Student Assessment

Chris Domaleski and Scott Marion
K-12 Schools
Standards and Accountability

Introduction

Student assessment, particularly state- and district-required assessment, has been increasingly prominent in U.S. education over the past 30 years but is not new. There has been assessment, often informal, since people began teaching one another. There are three general classes of assessment. It might be better termed "formative instruction" because it is a process used during instruction to provide feedback to students and teachers to improve learning.

Summative assessments, like the formal tests that most are familiar with, are administered after some period of instruction and used to evaluate how well students learned what was intended. Summative assessments are used at the student level and often in the aggregate at the classroom, school, district, and state levels. Summative assessments take many forms: performance tasks administered by teachers after a unit of instruction and the now ubiquitous large-scale standardized tests based largely on selected-response (e.g., multiple choice) test items, but many include some constructed response items (e.g., writing). Large-scale summative assessments are the main form used as policy instruments at the state and federal levels.

Interim assessments have occupied a niche in the educational testing landscape for more than 40 years. They are administered multiple times during the school year to evaluate students' knowledge and skills and inform educators and leaders' classroom, school, or district decisions. The most common interim assessments, offered by several testing companies, consist almost exclusively of selected-response questions and take the same form regardless of where they are administered.

The topic of student assessment is broad. It also includes a growing interest in assessing students' competence in important life skills, often termed 21st century skills, such as persistence, self-regulation, and critical thinking.

Validity is the most important criterion for judging the quality of any assessment. Importantly, validity is not an inherent feature of a test but instead tied to purpose and use. A test may have a high degree of validity for one use but little to none for another. Valid score interpretations also require that students have a fair opportunity to show what they know. Assessments should be as accessible and fair to all students as possible. We focus below on issues that policymakers are likely to encounter and have a well-developed research base.

Several topics we highlight below, particularly formative assessment, interim tests, and "beyond-academic" assessments, are often subsumed under discussions of balanced assessment systems. While there is significant writing about balanced assessment systems[1], little empirical evidence supports or refutes the effects of balanced assessment systems on student learning. This is largely due to the limited clarity on what exactly constitutes a balanced assessment system and the lack of high-quality models in practice. Therefore, we do not discuss balanced assessment systems as a separate topic, focusing instead on selected attributes. However, a well-designed, balanced assessment system is necessary to support a range of purposes and uses. No single assessment can serve more than one purpose well. This concept runs through many of our findings.

Key findings

Key finding #1: *Formative assessment, when well-implemented, has a documented and substantial influence on student learning and achievement.*

Formative assessment is a planned, ongoing process teachers use in learning contexts to elicit and use evidence of student learning to improve student understanding of intended learning goals[2]. Formative assessment is the one type of assessment that is documented to noticeably improve student learning[3]. Unfortunately, it is hard to do well. Formative assessment relies on teachers with excellent pedagogical knowledge and skills in schools with learning cultures and structures to facilitate collaborative practices.

Key finding #2: *Interim assessments, particularly those offered by commercial enterprises, are ubiquitous, but independent research indicates little to know documented relationship between these tests and student achievement.*

Interim assessments are administered multiple times during the school year to evaluate students' knowledge and skills and purportedly provide information to educators and education leaders to inform decisions at the classroom, school, or district level. Most school districts in the U.S. use at least one commercial interim assessment product. Yet, despite their widespread use, there is a paucity of evidence to support the myriad claims regarding their purported benefit to students and schools. On the other hand, many have documented the negative effects, such as over-testing and deskilling educators due to interim assessment use.

Key finding #3: *Large-scale assessments can be designed to be more accessible for all students, including students with disabilities and English language learners.*

Drawing on the principles of Universal Design for Learning (UDL), among other advances, assessments have been made more inclusive and accessible to all learners[4]. Employing these strategies has produced more valid measures of what students know and can do, especially for special education students and English learners. These approaches must be explicitly incorporated throughout the test design, development, and implementation phases to be most effective. Moreover, the effectiveness of these accessibility supports must be studied throughout implementation to ensure they are working as intended and to adjust as necessary.

