NCES strategic plan (specifically Goal 4: Embed the principles of diversity, equity, inclusion, and accessibility in all aspects of the Center's work, studies, and data products) elaborates the “critical need for NCES to continue to identify and monitor the full range of indicators that illuminate inequitable opportunities and inform solutions for helping all students reach their fullest potential.”¹

According to NASEM (2019), two types of equity indicators are necessary: (a) indicators of disparities in student achievement and engagement in schooling and (b) indicators that measure equitable access to resources and opportunities. We, along with many other researchers (Herman et al., 2000; Marion, 2020; NAE, 2019; NASEM, 2019), view gathering and reporting on an expanded range of OTL indicators, informed by contemporary research, as important in its own right—to track how students have opportunities and receive supports to actively participate in deep learning. Systematic differences in equitable OTL represent what some have called an “opportunity gap” (Carter & Welner, 2013). Most importantly,

¹ <https://nces.ed.gov/about/?sec=stratplan>

many of these opportunities are both malleable and actionable, increasing the likelihood that student learning can advance (NASEM, 2019). Documenting OTL indicators across time will provide a lens on the contexts of instruction and inequalities of access among students from different racial, socioeconomic, cultural, linguistic, gender, and disability groups, and it can inform policies and practices that support learning and achievement (National Education Policy Center [NEPC], 2023; Pullin, 2008). These indicators also can serve as leading indicators of academic progress because changes in these opportunities often precede improved student learning.

A third driver for this paper is the new NAEP Frameworks for Reading, Math, and Science, and their alignment with classroom instruction as implemented. All three frameworks draw on contemporary research on instruction and learning that acknowledges the importance of disciplinary learning in the social and cultural contexts of schools. These content frameworks will likely have implications for revising the Context Framework and specific survey items. In fact, NCES was directed² to develop an “issues paper” to reflect the contextual information priorities and identify data needs that are set forth in new content frameworks.

In the sections that follow, we first turn to a brief review of OTL conceptualized across time in terms of definition and policy, as well as the collection of OTL indicators in large-scale assessment programs such as NAEP. We then discuss current understandings about the factors that operate across contexts from society at large to the classroom that shape instruction and learning and their implications for OTL indices. The latter includes considerations of those contextual factors included in contemporary large-scale assessment programs such as NAEP and important missing factors that bear consideration for future index development and reporting in NAEP. With this backdrop, we discuss the current NAEP Contextual Framework, focusing on its limitations as well as suggestions for change. The paper continues with a discussion of possibilities for change in the assessment of OTL using the new NAEP content frameworks as a potential basis for doing so. It further provides an example of a newly revised and adopted content framework for NAEP (i.e., the 2026 Reading Framework) to illustrate what may be possible. The penultimate section considers issues of reporting and the use of OTL information with an emphasis on the development of scales for relevant OTL factors and their possible use in multilevel analyses of OTL factors associated with NAEP achievement. The final section summarizes the major points made in the preceding sections and offers some additional thoughts about collecting and using OTL information in the NAEP program and its importance to understanding educational progress for all segments of our student population.

² <https://nces.ed.gov/nationsreportcard/tdw/instruments/noncog.aspx>

OTL—A BRIEF HISTORICAL PERSPECTIVE

Our brief historical discussion of OTL and the NAEP assessment program includes three topics. The first two involve inclusion of OTL in educational policy discussions in the United States. We first consider the period prior to the advent of large-scale assessments of academic achievement as part of the educational accountability era ushered in by the No Child Left Behind (NCLB) Act (2001). We then consider the period following initiation of NCLB up to the present, with an emphasis on reactions to persistent findings of significant disparities in the academic achievement of different demographic groups. The third topic relates to the first two and is somewhat parallel in time. It focuses on the inclusion of so-called contextual or OTL factors in large-scale data collections focused on academic achievement nationally (NAEP) and internationally (First International Math Study [FIMS], Second International Math Study [SIMS], Trends in International Mathematics and Science Study [TIMSS], PISA, and Progress in International Reading Literacy Study [PIRLS]), including the contribution of such data to educational policy and equity discussions.

Pre-NCLB

As noted by Marion (2020), OTL is a more than 50-year-old concept that evolved from a focus on whether students have had sufficient access to instruction or content linked to particular concepts to a more robust conception regarding the conditions and resources provided inside schools that enable students to succeed. The general concept of OTL entered the educational policy realm in the early days of the standards-based reform movement during the latter part of the 20th century, with “debates over educational standards and whether schools had a responsibility to provide students with an adequate opportunity to learn before they could be held accountable for meeting such achievement standards. That debate came to a head in 1994 with the passage of the federal Goals 2000 legislation” (McDonnell, 1995, p. 306). This legislation resulted in a call for OTL-related policies that describe the resources that state and district policymakers should provide if students and teachers are to meet the high expectations prescribed by state content standards.

McDonnell (1995) reviewed efforts to use OTL indicators as policy instruments in the 1980s and 1990s, including a short-lived push for enacting school delivery standards, and accountability provisions associated with these standards. The National Council on Education Standards and Testing (1992) used the term “school delivery standards” instead of OTL to push policymakers to recognize that this matter was as critical to standards-based reform as were content standards and assessments: “School delivery standards should provide a metric for determining whether a school ‘delivers’ to students the ‘opportunity to learn’ well the material in the content standards” (p. E-5). The high watermark of these policy efforts was the Goals 2000: Educate America Act (1994), which called for standards to assess “the sufficiency or quality of the resources, practices, and conditions necessary at each level of the education system to provide all students with an opportunity to learn the material in voluntary national content standards or State content standards.”

McDonnell (1995) also highlighted the technical and political challenges that limit the use of OTL as a policy instrument, many of which continue to be relevant 30 years after the enactment of Goals 2000 in 1994. First, the precise definition of OTL can differ across contexts, from narrower binary indicators of curriculum coverage, breadth, or depth to

richer operationalizations involving school and classroom processes, pedagogical approaches and instructional practices, school resources, and a range of other elements of the instructional climate. Thus, the collection, interpretation, and reporting of OTL data, and its expected relationship to outcomes is not straightforward and may not be feasible without significant investment of resources and, where accountability is involved, political capital (McDonnell, 1995). At that time, neither resources nor political will could sufficiently make OTL a driving force.

Post-NCLB

NCLB ushered in an era of large-scale testing of the academic achievement of America's students accompanied by state and district accountability for the attainment of increasingly higher performance levels on tests of rigorous academic standards in mathematics, English language arts, and science. Results from those assessments have consistently highlighted significant and persistent disparities in the achievement of various demographic groups, including students whose backgrounds and living conditions reflected low socioeconomic status (SES); students with disabilities; and students from various racial and ethnic groups, including African American, Hispanic, and Native American students. Such demographic differences in achievement also have been regularly reported at national and state levels for NAEP assessments in reading and mathematics.

One of the purported goals of NCLB was to uncover disparities in educational attainment and focus resources on reducing and eliminating any inequalities that might be responsible for the results. Although an assessment and accountability policy such as NCLB, and its successor the Every Student Succeeds Act (ESSA), seeks to identify areas of need and inform resource allocation, it has been demonstrated in the ensuing 20-plus years that any goal related to equity of outcomes is not achievable by the reporting of outcome measures alone, regardless of the breadth or the level of detail provided regarding achievement; investment in teacher capacity building is essential (Elmore & Rothman, 1999). Moreover, informed decisions regarding policies and actions critically require documenting not only learning outcomes but also the resources and opportunities offered to learners to achieve those outcomes. As a 2022 report from the Aspen Institute Education and Society Program noted, "Opportunities to learn—the resources, experiences, and expectations students get access to—enable students to pursue their purpose, develop their agency, and contribute as community members and informed citizens" (p. 2). That report called on state leaders to take a strategic approach to collecting and making sense of OTL data, including analyses of disparities among groups, clear and actionable reporting mechanisms, and supports for continuous improvement. ESSA, which replaced NCLB, opened the door for states to include OTL measures in their accountability system. However, we have yet to see much movement in this direction because state accountability systems continue to be dominated by achievement test results, including measures of both academic status and growth.

Thus, it is not surprising that educational policy discussions have increasingly turned to the importance of measuring and monitoring OTL for achieving equity in the U.S. educational system. As mentioned earlier, large-scale achievement testing and accompanying accountability provisions, as mandated in federal reporting policies such as NCLB and ESSA, have failed to have much—if any—impact on closing achievement gaps. This is despite the fact that all demographic groups have shown some improvements in NAEP

performance scores over time, at least prior to the pandemic. Despite those gains, the racial, ethnic, and SES performance gaps remain substantial. It is worth mentioning that well before the current assessment and accountability era, McDonnell (1995) emphasized the value of OTL as a generative concept that can offer a vision of high-quality, equitable educational opportunities. OTL data also can help highlight differences in the educational experiences and opportunities afforded to different groups of students and how these might relate to disparities in achievement, both across and within groups. Thirty years after the Goals 2000 discussion (1994), a new wave of interest in OTL is best exemplified by a report from the National Academies calling for the development of systems to monitor educational equity (NASEM, 2019), along with proposals to develop subject-specific OTL standards in language, mathematics, and the arts among others (see, e.g., National Association for Music Education, 2020; National Council of Teachers of English, 2019; Wang et al., 2022). As defined in the NASEM report, equity requires that educational opportunities consider students' needs to counter "the effects of structural disadvantages that disproportionately affect different student groups" (NASEM, 2020, p. 1). What remains to be determined is the nature of those educational opportunities that matter most and the most appropriate and valid ways to measure and report them.