Key finding #4: *Durable or 21st century skills are essential for success in school and life but are challenging to measure well, particularly in large-scale settings.*

Experts agree that a range of skills is crucial for success in various post-secondary pursuits, including college, career, and citizenship[5]. Critical thinking, self-regulation, collaboration, creativity, and complex communication are common examples of these complex skills. There is a compelling case for supporting the development of local instruction and assessment systems to foster the development of 21st century skills. On the other hand, using large-scale, standardized assessments to assess these complex domains is challenging and carries a substantial risk of unintended negative consequences.

Background

The main policy drivers behind expanded testing requirements are associated with various reauthorizations of the Elementary and Secondary Education Act (ESEA), starting with the Improving America's Schools Act (IASA) in 1994 and especially with the No Child Left Behind (NCLB) Act of 2001.

There is a long history of using tests to make consequential decisions for students. These policies have proliferated in the 21st century for both students and schools. Accountability initiatives, such as requiring students to pass a high-school graduation test or as a "promotional gate" such as a 3rd grade reading requirement, have become more common. As part of school accountability systems, school-level consequences intensified in the United States following the 2001 passage of NCLB. The evidence associated with consequential uses of large-scale assessments for students and schools (or other entities) is addressed in this chapter.

Once-per-grade span statewide summative achievement testing of mathematics and reading/English language arts was first mandated by IASA, which was then increased to grades 3-8 and once in high school under NCLB. States were required to assess science annually at least once per grade span under NCLB. Essentially, the same assessment requirements were carried over to the Every Student Succeeds Act (ESSA). NCLB also required the implementation of high-stakes school accountability systems, which increased the pressure to perform well on state summative assessments. School and district leaders responded to this pressure in many ways, one of which was relying on district-required interim assessments to hopefully help improve student learning. The proliferation of both statewide and interim assessments often crowded out classroom and formative assessments.

Subsequently, we discuss the extensive research literature associated with formative and interim assessments. We also summarize the research associated with efforts to improve the fairness and accessibility of large-scale assessments and the growing interest in trying to assess 21st century skills. It is important to recognize the challenges associated with researching the effects of assessment policies and practices. First, in educational research, it is extremely rare to conduct the type of experiments necessary to support causal inference, which requires randomly selecting participants from a population and randomly assigning participants (e.g., teachers or students) to the "treatment" or "control" conditions. For example, it would be challenging for a state to randomly assign a subset of students, classrooms, schools, or some other unit to test all students in grades 3–8 while randomly assigning another subset not to have to test.

Perhaps more important, assessments are rarely the intervention or treatment. They are enmeshed in a complex context of other aspects of the education system that have considerable implications for how one judges the efficacy of the assessment initiative. For example, there has been a recent expansion of assessments for young students to screen for dyslexia and to provide additional information about students developing early reading skills. However, these testing programs are one component of a much larger initiative, including updated reading programs, intense professional development for teachers, summer school for struggling readers, and grade retention policies. Disentangling the effects of the assessment component is almost impossible, which is one reason why we do not discuss the efficacy of K–3 reading assessments in this chapter.

Evidence

Key Finding #1: *Formative assessment, when well-implemented, has a documented and substantial influence on student learning and achievement.*

Formative assessment is the one type of assessment that is documented to noticeably improve student learning. The consensus definition of formative assessment was updated in 2018 by a Council of Chief State School Officers working group as "…a planned, ongoing process used by all students and teachers during learning and teaching to elicit and use evidence of student learning to improve student understanding of intended disciplinary learning outcomes and support students to become self-directed learners[6]."

Black and Wiliam[7] first tried in 1998 to quantify the influence of formative assessment on student outcomes by conducting a large-scale research synthesis. They reported an average effect size of approximately 0.4, which is relatively large for an educational intervention. This means that the average score of the group experiencing formative assessment was at about the 65th percentile of the distribution of scores of the group not experiencing formative assessment.

Subsequent meta-analyses[8] using more rigorous methods than those applied by Black and Wiliam found slightly lower but still meaningful effect sizes ranging from approximately 0.25 to 0.35 standard deviation units, depending on the subject area and grade span.

Several researchers have examined the features of the formative assessment process most associated with improved achievement. Hattie and Timperley (2007)[9] noted that the quality, timing, and specificity of the feedback to students was critical to the influence on student learning and achievement. As part of their extensive meta-analyses, Lee and colleagues (2020) identified the features of formative assessment most associated with the overall positive effects. Unsurprisingly, supporting students' capacity for self-assessment had more influence than any other feature they identified. Cases where interventions were planned compared to unplanned (or not enacted) and where teachers received ongoing professional development also had strong positive effects.

The major challenge associated with formative assessment is that it is hard to do well. Formative assessment relies on well-trained and supported teachers working in schools with learning cultures and structures to facilitate collaborative practices. Shepard noted, "…ambitious teaching practices, framed by sociocultural theory, are essentially one and the same as equitable [and formative] assessment practices[10]." Ambitious teaching practices require considerable expertise and experience to implement well. Furtak and colleagues[11] carried out an extensive intervention study. They conceptualized formative assessment as consisting of four main dimensions: designing formative assessment tasks, asking questions to elicit student thinking, interpreting student ideas, and providing feedback that moves student thinking forward. After three years of intensive efforts, they were able to help participating teachers significantly improve their skills on all four dimensions.

Many other studies have been shown to improve teachers' formative assessment knowledge and skills. In general, it requires multiple years of consistent professional learning opportunities, considerable support (often from outside experts), and structures within schools such as common planning time for educators to collaboratively learn from one another. Given the turnover among leaders and teachers and the financial challenges associated with providing the structural support necessary to sustain high-quality formative assessment practices, it is not surprising that it is not more widespread.

Does this mean that policymakers and education leaders should give up on formative assessment? Many have turned to interim assessments to reduce the effort required to implement formative assessment at scale. However, as documented below, interim assessment has not been shown to consistently improve student learning, so if educational leaders are looking to use assessment to improve learning and teaching in meaningful ways, formative assessment is the best hope. Therefore, leaders should appropriately support ongoing professional learning and create the structures necessary to implement and sustain ongoing formative assessment practices.

Key finding #2: *Interim assessments, particularly those offered by commercial enterprises, are ubiquitous, but independent research indicates little to no documented relationship between these tests and student achievement.*

Interim assessments are administered multiple times during the school year to evaluate students' knowledge and skills and provide information to educators and education leaders to inform classroom, school, or district decisions. Interim assessments are used for a variety of intended purposes[12]. Since the passage of NCLB, they have proliferated based on claims that they could help school personnel identify

students at risk of not meeting key performance thresholds to support their achievement. It is hard to find reliable information on the number of districts in the U.S. using interim assessments. However, the CEO of one of the leading interim assessment companies[13] recently told one of us that approximately 92% of school districts in the U.S. use at least one commercial interim assessment product. Despite their widespread use, there is little evidence to support the myriad claims regarding their purported benefit to students and schools[14].

One well-known, independent, large-scale, randomized controlled study used data from Indiana to study the effects on student achievement after a randomly assigned set of districts implemented interim assessments. The researchers found no benefit from using interim assessments in reading or mathematics, grades 3–8, compared to control schools, but they found significant negative effects for students in early elementary school[15].

In another large-scale experiment, researchers at Johns Hopkins[16] conducted a study involving their interim assessments and other data-driven intervention products. The results of these studies were generally not significant after the first two years. However, limited significant improvements were found for some grades and subjects in years 3 and 4 after a variety of programmatic improvements were implemented by year 3. The complete set of interventions led to increased student achievement, not just the interim assessments.