OTL and Large-Scale Assessment

There is a long tradition of collecting OTL data as part of surveys of educational conditions and as components of testing programs. As noted earlier, the collection and reporting of contextual data contributed to the policy debates involving evidence of substantial inequities in student achievement outcomes on various standardized tests of academic achievement.

The concept of OTL first originated in the 1960s in FIMS (1964) and the subsequent SIMS in the early 1980s to ensure the validity of cross-national comparisons in International Association for the Evaluation of Educational Achievement (IEA) studies of mathematics achievement (see Floden, 2002). OTL was operationalized as a measure of curricular coverage—whether students in the sampled countries had opportunities to study particular topics or solve particular types of problems included in the test—known as “curriculum exposure.” As such, OTL was useful to explain differences in achievement across countries, essentially asking if the assessment was “appropriate” for individual countries. It also was useful for comparing curricular foci across and within participating countries, making this measure of OTL important in its own right, regardless of achievement results. In fact, Floden (2002) noted that OTL was first treated as a “given” background variable (FIMS) and then considered as a malleable and actionable factor that included organizational and teaching variables in addition to curricular choices (SIMS, 1976–1982). As we noted earlier, interest in malleable and actionable OTL indicators are currently reemerging as valuable outcomes in their own right, a position reminiscent of Floden’s early insights.

Since the initial collection of contextual factors in the early IEA mathematics surveys, the collection of these so-called contextual factors has become a consistent feature of multiple large-scale achievement testing programs, including NAEP, TIMSS, PISA, and PIRLS. All four of these large-scale assessment programs include at least a student, teacher, and school questionnaire, and all, except NAEP, include a home questionnaire. The contextual frameworks associated with each assessment refer to OTL as an important element of the assessment for interpreting achievement results, informing research, and improving

education. However, only the PISA program directly defines OTL and produces reports that refer to OTL. Acknowledging that OTL should account for the quality of instruction (e.g., cognitive activation, presentation of lessons), quantity of instruction (e.g., content exposure, content emphasis), and a range of other variables related to domain-specific and general achievement (e.g., disciplinary climate, teacher support), the PISA 2022 Context Questionnaire Framework expanded its definition of OTL for 2022:

OTL could be defined as all contextual factors that capture the cumulative learning opportunities a student has been exposed to at the time of the assessment (Bertling, Marksteiner & Kyllonen, 2016). These contextual factors may comprise both learning opportunities at school and informal and formal learning opportunities outside of school. (OECD, 2023a, p. 180)

Each assessment organizes contextual data and reporting categories somewhat differently. For example, the PIRLS 2021 Context Questionnaire Framework (Mullis & Martin, 2019) details four questionnaires administered to students, parents, teachers, and school principals that provide data for PIRLS reporting. Overall, the questions target three types of data: demographic data (e.g., gender, SES) used for reporting results for subgroups; background data about the status of students, parents, and the home (e.g., students' reading skills when they enter primary grades, library resources, teacher credentials); and student reading behaviors and instruction (e.g., reading enjoyment, organization for instruction). The PIRLS 2021 report (Mullis et al., 2023) combined the data from sets of related questions into scales, rather than reporting results of individual questions, which results in nine scales across three underlying constructs: home environment support; school composition, resources, and climate; and students' reading attitudes and behaviors. The scales are reported with respect to student achievement.

In contrast to PIRLS, the PISA 2022 Context Questionnaire Framework (OECD, 2023a) organized contextual data into two categories of constructs: (a) domain-specific constructs related to the tested domain and (b) general constructs related to achievement across multiple domains. Data from student and school questionnaire items are categorized into 21 modules for reporting in five educational policy areas: student background; student beliefs, attitudes, feelings, and behaviors; teaching practices and learning opportunities; school practices, policies, and infrastructure; and governance, system-level policies, and practices. As with PIRLS, taken as a whole, the questionnaire items and the constructed reporting categories tap a broad range of data for different purposes (e.g., reporting by subgroup, the nature of disciplinary instruction, system policies). The PISA 2022 Contextual Framework suggests that three modules most closely align with OTL: organization of learning in school, exposure to mathematics content, and mathematics teacher behaviors. Note, however, that the 2022 results for contextual data were not reported using these modules.

The NAEP Contextual Framework organizes contextual data into four categories: demographic reporting categories, SES, public policy contextual factors, and subject-specific contextual data (NAGB, 2013). In 2019, NAEP reported the following scales for both reading and mathematics subject-specific data: student internet access and digital technology at home, content/instructional emphasis in [subject], confidence in [subject] knowledge and skills, and teacher satisfaction and views of school resources. In comparison, the reporting categories in 2017 were different: persistence in learning, academic self-discipline, enjoyment

of complex problems, confidence in [subject] knowledge and skills, and interest/enjoyment in [subject].³ Although these reporting categories span a wide range of variables, some of which may be outside OTL (e.g., student and home characteristics), our point is that changes in reporting categories would make it difficult for NAEP to monitor shifts in OTL across time and within/across groups.

Overall, the PIRLS, PISA, and NAEP frameworks include a mix of concepts, indicators, and reporting categories. Although some of the content categorized as “contextual” might represent malleable and actionable OTL, other content targets indicators used for creating subgroups or for reporting more static indicators of schools and classrooms.

We turn next to considerations of sociocultural perspectives on learning and instruction that have emerged in the last 50 years. Our aim in the following brief summary of relevant theory and research is to (a) bring together concepts of OTL contextual indicators and these understandings and (b) consider implications for the validity of interpretations of NAEP results and issues of equity.

³ https://nces.ed.gov/nationsreportcard/tdw/instruments/2017/noncog_indices_read_g4_2017.aspx

RECONSIDERATION OF OTL: IMPLICATIONS OF CONTEMPORARY UNDERSTANDINGS ABOUT INSTRUCTION AND LEARNING

Much has changed in this century about our conceptions of knowing and learning and the implications for classroom instruction and assessment. Three such developments have implications for considerations of past, present, and future conceptions and indices of OTL. What follows is a brief encapsulation and interweaving of the three developments and their implications for OTL.

- Theories, models, and data on disciplinary knowing and learning have changed substantially and are best represented by a broad sociocultural perspective on the general nature of knowing and learning combined with theory and research on the disciplinary organization of knowledge and skills and discipline-specific learning progressions where feasible.
- Standards and expectations for academic proficiency have changed substantially and now demand more complex understandings as well as integrated knowledge of core disciplinary knowledge and practices and the capacity to transfer and apply that knowledge in new contexts. The new standards have major implications for the design and implementation of curriculum and instruction, along with corresponding principled approaches to assessment and measurement.
- Coordination and integration of curriculum, instruction, and assessment are essential to achieve equitable and coherent classroom learning environments and are best accomplished when all three derive from conceptual models and empirical data on disciplinary knowing and learning.

The past two decades include the publication of important syntheses regarding the general nature of learning and knowing plus the specifics of learning and knowing in the academic disciplines (e.g., NASEM, 2018; National Research Council [NRC], 2000). Within that broader literature, the argument is that sociocultural theory provides the most compelling general theory of knowing and learning because it accounts for how meanings, purpose, values, and motivation are jointly developed as part of learning and helps understand the ecology of classroom instruction and learning (e.g., NASEM, 2018; Shepard et al., 2018). In this perspective, learning takes place as individuals participate in the practices of a community, using the tools, language, and other cultural artifacts of that community. These practices, artifacts, and ways of knowing and relating to others in classrooms do not occur in isolation. They bear relationships to the experiences, practices, and value systems in the surrounding world and experience beyond the school walls that classroom participants bring with them. Thus, learning is “situated” within a specific classroom community setting in which participants engage in the particularly emphasized practices of that community for the teaching, learning, and assessment of academic disciplines, including reading/literacy, mathematics, science, and history, and also learn to respect and draw connections to each other’s lived experiences, shared from home and the community.

As argued by Shepard et al. (2018), sociocultural theory offers a powerful, integrative account of how motivational aspects of learning—such as self-regulation, self-efficacy, sense of belonging, and identity—completely intertwine with cognitive development. Sociocultural

theory helps us see how knowledge and self-beliefs are jointly developed in communities of practice, and, correspondingly, how sense of self and meaningful participation may be harmed in unsupportive learning environments in which children are labeled as unable or deficient learners.

Sociocultural approaches make it possible to design studies of instructional equity in educational settings by attending both to who learners are when they join a community and who they might become. That is, these approaches consider what is at stake for learners when they invest in developing new knowledge or skills in practice in a particular context given an implemented curriculum and its compatibility with students' responsiveness to those learning opportunities. Countless studies have documented the lack of learning that occurs when students are treated as deficient or when members of nondominant communities by race, language, or gender identity are asked to park their identities at the door to join the mainstream school or college culture (see, e.g., Eccles & Wigfield, 2020). In contrast, sociocultural approaches such as funds of knowledge (Moll et al., 1992) and cultural modeling (Lee, 1995) pay attention to students' everyday practices and connect "varied repertoires of practice" with academic disciplinary practices (Nasir et al., 2014).

The sociocultural perspective has important implications for the teaching and learning of academic disciplines in school. For example, the disciplines are distinct communities that engage in shared practices of ongoing knowledge creation, understanding, and revision. Thus, it is not surprising that a sociocultural perspective is reflected in the most recent disciplinary standards for primary and secondary education. For example, the science framework (NRC, 2012) and the derivative Next Generation Science Standards (NGSS Lead States, 2013) call for integrated development of socially negotiated science practices, cross-cutting concepts, and core ideas of the discipline. They posit that proficiency involves much more than being able to recite core ideas; proficiency requires application of those ideas to explain phenomena in the world and solve problems using the kinds of practices that disciplinary experts use. A similar argument is possible about the mathematics standards (Common Core State Standards, 2010), with their articulation of both mathematical practices and core disciplinary content and their integration in solving problems and constructing explanations. The recent NAEP frameworks for science and mathematics incorporate the ideas of multiple dimensions of knowing, including core disciplinary content and disciplinary practices and their integration for purposes of learning and assessment. The Reading Framework for the 2026 NAEP (NAGB, 2021b) acknowledges that reading comprehension involves both cognitive and social processes to make meaning. Further, that framework expands the role of disciplinary reading by systematically sampling comprehension in literature, social studies, and science and by having students read to solve problems and use and apply knowledge gained through reading.

Sociocultural theory has multiple implications for curriculum, instruction, and assessment in the classroom at a broad level, including the role that instruction and assessment play in the culture of the classroom learning environment, how students come to understand the processes and practices that constitute them in the classroom, and how students and teachers use them to advance their trajectories of learning. Sociocultural theory posits that in addition to probing disciplinary knowledge, a key role for assessment is to elicit students' relevant interests, experiences, and identities, as well as the classroom contexts and interactions in which they learn; all help inform instructional practices and impact learning

that can help students develop identity and become members of the disciplinary community. This framing of learning has major implications for issues of equity and fairness.

Goldman and Lee (2024) summarized much of the current understandings related to sociocultural theory and research on human learning and development. Box 3-1 from their chapter summarizes the foundational principles of human learning and development that have implications for teaching, learning, and assessment.

BOX 3-1

Foundational Principles of Human Learning and Development

- Learning entails dialogic relations among thinking, emotional salience attributed to experience, and perceptions of the self along multiple dimensions
 - Thinking and the role of prior knowledge in new learning
 - Conceptual knowledge
 - Procedural knowledge
 - Epistemology
 - Emotional salience
 - Perceptions of safety and self-efficacy (e.g., growth mindset)
 - Perceptions of relevance
 - Perceptions of the self
 - As an individual
 - As a member of cultural communities of practice (family; social networks; interest networks; and institutional settings such as schools, community organizations, age cohorts, etc.)
- Relationships matter
- Participation in routine cultural practices within and across settings, within and across time
 - Affordances of artifacts, belief systems, and practices in cultural communities
 - Relationships across different cultural communities of practices
 - Where learners are in the life course
- Learning is malleable across the life course
- Diversity in developmental pathways is essential for the human species

The principles for human learning and development are broad statements about how to conceptualize and operationalize various facets of schooling, including curriculum, classroom instruction, and assessment. Although sociocultural theory establishes some broad principles about the nature, roles, and uses of instruction and assessment in education in general, and classrooms in particular, the specifics of instruction and assessment need to be supplemented by discipline-specific models of learning, including learning progressions (i.e., trajectories; see Shepard et al., 2018). Two of the most extensive discussions of the learning progression construct and its implications for curriculum, instruction, and assessment are in the Consortium for Policy Research in Education reports *Learning Progressions in Science* (Corcoran et al., 2009) and *Learning Trajectories in Mathematics* (Daro et al., 2011). As described therein, learning progressions are empirically grounded and testable hypotheses about how students' understanding of, and ability to use, core concepts and explanations and related disciplinary practices grow and become more sophisticated across time with appropriate instruction.

Not surprisingly, a parallel body of work on quality instruction aligns with a sociocultural framing of disciplinary learning and knowledge development. As discussed in a recent chapter by Ruiz-Primo and Furtak (2024), many instructional approaches have been described under the umbrella of “robust” or “ambitious” teaching (Goldman & Lee, 2024; Hammond, 2021; Lampert et al., 2013; Smith et al., 2001; Shepard, 2021; Windschitl et al., 2018). Common principles include

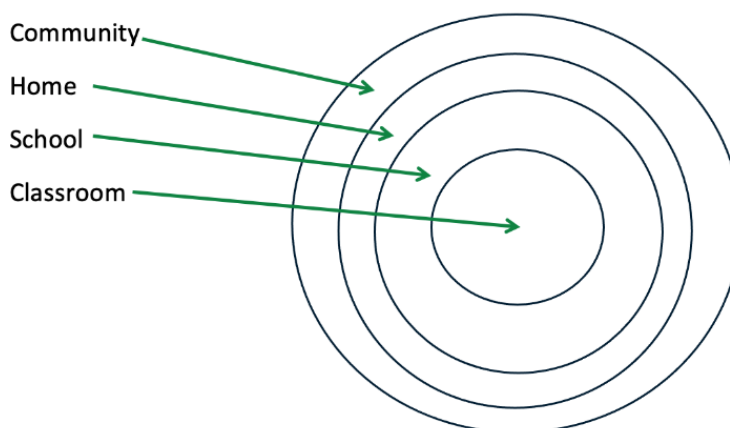
- centering the interests, funds of knowledge, and experiences of students from diverse cultural and linguistic backgrounds to new learning;
- engaging students in rich, authentic tasks with scaffolds to support their participation and conceptual understanding;
- inviting students to be active co-constructors of and participants in their learning through productive classroom discourse that involves reasoning, explaining, analyzing, and justifying;
- developing students’ disciplinary knowledge, discourse, and practice in a community of learners;
- fostering student agency, participation, and self-regulation;
- developing norms of respect, caring, responsibility, and improvement; and
- including assessment designed and enacted in alignment with these goals.

Ambitious teaching principles are grounded in sociocultural theories of learning (Shepard, 2021), situate learning in a cultural context organized by tools and routines (e.g., Engeström, 2001; Greeno, 2006), and conceptualize learning as changes in participation in practice (Lave & Wenger, 1991). A sociocultural view compels us to consider (a) classrooms as organized disciplinary communities of practice that attend to the interconnected cognitive, social, emotional, and cultural facets of learning and development (Goldman & Lee, 2024) and (b) both teachers and students as key participants who bring their previous knowledge, identities, and lived experiences into learning environments (Gutiérrez & Rogoff, 2003). All this, of course, takes place in discipline-specific contexts in which teachers, as well as students, vary in their cognitive, social, and cultural ways of participating and understanding (Cohen et al., 2018; Grossman et al., 2014).

Implications of a Sociocultural Framing of Learning and Instruction for OTL

From a sociocultural perspective, performance on assessments, such as previous NAEP achievement tests and those that will be forthcoming based on new content frameworks, is the product of factors operating at multiple levels, ranging from the larger community and the home to the district, school, and classroom context, as illustrated in Figure 1.

Figure 1. Sociocultural Framing of Learning and Instruction



Clearly, a sociocultural framing of OTL recognizes the relevance of the types of variables that historically are among the contextual factors used to develop items in previous surveys for NAEP and other large-scale assessment programs. What is missing, however, is a focus on the key OTL factors that are most proximal to learning and instruction and that contribute to the sociocultural context of the classroom.

Our focus in this paper is on one set of OTL indicators often referenced as “quality of instruction.” From a sociocultural perspective with respect to assessment, indicators of instructional quality include—and go beyond—traditional measures of alignment of tested material with taught curriculum, time, practice with specific tasks and test formats, and so forth. Specifically, we suggest that indicators of the quality of instruction should be grounded in ambitious teaching principles as well as disciplinary content, ways of knowing, and pedagogy. They should include attention to students’ academic, social, and cultural needs as they interact with teachers and peers to put knowledge into practice—all in the service of meaningful learning. From this foundation, we will next review NAEP’s existing survey questionnaires and new frameworks for guidance on improving measures of OTL in the NAEP program and in interpreting the meaning of NAEP achievement test outcomes. Our purpose is to consider whether a class of indicators more proximal to the contexts for instruction and learning, and rooted in what we know about ambitious teaching and disciplinary learning, could yield information about critical factors that contribute to disparities in educational outcomes while also suggesting interventions known to be effective to impact learning for a culturally diverse population of learners.

NAEP AND PISA CONTEXTUAL FRAMEWORKS AND SURVEY QUESTIONNAIRES

This section will briefly explore how contemporary understandings of the social contexts of instruction and learning are reflected in 2024 NAEP teacher and student surveys and how new or revised questions might better tap OTL that are proximal to classroom experiences and practices. In doing so, we also take a brief look at the PISA Questionnaire Framework and a sample of PISA items for questions that tap proximal measures of instruction.

NAEP

In 2012, the NAGB adopted a policy statement on *NAEP Background Questions and the Use of Contextual Data in NAEP Reporting*, updating the existing approach to collecting background data to be more expansive. An expert review committee working on the 2013 framework suggested that making greater use of contextual data could “provide rich insights into a wide range of important issues about the nature and quality of American primary and secondary education” (NAGB, 2013, p. 5). As a result, the *Contextual Information Framework for the National Assessment of Educational Progress* (2013) was written and approved—a framework still in place today.