Perhaps the most important set of studies emerged from a Consortium on Policy Research in Education (CPRE) project[17]. The CPRE studies focused on district-required "benchmark" assessment, a type of interim assessment administered on six-week cycles. These benchmarks were related to each district's curriculum, making them more instructionally useful than curriculum-agnostic commercial interim assessments. They reported that teachers used the information from the interim tests to focus on remediating procedural or similar lower-level skills rather than gaining substantive insights about student learning. This occurred partly because the benchmarks were constructed only of selected-response items, making it impossible for teachers to observe student thinking. In the CPRE studies, teachers could see the actual test items and student responses (correct/incorrect), which is not true for commercial interim assessments. They used this information to group students for re-teaching based on low performance on specific items. However, this did not translate to noticeable improvements on the state summative assessment or other important outcome measures[18].

On the other hand, interim assessments have been helpful as outcome measures to study the effects of a particular intervention or event (e.g., COVID-19 disruptions) on student achievement. Depending on the type of study, the ability to track student performance throughout the school year and across school years and their widespread use may be a positive case for interim assessments[19]. That said, many of the claims made by test publishers about interim assessments are about instructional uses, and those claims must be supported by a rigorous research base for policymakers to continue to support such uses.

Key finding #3: *Large-scale assessments can be designed to be more accessible for all students, including students with disabilities and English language learners.*

Test developers and practitioners must create opportunities for a wide range of learners, including students with disabilities (SWDs) and English language learners (ELs), to show what they know and can do on educational assessments. One established approach is to provide effective accessibility and accommodation features. Common features generally relate to setting, timing, presentation, and response; examples appear in Table 1.[20]

Table 1. Examples of accessibility and accommodation features

Category	Examples
Setting	small group or individual administration, separate testing location
Timing	extended testing time, flexible scheduling, frequent breaks
Presentation	Braille or large print, clarify directions, oral reading, translation
Response	Use of scribe, respond in native language, calculator

A seminal report by the National Research Council (NRC) reviewed multiple studies on the effects of accommodations on test performance for ELs and SWDs.[21] Many studies included in the NRC's review were based on an experimental design in which the focal student group and a comparison group were administered tests with and without various accommodations. The results were evaluated to determine if there were interactions or a differential boost, which demonstrated that score gains under accommodated conditions for SWDs exceeded the gains of the comparison groups. One seminal study found significant gains for SWDs randomly assigned to receive a read-aloud accommodation on a fourth-grade mathematics test compared to a control group.[22] Similar patterns were observed for studies that compared accommodations for ELs to native English speakers. For example, another key study examined performance on linguistically simplified mathematics word problems, revealing greater score increases for EL students over a comparison group.[23]

Results from the studies reviewed in the NRC report and elsewhere in the literature are inconsistent. Research indicates that multiple factors can influence the degree to which accommodations are effective and appropriate, such as the background factors of the examinee, the nature of the accommodation, and the characteristics of the assessed construct. Sireci et al. (2005) reviewed 28 empirical studies of accommodations and found that most studies showed gains for SWDs that exceeded those of general education students.[24] The finding was relatively consistent for extended testing time, but the effectiveness of many other accommodations varied. The authors called for more research focused on how the accommodation is thought to function in instruction and assessment and the degree to which accommodated administrations' results support the assessment's intended interpretations and uses. Such findings underscore the need to address the accessibility early and throughout the development process.

UDL has been used as a framework for making assessments more accessible and fairer. The key principles of UDL are inclusive population, precisely defined constructs, accessible items, amenable to accommodations, clear instructions and procedures, and maximum readability and legibility.[25] Subsequent scholarship has examined how UDL can be refined and improved for assessment and how these features impact student performance. For example, CAST emphasizes the importance of providing multiple means of engagement, representation, and expression to provide access and opportunities appropriate to every learner.[26] A National Center on Educational Outcomes report reviewed 76 resources on universal design for assessment from 1985–2023 and found that the application of UDL principles has been uneven.[27] For example, there is often a lack of clarity about the guiding framework and the design elements implemented. For this reason, evidence regarding the efficacy of UDL is limited. The study surfaced the need for design and development initiatives that go beyond accessibility features and accommodations and that serve a wide range of students, including SWDs, ELs, ELs with disabilities, and students with significant cognitive disabilities. Such improvements will help the field better understand the efficacy of UDL applications for different learners.