The 2013 framework provides a brief history of NAEP’s collection of contextual information beyond the academic portion of the assessment. It defines the purposes and scope of NAEP’s system of collecting contextual information plus the criteria for reporting. One important influence on the framework was NCLB, which aimed to close the performance gaps among specific groups of students. As a result, NAEP was positioned to not only identify gaps in achievement but also provide descriptive information that would allow NAEP to track changes in contextual factors related to achievement and the distribution of educational resources (NAGB, 2013).

Here we briefly review a subset of NAEP Reading student and teacher questionnaire items that fall under the heading of “subject-specific contextual data outlined in the 2013 framework. According to the 2013 framework, NAEP collects “contextual information” to

- report achievement results by community;
- inform educational policy by describing contexts for learning (i.e., OTL);
- support research into factors related to student achievement;
- support technical functions of NAEP, such as evaluating potential bias, sampling, and weighting; and
- prepare estimates of student proficiency distributions (NAGB, 2013, pp. 16–17).

The term “OTL” is further defined in the framework as follows: “Broadly, this involves the content specified in the curriculum, whether and how that content actually is taught, students’ propensity to learn, as well as home and school factors that can enhance learning” (NAGB, 2013, p. 16). This definition leaves substantial room for interpretation and consideration of updated research.

The purposes for collecting contextual information are still relevant today, as is the broad definition of OTL. However, as noted throughout this paper, definitions have expanded to include a more interactive, contextually influenced learning environment. To gain a better idea of how new NAEP frameworks have operationalized contextual factors, we categorized general teacher and student survey items administered as part of both the 2019 NAEP Reading and Math assessments and the content-specific items for mathematics using fine-grained categories related to instruction.⁴ The data in Table 1 represent all possible responses to the survey. That is, for questions with one stem (e.g., how often do you do each of these things . . .) and many parts (e.g., summarize, read alone, write answers to questions), each response is counted as a question. As we reviewed the content of the questionnaire items, we considered how more up-to-date understandings of the cognitive, social, and cultural contexts of instruction and learning are reflected in the questions targeting classroom instructional interactions. Later, Table 1 provides tabulations according to our categories for eighth-grade mathematics.

Table 1. Tabulation of NAEP Questionnaire Context Variables for Eighth-Grade Mathematics

Contextual indicator	# of student items		# of teacher items	
	General	Discipline specific	General	Discipline specific
Demographic information (student, school, district)	2		3	
Student factors (engagement, motivation)	14	4		
Teacher factors (qualifications)			33	
School/classroom factors (resources)	2		9	7
Community factors (home resources)	18			
Instruction quality—general teaching behaviors	3	2	9	5
Disciplinary curriculum and content emphasis		1		20
Disposition toward specific discipline		12		
Out-of-school experiences with discipline		5		2
Other				

Our preliminary review of a subset of survey questions and categories (those targeting classroom instruction) provides guidance on revisions or additions that reflect the cognitive, social, and cultural contexts of classroom learning. The following summarizes the results from our analyses:

- Student factors (e.g., effort, self-efficacy, identity, engagement, motivation) are disproportionately general rather than subject specific. Because these subject-matter content and classroom interactions influence student dispositions, it would be useful to situate more of these in classroom instruction of the NAEP-tested subject.

⁴ Since the writing of this paper, NCES new Survey Questions have been developed in conjunction with the new Reading and Mathematics Frameworks. It would be useful to review these in light of the recommendation made in this report (See <https://nces.ed.gov/nationsreportcard/experience/archive.aspx>)

- Student questions targeting dispositions toward specific disciplines ask about enjoyment and effort in mathematics. Many ask about enjoyment of specific mathematical “topics.” It would be helpful to know more about engagement, rather than enjoyment, with mathematical concepts, practices, problems, and use—more about mathematical identity.
- School/classroom factors primarily ask teachers about how well technological devices work, not what students do with them, and not in mathematics classes. Other general questions ask about classroom conditions (e.g., overcrowding, space). Teacher disciplinary questions ask about access to mathematics materials; they don’t ask how students use them.
- Instructional quality questions were asked of students and teachers. Student questions aimed at remote learning and computer use in mathematics classes. Teacher questions asked about school-level collaboration in general and about grouping and differentiating instruction in mathematics classes.
- Teacher questions about disciplinary curriculum and content asked about emphasis placed on teaching particular mathematics topics, practices, and processes. Again, these do not investigate how teachers engage students during instruction. It would be helpful to know how this aligns with students’ responses to similar questions. Also important would be learning more about how teachers teach and interact with students.

With our focus on classroom interactions (teacher-student, student-student) that impact learning, we suggest revising several NAEP survey questions to highlight the nature and character of instruction and, more directly, OTL. In addition to knowing what students do, we want to learn more about how they experience learning—how teachers support them to learn, build knowledge, and develop identities as learners in the subject being tested. In addition, consideration of the cultural and social nature of learning in the NAEP-tested discipline should be explored.

PISA

In addition to examining NAEP items, we briefly reviewed PISA, an international assessment administered to 15-year-olds in 79 countries and economies, for guidance about measuring and reporting indicators related to OTL. Questionnaire items are allocated more than twice the amount of time as compared with NAEP (36 minutes vs. 15 minutes) and cover a broader range of topics than those represented on NAEP. PISA’s perspective on OTL aligns with understandings of the sociocultural contexts of learning. The PISA 2022 Questionnaire Framework (OECD, 2023a) broadened its definition and perspective on OTL beyond earlier iterations to include both school and home experiences. It calls particular attention to classroom-based instruction as the “immediate and core” setting of formal education and further notes the importance of “quality of instruction” in addition to mere quantity (p. 16). Thus, the framework calls attention to effective instruction, particularly how teachers facilitate student thinking and understanding of subject-matter content and concepts and the “proximal variables” such as classroom characteristics and teaching and learning practices that more closely relate to student achievement than school- or system-level variables. Teacher and student survey questions include conventional OTL indicators such as content coverage,

content exposure, teachers' qualifications, time, and classroom climate. It includes both general teaching practices and, in 2022, domain-specific constructs such as the following:

- Disciplinary climate in mathematics
- Mathematics teacher feedback
- Structure of mathematics instruction
- Cognitive activation in mathematics
- Teachers' use of assessment and mathematics teacher feedback
- Mathematics teacher qualification and in-house professional development

In addition, PISA 2022 asked students how often they had seen and had experiences with tasks similar to those on the test (e.g., complex reasoning, math problems with more than one solution).

PISA 2018 (OECD, 2019) did not report out directly about OTL. However, separate chapters in the report cover student questionnaire topics related to OTL, and most are discipline specific:

- Teacher enthusiasm
- Teachers' support and teaching practices
- Teacher behavior and student learning
- Student co-operation and competition
- Sense of belonging at school
- Parent involvement in school activities
- Students' life satisfaction and meaning in life
- Students' feelings
- Students' self-efficacy and fear of failure
- Growth mindset

A closer look at the structure of several of the student questions provides interesting models that get closer to interactions between teachers and students during classroom instruction for the tested subject. The following are some examples.

Student Questionnaire: Teacher Enthusiasm (agree-disagree)

- "It was clear to me that the teacher liked teaching us."
- "The enthusiasm of the teacher inspired me."
- "It was clear that the teacher likes to deal with the topic of the lesson."
- "The teacher showed enjoyment in teaching."

Student Questionnaire: Teachers' Support and Teaching Practices (how often)

- “The teacher shows an interest in every student’s learning.”
- “The teacher gives extra help when students need it.”
- “The teacher helps students with their learning.”
- “The teacher continues teaching until students understand.”

Student Questionnaire: Teacher Feedback (agree-disagree)

- “The teacher made me feel confident in my ability to do well in the course.”
- “The teacher listened to my view on how to do things.”
- “I felt that the teacher understood me.”
- “The teacher tells me in which areas I can still improve.”
- “The teacher tells me how I can improve my performance.”

In summary, across both NAEP and PISA, many questions relate to OTL but are not worded or situated closer to actual teaching and learning experiences in the tested discipline and do not reflect the complexity of ambitious teaching. Moreover, we do not get a sense of the social and cultural nature of either instruction or learning. Clearly, it is difficult for large-scale assessments to gather such data. Nevertheless, issues of equity and interpretations of NAEP achievement results require that we try to do so. The next section considers how this might be done by considering the NAEP’s current disciplinary frameworks as guides for development of survey items.

NAEP DISCIPLINARY FRAMEWORKS AS GUIDES FOR FUTURE OTL INDICATORS

Guidance for rethinking OTL indicators is in the new NAEP Frameworks for Reading, Mathematics, and Science. As noted in the *NAEP Technical Documentation for Contextual Items (Survey Questionnaires)*, “When a new assessment framework is approved, or when new policy issues are identified for NAEP to address, NCES conducts a literature review and develops an issues paper to reflect the new priorities, and identifies the data needed to address the issues.”⁵

As of this writing, we do not have access to either the Math or Reading issues papers. So based only on the frameworks as published, we briefly describe how they align with sociocultural perspectives on knowing, learning, and instruction, and, specifically, how they might inform the development of OTL indicators. We begin with an overview of how the Mathematics Framework addresses the role of sociocultural perspectives in learning. Then, using the Reading Framework as an example, we look in more detail at how that framework might inform OTL survey questions and data collection.

The Mathematics Framework Development Panel was charged by its Visioning Panel to address issues of OTL. It recommended an “expansive conception of opportunities to learn” mathematics content, processes, and practices as well as updates to the contextual survey items. The Mathematics Framework (NAGB, 2021a), itself, provides background on research-based factors that affect student learning, including time, mathematics content and practices, teacher factors, instructional organization and strategies, and instructional resources. It further points to the social and cultural influences on instruction and learning that shape students’ interests, motivation, mathematical identities, and sense of community. All these factors align with the cognitive and sociocultural perspectives discussed earlier. The framework raises concern that without proper attention to gaps in students’ learning opportunities, “achievement gaps” can be misinterpreted as differences in student ability rather than differences caused by unequal and inadequate OTL.

Similar to the Mathematics Framework, the 2026 Reading Framework incorporates several priorities of contemporary research on instruction and learning, including disciplinary literacy; the role of affect, motivation, and agency in shaping readers’ performance; the role of social and cultural experiences in human development and learning, particularly in reading comprehension; and the principles of Universal Design of Assessments to embody the experiences of students from diverse backgrounds (NAGB, 2021b, p. 4). The framework also incorporates contemporary research into an expanded definition of reading comprehension, acknowledging the influence on meaning-making of different purposes for reading, students’ prior experiences and knowledge, and text types (e.g., digital) and tasks. All are potential sources of information about OTL. Reading comprehension, as defined for the NAEP Reading Assessment, attends to the cognitive, affective, and contextual influences on meaning-making:

⁵ <https://nces.ed.gov/nationsreportcard/tdw/instruments/noncog.aspx>

Reading comprehension is making meaning with text, a complex process shaped by many factors, including readers’ abilities to

- engage with texts in print and multimodal forms;
- employ personal resources that include foundational reading skills, language, knowledge, and motivation; and
- extract, construct, integrate, critique, and apply meaning in activities across a range of social and cultural contexts. (NAGB, 2021b, p. 10; emphasis in original)

In addition to the expanded definition of comprehension, the framework’s new section on contextual variables provides specific recommendations for the collection of OTL data. A summary of these appears as Exhibit 4.2 in the framework: a table of reader characteristics and environmental characteristics that shape students’ OTL (see Table 2). Data from these variables can help researchers, educators, and policymakers interpret differential performance on NAEP. The data could be collected through traditional questionnaire items and, most recently, in digital form through process data recorded while students are completing the test.

Table 2. Contextual Variables Identified in 2026 Reading Framework (NAGB, 2021b)

Variables	Source		
	Student Questionnaire	Teacher/Administrator Questionnaires	Process Data
Reader Characteristics			
Cognitive and Metacognition			
Cognitive strategies	✓	✓	✓
Metacognitive strategies	✓		✓
Topical knowledge	✓	✓	
Engagement and Motivation			
Volume of reading	✓	✓	✓
Reading for enjoyment	✓	✓	
Motivations for reading	✓	✓	
Environmental Characteristics			
Reports of School and Community Resources			
School social support	✓	✓	
Belonging in school	✓	✓	
Participation in out-of-school reading/literacy activities	✓		
Reports of Teacher, Instructional, and Classroom Supports			
Teacher support for reading engagement	✓	✓	
Teacher support for motivation	✓	✓	
Teacher support for students’ background experiences	✓	✓	
Program and curricular support for reading development	✓	✓	

The contextual variables in Table 2 represent important and several new, possible OTL indicators for examining interactions more proximal to instruction and learning as well as students’ dispositions and sense of belonging. One of the most useful categories for the purposes of this paper is “Reports of Teacher, Instructional, and Classroom Supports.” Here we imagine survey items that tap the exchange between teacher and student—“teacher support for”—that could be asked of both teachers and students. Although not worded in the same way in this chart, we can imagine structuring questions that ask about learning and

instruction of cognitive strategies, metacognitive strategies, belonging in school, and so forth. For example, it would be important to know what contributes to students' sense of belonging by situating questions in a discipline-specific lesson (e.g., students helping one another solve mathematics problems, a teacher who answers questions, being recognized during reading discussions, being able to use home language to explain a character's motive). Similarly, if process data reveal that students are not using metacognitive strategies while reading, it would be useful to know if they had opportunities to learn those strategies.

The second feature of the Reading Framework that should inform the development of contextual items is in the updated definition of reading comprehension and the design of the assessment. Table 3 gives examples of the changes found in the 2026 Reading Framework and how those might shape new contextual items designed to gather information about instruction-learning interactions in the social contexts of classrooms.

Table 3. Possible OTL Questions Derived From 2026 Reading Framework

New Reading Framework	Possible OTL concepts and survey items for teacher-student interactions ^a
<i>Disciplinary Contexts & Texts:</i> Students read passages in three subjects: literature, social studies, science. Scores will be reported for reading in these different contexts.	What is the nature of the support that students get for learning how to read in these different disciplines? What is the nature of their engagement? • My teacher helps me understand when I read my social studies material. • My teacher shows us how to find evidence to support our answers in science class. • I talk with my classmates about the ideas in the stories we read. • My teacher asks us to explain how we figured out an answer to a problem in math. • My teacher reads our social studies book aloud to the whole class. <i>(These are especially important in middle school where the questions should target teacher support in non-ELA classes.)</i>
<i>Comprehension Targets:</i> Students must respond to questions that ask them to use and apply what they have learned. <i>Purpose:</i> Reading to Solve a Problem	What types of tasks are students doing that require using information from text? How are teachers facilitating this purpose for reading? Have they engaged in tasks such as these in class? • I use information from two or more books or online sources when I work on a science project. • I use information that I read in books or online to solve problems. • After we read, my teacher helps us talk about what we think the story means. • My teacher encourages us to share different ways to solve problems. <i>(These items cross with disciplinary differences.)</i>
<i>Text and Response Formats:</i> Students read from digital texts, manipulate them, responding digitally (briefly and more extensively).	In the classroom, how do teachers use digital devices with students? What kinds of work (tasks) do students do on a computer or other digital device? How do teachers support students to produce written work on a computer? • My teacher shows me how to search for information on the computer. • After reading, we write answers to questions on a tablet. • How do you use computers in your classroom – to search for information? – to write stories? – to play games? – to do worksheets? <i>(These items cross with disciplinary differences, especially in middle school.)</i>

^a These examples are framed as student survey questions. Similar questions could be posed to teachers.

Overall, the new NAEP content frameworks should be particularly useful in rethinking OTL indicators that could provide critical contextual information for interpreting NAEP results and generating insights into possible actions to achieve more equitable educational outcomes. Once collected, NAEP will need to create new models for reporting such data that are at once helpful but are not misunderstood to be causal. In the next section, we briefly consider some of the reporting and use issues as NAEP explores new and more valid OTL indices.

CONSIDERATIONS RELATED TO GENERATING AND USING OTL DATA

Much of our discussion has focused on OTL indicators as they have been developed and implemented to date and how to modify and/or expand indicators for future data collection purposes. Obviously, the collection of such contextual information is important for a variety of purposes in a program such as NAEP, not the least of which is reporting specific results for consideration in various analyses of demographic and other factors relating to achievement in specific disciplinary areas. Although the NAEP program is limited in the types of analysis and reporting it typically does, it is certainly worth considering what gets reported and how others will use those results in analyses pursuing specific interpretive work on the OTL factors associated with variation in performance on NAEP tests. In the discussion that follows, we offer two considerations regarding the collection and reporting of OTL indicators in NAEP assessments: (a) continued development and reporting of interpretive scales and (b) relationships among variables and scales.

Beyond Item Data: Construction of OTL-Related Scales

Historically, NAEP designed its survey questionnaires around single questions, and the questionnaire results were reported for single questions. In 2014, the program moved toward an enhanced design and reporting approach to examine information of key interest to NAEP audiences. Specifically, although some survey questions are still analyzed and reported as single items (e.g., gender), several questions on the same topic are combined into indices measuring a single underlying construct or concept. This approach aims to provide both breadth and depth of coverage of the factors being examined.

As noted in the Contextual Information Framework for NAEP (NAGB, 2013), the current approach to survey questionnaire design and reporting directly addresses NAGB's policy principles laid out in their 2012 policy statement, particularly the principles that "NAEP reporting should be enriched by greater use of contextual data derived from background or non-cognitive questions asked of students, teachers, and schools" (National Assessment Governing Board, 2012, p. 2). This approach also aligns with the design and reporting approaches followed in large-scale international assessments, including PISA, TIMSS, and PIRLS.

Index scores for contextual factors have been reported for several NAEP assessments, including the 2019 Mathematics, Reading, and Science assessments.⁶ Indices were reported for Grade 4 and Grade 8 Reading and Grade 4 and Grade 8 Mathematics. The Grade 4 mathematics indices included represented both subject-specific and more general learning behaviors. For example, mathematics-specific indices included (a) students' confidence in mathematics knowledge and skills; (b) students' performance goals in mathematics; (c) students' mastery goals in mathematics; (d) students' interest/enjoyment in mathematics. More general learning indices include (a) persistence in learning; (b) students' enjoyment of complex problems; and (c) students' academic self-discipline. Consistent with our earlier reasoning, the nature of the survey items and the derived indices, even those that are discipline-specific, do not reflect actual experiences related to the processes of classroom teaching that might affect learning outcomes. For example, the four-item index of enjoyment

⁶ https://nces.ed.gov/nationsreportcard/tdw/instruments/2019/noncog_indices_math_g4_2019.aspx

of complex problems includes the following questions: “(a) I enjoy situations where I will have to think about something, (b) I enjoy thinking about new solutions to problems, (c) I like complex problems more than easy problems, and (d) I like activities that challenge my thinking abilities.” (NAEP Mathematics Report Card, n.d.). In the 2012 PISA assessment (OECD, 2014), OTL items focused on content (e.g., experience with mathematical tasks, familiarity with mathematical concepts, and exposure to types of mathematical tasks in lessons and in tests), teaching practices (e.g., teacher-directed instruction, student orientation, formative assessment), and teaching quality (e.g., classroom organization and management, teacher emotion and social support, cognitive activation). In mathematics, three indices were developed from the student questionnaire items: formal mathematics, an indicator of the extent of student exposure to algebra and geometry topics in classroom instruction; word problems, an indicator of the frequency that students encountered word problems in their mathematics schooling; and applied mathematics, an indicator of the frequency students in their mathematics classes encountered problems that required the application of mathematics either in a mathematical situation or in a real-world context. An interesting use of these data appeared in a recent paper by Wang et al. (2022). They used PISA 2012 data provided by 470,745 fifteen-year-old students. They argued that results from psychometric analyses and multilevel structural equation modeling indicated that (a) OTL measured by familiarity with mathematical concepts displayed high reliability and validity, (b) student-reported OTL exhibited high reliability at the school level, (c) OTL significantly predicted achievement at the student and school levels, and (d) OTL mediated the relationship between self-efficacy and mathematics achievement at the student level.

As detailed earlier, PISA continues to broaden and refine its notion of OTL to include learning opportunities at school and outside school. Most recently, PISA conceptualized OTL in three different modules, identified as

opportunities provided through the ways in which student learning is organized (Module 14), opportunities defined based on the mathematics content students are exposed to (Module 15), and opportunities created based on the behaviours teachers exhibit in the classroom (Module 16)” (OECD, 2023a, p. 181.

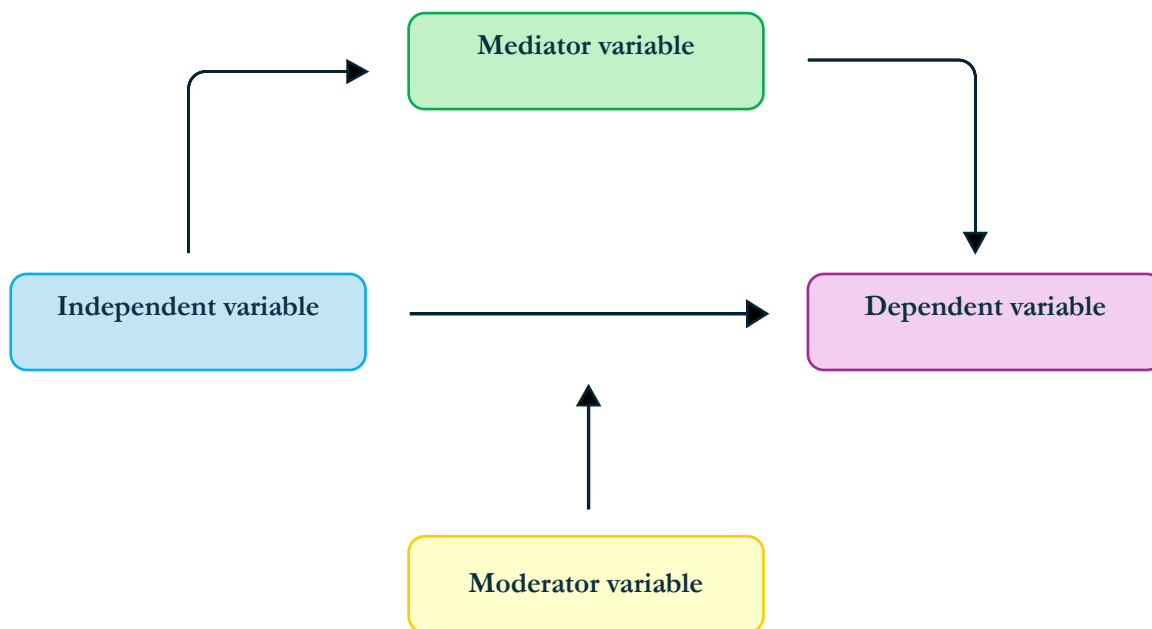
To date, however, PISA has not reported results using these modules.

Multivariate Frameworks for the Analysis of OTL and Academic Achievement

As summarized earlier, it may prove useful to organize the sets of variables investigated through surveys and other data sources into indices to pursue analyses of relationships between OTL factors and achievement scores. The example mentioned previously for the 2012 PISA mathematics results is a case in point. We think that consideration might be given to various multivariate analytic frameworks. One such example is the use of moderator and mediator analyses of the effects of specific OTL factors on academic achievement scores. The efficacy research literature contains numerous studies using such a framework to understand the effects of various factors, including experimental interventions, on valued outcomes—educational and otherwise. In this analytic framework, a mediator variable “explains” the relationship between two variables (i.e., the mediator variable captures relevant situational features seemingly responsible, in whole or in part, for the relationship);

see the examples after Figure 2. A moderator variable, in contrast, affects the direction or strength of the relationship between those two variables. A moderator relationship, also known as an interaction or conditioning effect, is one in which the independent and dependent variables have a relationship, but by adding the third, or a moderating variable, the relationship changes. Moderators point out the when, who, or under what circumstances the relationship holds (Figure 2).

Figure 2. Use of Moderators and Mediators in the Analysis of Relationships Among Variables



Adapted from <https://www.scribbr.com/methodology/mediator-vs-moderator/>

An example related to academic achievement may help clarify this role regarding the types of OTL indicators collected in large-scale assessment programs and analyses of the relationships between such variables and test performance. One frequent indicator collected in various large-scale assessment programs is information about the coverage of specific content, such as encountering specific types of problems (e.g., word problems) in the mathematics curriculum and mathematics performance on a large-scale survey such as PISA (e.g., Wang et al., 2022). A mediator variable could be students' motivation and interest in mathematics problem-solving lessons. Thus, any relationship between types of mathematics problems presented and test performance is mediated by how engaged students are in the instructional activities. Another mediating variable for the impact of mathematics problem solving could be the manner in which students participate in activities with those problems in the classroom—such as collaborative problem solving and discussion of the reasoning behind their answers—which mediates the intensity of the effect of the measure of problem-solving experience on mathematics achievement. In this example, a moderator variable could be the individual students teachers' disciplinary specialization and/or the teacher's sense of self-efficacy in teaching mathematics. Thus, it is possible that the effect of the mathematics problem-solving variable is not only mediated by the nature of students' experiences in the classroom in engaging with problem solving but also moderated by factors such as teacher

experience in creating efficacious student interactions and/or other variables such as class size and curricular materials.

There are at least two ways to use this framework in considering work done to date in collecting and analyzing OTL information in large-scale assessment programs. First, it can help sort variables or indices into one of the two possible categories of variables. It may well be the case that many of the factors and indices for which data have been obtained fall into the category of moderator variables; at a sociocultural analysis level, they may be relatively distal from those instructional factors actually transpiring in the classroom that directly influence student learning. For those factors and indices in which some connection is assumed between the factor and academic achievement, such as coverage of specific types of mathematics content, what may be lacking is information in the form of a mediating variable, such as how students engage with the content, which explains or clarifies the nature of the relationship between the OTL variable and achievement. The mediating variable may be the primary variable of interest for purposes of an OTL interpretive analysis.

The second way in which the mediator-moderator analytic framework may help in future thinking about OTL in the NAEP program is identifying needed directions for the collection of OTL indices that will have interpretive value in understanding variation in NAEP outcomes. In other words, it may help identify those types of OTL data reflecting factors that generally lie much more proximal to actual teaching and learning interactions, which have been identified from research on learning and instruction as conceptually important for learning and achievement outcomes from a sociocultural perspective on teaching and learning (e.g., NASEM, 2018).

We offer the mediator-moderator framework with some considerable caution, not the least of which is attempting to draw strong inferences of causality given the nature of any available data sources in a program such as NAEP. Furthermore, one person's mediator variable may well be another person's moderator variable given differing perspectives on the constructs and measures involved. This caution should be observed for any multivariate approach to the analysis of variables related to academic achievement outcomes on a large-scale assessment such as NAEP. In addition, various OTL indices should not be construed as predictor variables for interpretive analyses of variation in NAEP test score results. As we have argued elsewhere in this paper, OTL indices are important in their own right and thus could well serve as the outcomes to be predicted and understood using a mediator-moderator or other multilevel analytic framework. Perhaps the most important final point we can make in this section is to note that the NAEP program needs to pursue development and refinement of OTL indices that reflect more proximal factors known to impact learning and achievement that researchers can then use in various multivariate analyses of important educational outcomes, not the least of which is OTL.

FINAL THOUGHTS AND CONSIDERATIONS

Our review of OTL, and our particular focus on OTL indicators that consider the cognitive, social, and cultural contexts of classroom instruction, provide insights about possible ways forward and considerations for future work in the NAEP program. We summarize some of our key insights in this section and provide some implications.

- The NAEP Contextual Information Framework was last updated in 2013. It should be updated. In doing so, we suggest that the developers of the new framework consider both research and theory on the complex sociocultural nature of disciplinary instruction and learning. Such a foundation will illuminate the many factors that influence the learning of students from diverse backgrounds. It widens the lens on OTL. In addition, we suggest that newer NAEP content frameworks (i.e., reading, math, science) should figure into thinking about contextual variables. Finally, we suggest that the updated contextual framework distinguish the different categories of contextual information and their utility for interpreting NAEP assessment results. For example, demographic variables used for reporting subgroup performance (e.g., ethnicity, gender, SES, EL status); background variables that describe somewhat “static” conditions related to achievement (e.g., number of students enrolled in specific classes, access to home internet, teacher background, student perception of test difficulty); and OTL variables that describe teaching and learning conditions and practices that may be malleable and actionable based on policy and/or practice. These distinctions will highlight the differences among contextual variables and the importance of OTL as distinct from the other categories of contextual information.
- As much as possible, OTL survey questions should be designed to reveal classroom interactions that influence learning—teacher-student, student-student—all embedded in discipline-specific activity structures and processes specified in updated NAEP subject-matter frameworks. Currently, few survey items ask about these important classroom interactions. Existing items should be reviewed/revised and new items developed that measure the nature of classroom instruction. Questions should attend to the cognitive, social, cultural, and emotional experiences of students and teachers as they interact in the classroom. Neither teaching nor learning is discipline neutral; students’ interests, prior experiences, identities, or engagement also are not neutral. Situating questions closer to the “action” of the classroom instructional environment may provide additional insights into opportunities that foster or hinder learning for particular groups of students.
- NCES/NAEP should develop strategies to monitor and report subject-specific and general indicators of OTL across time. In addition to these data, serving as the context to interpret achievement results and helping educators and policymakers examine students’ equitable access to ambitious teaching practices, they constitute measures of educational progress that complement NAEP achievement results. The OTL indices also will provide a rich database for secondary analyses conducted by NCES and educational researchers. NCES should make NAEP users (educators, policymakers, journalists) aware of OTL indicators and results and actively support

discussions regarding educational progress in OTL for students from diverse backgrounds.

- We support the continued use of OTL-related scales and indices for reporting in general and within subgroups because such scales have greater interpretive value than single item responses. We believe that new scales will need to be created to capture features of important interactions and processes operating in the classroom learning environment. We also suggest considering the relationship of such scales to each other in terms of their value for prospective multilevel analyses.
- NAEP should explore how to take advantage of digital formats to design survey questions and gather OTL data. For example, process data collected during NAEP Reading could be used in conjunction with survey questions to examine reading processes used during test taking and opportunities to learn those processes during instruction. Another possibility is to design the questionnaires in a branching format, in which students or teachers who respond as having had particular experiences in classroom instruction can be routed to more in-depth questions about those experiences. Finally, the digital format might allow students to see representations of items/tasks they have just encountered on the assessment to ask about familiarity.
- A particular challenge for understanding OTL and NAEP Reading achievement is the important addition of tasks that assess reading in the disciplines (literature, science, social studies). The Reading Framework clearly lays out the different purposes, understandings, and uses of disciplinary texts and calls for reporting separate scales. If OTL data are to add context to scale interpretation, it will likely need to be collected from subject-matter teachers at the middle and high school levels and to differentiate interactions by text types at the fourth-grade level. We imagine that similar considerations of understanding OTL will exist for mathematics and science achievement given the nature of those new frameworks and the types of tasks to be included in the forthcoming assessments.

Although some of these ideas may follow the typical survey questionnaire development process, others may require special studies. In addition, once available, OTL will provide a rich database for additional study. The following are some initial thoughts about special studies.

- Although there is a good deal of research on supportive classroom social contexts and instructional practices that contribute to discipline-specific student engagement and learning, NAEP will need to conduct a program of research to establish the validity and reliability of OTL indicators. We imagine these will include in-depth literature reviews, case studies, and statistical analyses.
- Studies should be conducted to examine the relation of OTL indicators and NAEP achievement results to determine which items/factors (clusters of items) are useful in understanding NAEP performance on discipline-specific assessments. Research should consider oversampling specific groups of students, especially those historically underachieving, to allow for further multivariate analyses.

- To more fully capture the classroom context of learning and instruction, NAEP should consider exploratory studies of the relationship between teacher and student surveys that target classroom interactions.
- NAEP's current interest in special studies related to the design of culturally responsive and culturally relevant assessments could be joined with studies of culturally relevant opportunities to learn in the disciplines. Such couplings of special studies could contribute to parsimonious use of intellectual and fiscal resources and address important issues of equity.

Concluding Thoughts

We are confident that the NAEP program can improve its coverage of important OTL issues by expanding its range of contextual questions regarding how students and teachers experience and engage in discipline-specific classroom instruction. These data will not only help educators and policymakers interpret NAEP achievement results but also provide a way to monitor, across time, the learning opportunities provided to students from diverse backgrounds. Accordingly, it is worthwhile to explore the design of special studies within the existing NAEP program or in collaboration with the NAEP program that explore more carefully OTL experiences. Such attention to OTL and related research is very much needed to appropriately recognize and take up the challenges of educating underserved racially, culturally, and linguistically diverse educational communities whose assets and promise are prominent in our public consciousness.

It is important that any follow-on recommendations from this paper taking up the development and use of new OTL items proposed for contextual questionnaires as part of the formal NAEP program adhere to item performance demands and content coverage restrictions required by federal law and Office of Management and Budget clearance procedures. That said, we hope that far more useful contextual information indices, consistent with a sociocultural perspective on teaching, learning, and achievement, can be created and implemented in the NAEP program. We also believe that measures of progress in OTL, and their relationships to academic progress as measured by NAEP's assessments, should become a far more prominent part of NAEP if it is to truly inform the nation about educational progress, especially for those historically most disadvantaged by society.

REFERENCES

- Aspen Institute Education and Society Program. (2022). *Opportunity to learn, responsibility to lead*. <https://www.aspeninstitute.org/publications/opportunity-to-learn-responsibility-to-lead/>
- Bennett, R. E. (2023). Toward a theory of socioculturally responsive assessment. *Educational Assessment*, 28(2), 83–104. <https://doi.org/10.1080/10627197.2023.2202312>
- Carter, P. L., & Welner, K. G. (Eds.). (2013). *Closing the opportunity gap: What America must do to give every child an even chance*. Oxford University Press.
- Cohen, J., Ruzek, E., & Sandilos, L. (2018). Does teaching quality cross subjects? Exploring consistency in elementary teacher practice across subjects. *AERA Open*, 4(3). <https://doi.org/10.1177/2332858418794492>
- Corcoran, T. B., Mosher, F. A., & Rogat, A. (2009). *Learning progressions in science: An evidence-based approach to reform* (CPRE Research Report # RR-63). Consortium for Policy Research in Education. <https://files.eric.ed.gov/fulltext/ED506730.pdf>
- Daro, P., Mosher, F. A., & Corcoran, T. B. (with Barrett, J., Battista, M., Clements, D., Confrey, J., Daro, V., Maloney, A., Nagakura, W., Petit, M., & Sarama, J.). (2011). *Learning trajectories in mathematics: A foundation for standards, curriculum, assessment, and instruction* (CPRE Research Report # RR-68). Consortium for Policy Research in Education. <https://files.eric.ed.gov/fulltext/ED519792.pdf>
- Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, 61, Article 101859. <https://doi.org/10.1016/j.cedpsych.2020.101859>
- Education First. (2022). “There is a bigger story going on here.” How are states using broader measures of student success to address education disparities? <https://www.education-first.com/wp-content/uploads/2022/09/Ed-First-Equity-Measures-Landscape-Overview-June-2022.pdf>
- Elmore, R. F., & Rothman, R. (Eds.). (1999). *Testing, teaching, and learning: A guide for states and school districts*. National Academies Press. <https://files.eric.ed.gov/fulltext/ED447172.pdf>
- Engeström, Y. (2001). Expansive learning at work: Toward an activity theoretical reconceptualization. *Journal of Education and Work*, 14(1), 133–156. <https://doi.org/10.1080/13639080123238>
- Floden, R. E. (2002). The measurement of opportunity to learn. In National Research Council, *Methodological advances in cross-national surveys of educational achievement* (pp. 231–266). The National Academies Press. <https://doi.org/10.17226/10322>

- Goals : Educate America Act of 1994, Pub. L. No. 103-227. 108 Stat. 125 (1994).
<https://www.congress.gov/103/statute/STATUTE-108/STATUTE-108-Pg125.pdf>
- Goldman, S. R., & Lee, C. D. (2024). Human learning and development: Theoretical perspectives to inform assessment systems. In S. Marion, J. W. Pellegrino, & A. Berman (Eds.), *Reimagining balanced assessment systems* (pp. 48–92). National Academy of Education. https://naeducation.org/wp-content/uploads/2024/04/Full-Report_Reimagining-Balanced-Assessment-Systems.pdf
- Goodman, M. J., & Walton, A. (2023, June 1). New NCES tool explores how differential learning opportunities may influence achievement gaps. *NAEPPlus+*.
https://nces.ed.gov/nationsreportcard/blog/achievement_gaps_tool.aspx
- Greeno, J. G. (2006). Learning in activity. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 79–96). Cambridge University Press.
- Grossman, P., Cohen, J., Ronfeldt, M., & Brown, L. (2014). The test matters: The relationship between classroom observation scores and teacher value added on multiple types of assessment. *Educational Researcher*, 43(6), 293–303.
<https://doi.org/10.3102/0013189X14544542>
- Gutiérrez, K. D., & Rogoff, B. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher*, 32(5), 19–25.
<https://doi.org/10.3102/0013189X032005019>
- Hammond, Z. (2021). On equity and student engagement. *ASCD*, 79(4).
<https://www.ascd.org/el/articles/zaretta-hammond-on-equity-and-student-engagement>
- Herman, J. L., Klein, D. C. D., & Abedi, J. (2000). Assessing students' opportunity to learn: Teacher and student perspectives. *Educational Measurement, Issues and Practice*, 19(4), 16–24.
<https://doi.org/10.1111/j.1745-3992.2000.tb00042.x>
- Lampert, M., Franke, M. L., Kazemi, E., Ghouseini, H., Turrou, A. C., Beasley, H., Cunard, A., & Crowe, K. (2013). Keeping it complex: Using rehearsals to support novice teacher learning of ambitious teaching. *Journal of Teacher Education*, 64(3), 226–243.
<https://doi.org/10.1177/0022487112473837>
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Harvard University Press.
- Lee, C. D. (1995). A culturally based cognitive apprenticeship: Teaching African American high school students skills in literary interpretation. *Reading Research Quarterly*, 608–630.
- Marion, S. (2020). *Using opportunity-to-learn data to support educational equity*. National Center for the Improvement of Educational Assessment.
<https://files.eric.ed.gov/fulltext/ED611325.pdf>

- McDonnell, L. M. (1995). Opportunity to learn as a research concept and a policy instrument. *Educational Evaluation and Policy Analysis*, 17(3), 305–322.
<https://doi.org/10.3102/01623737017003305>
- Moll, L. C., Amanti, C., Neff, D., & Gonzalez, N. (1992). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. *Theory Into Practice*, 31(2), 132–141. <https://doi.org/10.1080/00405849209543534>
- Moss, P. A., Pullin, D. C., Gee, J. P., Haertel, E. H., & Young, L. J. (2008). *Assessment, equity, and opportunity to learn*. Cambridge University Press.
- Mullis, I. V. S., & Martin, M. O. (Eds.). (2019). *PIRLS 2021 assessment frameworks*.
<https://timssandpirls.bc.edu/pirls2021/frameworks/>
- Mullis, I. V. S., von Davier, M., Foy, P., Fishbein, B., Reynolds, K. A., & Wry, E. (2023). *PIRLS 2021 international results in reading*. Boston College, TIMSS & PIRLS International Study Center. <https://doi.org/10.6017/lse.tpisc.tr2103.kb5342>
- National Education Policy Center (2023). Educational accountability 3.0: Beyond ESSA. Beyond Test Scores Project and National Education Policy Center.
<https://nepc.colorado.edu/publication/resource-document-beyond-essa>
- NAEP Mathematics Report Card. (n.d.).
https://www.nationsreportcard.gov/math_2017/student-experiences?grade=4
- Nasir, N. S., Rosebery, A., Warren, B., & Lee, C. D. (2014). Learning as a cultural process: Achieving equity through diversity. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (2nd ed., pp. 686–706). Cambridge University Press.
- National Academies of Sciences, Engineering, and Medicine. (2018). *How People Learn II: Learners, Contexts, and Cultures*. The National Academies Press.
<https://doi.org/10.17226/24783>
- National Academies of Sciences, Engineering, and Medicine. (2019). *Monitoring educational equity*. The National Academies Press. <https://doi.org/10.17226/25389>
- National Academies of Sciences, Engineering, and Medicine. 2020. *Building Educational Equity Indicator Systems: A Guidebook for States and School Districts*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/25833>
- National Academy of Education. (2021). *Educational assessments in the COVID-19 era and beyond*. <https://files.eric.ed.gov/fulltext/ED618967.pdf>
- National Assessment Governing Board. (2012). *NAEP background questions and the use of contextual data in NAEP reporting* [Policy Statement].
<https://www.nagb.gov/content/dam/nagb/en/documents/policies/collection-report-backg-data.pdf>

- National Assessment Governing Board. (2013). *Contextual information framework for the National Assessment of Educational Progress*.
<https://www.nagb.gov/content/dam/nagb/en/documents/publications/frameworks/contextual-information/contextual-information-framework.pdf>
- National Assessment Governing Board (2021a). Mathematics framework for the 2026 National Assessment of Educational Progress.
<https://www.nagb.gov/content/dam/nagb/en/documents/publications/frameworks/mathematics/2026-math-frameowork/NAEP-2026-Mathematics-Framework-Combined.pdf>
- National Assessment Governing Board. (2021b). *Reading framework for the 2026 National Assessment of Educational Progress*.
<https://www.nagb.gov/content/dam/nagb/en/documents/publications/frameworks/reading/2026-reading-framework/naep-2026-reading-framework.pdf>
- National Association for Music Education. (2020). *National opportunity to learn standards*.
<https://nafme.org/wp-content/uploads/2020/08/NAfME-Opportunity-to-Learn-Standards-2020.pdf>
- National Council of Teachers of English. (2019). *Position statement on the opportunity to learn*.
<https://ncte.org/statement/opptolearnstandards/>
- National Council on Education Standards and Testing. (1992). *Raising standards for American education: A report to Congress, the Secretary of Education, the National Education Goals Panel, and the American people*. Department of Education.
<https://files.eric.ed.gov/fulltext/ED338721.pdf>
- National Education Policy Center. (2023). *Educational accountability 3.0: Beyond ESSA*. Beyond Test Scores Project and National Education Policy Center.
<https://nepc.colorado.edu/publication/resource-document-beyond-essa>
- National Research Council. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. The National Academies Press. <https://doi.org/10.17226/9853>
- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. The National Academies Press. <https://doi.org/10.17226/13165>
- National Research Council. (2018). *How people learn II: Learners, contexts, and cultures*. The National Academies Press. <https://doi.org/10.17226/24783>
- NGSS Lead States. (2013). *Next generation science standards: For states, by states*. The National Academies Press. <https://doi.org/10.17226/18290>
- Organisation for Economic Co-operation and Development. (2014). *PISA 2012 technical report*. OECD Publishing. <https://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf>

- Organisation for Economic Co-operation and Development. (2019). *PISA 2018 assessment and analytical framework*. OECD Publishing. <https://doi.org/10.1787/b25efab8-en>
- Organisation for Economic Co-operation and Development. (2023a). *PISA 2022 assessment and analytical framework*. OECD Publishing. <https://doi.org/10.1787/dfe0bf9c-en>
- Organisation for Economic Co-operation and Development. (2023b). *PISA 2022 results (Volume II): Learning during – and from – disruption*. OECD Publishing. <http://doi.org/10.1787/a97db61c-en>
- Pullin, D. (2008). Assessment, equity, and opportunity to learn. In P. A. Moss, D. C. Pullin, J. P. Gee, E. H. Haertel, & L. J. H. Young (Eds.), *Assessment, equity, and opportunity to learn* (pp. 333–352). Cambridge University Press. <https://doi.org/10.1017/CBO9780511802157>
- Rothman, R. (1993). Researchers seek to measure ‘opportunity’ standards. *Education Week*. <https://www.edweek.org/education/researchers-seek-to-measure-opportunity-standards/1993/04>
- Ruiz-Primo, M.-A., & Furtak, E. M. (2024). Classroom activity systems to support ambitious instruction and assessment. In S. Marion, J. W. Pellegrino, & A. Berman (Eds.), *Reimagining balanced assessment systems* (pp. 93–131). National Academy of Education. https://naeducation.org/wp-content/uploads/2024/04/Full-Report_Reimagining-Balanced-Assessment-Systems.pdf
- Shepard, L. A. (2021). Ambitious teaching and equitable assessment: A vision for prioritizing learning, not testing. *American Educator*, 45(3), 28–37. <https://files.eric.ed.gov/fulltext/EJ1321974.pdf>
- Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational Measurement: Issues and Practice*, 37(1), 21–34. <https://doi.org/10.1111/emip.12189>
- Smith, J. B., Lee, V. E., & Newmann, F. M. (2001). *Achievement in Chicago Elementary Schools* (p. 52). Consortium on Chicago School Research.
- Wang, F., Wang, Y., Liu, Y., & Leung, S. O. (2022). Investigating the measurement of OTL in PISA 2012 and its relationship with self-efficacy and mathematics achievement: Doubly latent multilevel analyses. *Scandinavian Journal of Educational Research*, 63(6), 837–852. <https://doi.org/10.1080/00313831.2022.2070929>
- White House. (2021, January 20). *Executive order on advancing racial equity and support for underserved communities through the federal government* [Executive Order 13985]. <https://www.govinfo.gov/content/pkg/FR-2021-01-25/pdf/2021-01753.pdf>
- Windschitl, M., Thompson, J., & Braaten, M. (2018). *Ambitious science teaching*. Harvard University Press.