Contemporary research is expanding on how technology can better support accessibility for all learners. For example, customized accessibility features appropriate to each learner can be embedded in items and forms.[28] The most recent Question and Test Interoperability standards (QTI3) include a standardized mechanism to address examinees' personal needs and preferences (PNP).[29] These advancements help ensure that the tools and support appropriate for each test taker are available and activated for each test administration.

Key finding #4: *Durable or 21st century skills are essential for success in school and life but are challenging to measure well, particularly in large-scale settings.*

In a rapidly changing world where artificial intelligence, among other advances, is changing how work is accomplished, the need for distinctly 'human' skills beyond traditional academic or cognitive domains is elevated. Experts across multiple fields emphasize the importance of perseverance, collaboration, critical thinking, self-regulation, and complex communication.[30] [31] [32]

Research shows a relationship between these broader skills and various educational, career, and health outcomes[33]. A meta-analysis by Durlak and colleagues reviewed 213 school-based social emotional learning (SEL) programs involving 270,034 K-12 students and found evidence of improved social and emotional skills, attitudes, behaviors, and academic performance[34]. Another meta-analysis by Cipriano and colleagues synthesized 424 studies from 53 countries focusing on SEL interventions. While there was noteworthy variation in the results, the researchers concluded that participants "experienced significantly improved skills, attitudes, behaviors, school climate and safety, peer relationships, school functioning, and academic achievement[35]."

While widely acknowledged as necessary, these skills can be difficult to identify, much less measure well, particularly in large-scale, standardized settings. The wide range of terms and frameworks used contributes to this challenge. Harvard's Ecological Approaches to Social Emotional Learning (EASEL) laboratory created a taxonomy project that compares over 40 SEL frameworks addressing over 200 terms associated with key skills[36].

Regardless of the terms, there is a compelling case for supporting the development of local instruction and assessment systems to foster the development of these skills. Drawing on findings from 12 independent meta-analyses of school-based programs, Greenberg (2023) identified four elements needed to implement high-quality SEL programs:

- Sequenced: They involve a developmentally coordinated set of activities
- Active: Active learning helps students master new skills
- Focused: Programs intentionally develop personal and social skills
- Explicit: The specific skills are identified, taught, and practiced[37]

On the other hand, using large-scale, standardized assessments to assess these complex skills is quite challenging.[38] Evans et al. (2020) noted that barriers include poorly defined constructs, limited understanding of how skills develop, interactions among skills, and the appropriateness of separating skills from context.[39] The authors also cite concerns about the sufficiency of the evidence to generalize about performance and the limited knowledge about cultural validity and equity. Simply put, more evidence is needed to develop direct measures of 21st century skills that are reliable, valid, fair, and comparable for all learners.

Accordingly, many scholars have cautioned against using measures of SEL or related constructs for consequential purposes, such as by including outcomes as components of student, educator, or school accountability systems.[40] For this reason, the most promising measurement practices for 21st century skills will likely be classroom-based grounded in local contexts and used for formative purposes.

Endnotes and References

[1] National Research Council. 2001. Knowing What Students Know: The Science and Design of Educational Assessment; National Research Council. 2006. Systems for State Science Assessment. Washington, DC: The National Academies Press; National Research Council. 2014. Developing Assessments for the Next Generation Science Standards. Committee on Developing Assessments of Science Proficiency in K-12. Board on Testing and Assessment and Board on Science Education, James W. Pellegrino, Mark R. Wilson, Judith A. Koenig, and Alexandra S. Beatty, Editors. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press; Marion, S. F., J. W. Pellegrino, and A. I. Berman (Eds). 2024. Reimagining Balanced Assessment Systems. Washington, DC: National Academy of Education.

[2] Council of Chief State School Officers. 2018. Revising the definition of formative assessment (p. 2).

[3] Black, P., and D. Wiliam. 1998. Assessment and classroom learning. Assessment in Education: Principles, Policy & Practice 5(1): 7–73; Kingston, N., and B. Nash. 2011. Formative assessment: A meta-analysis and a call for research. Educational Measurement: Issues and Practice 30(4): 28–37; Lee, H., H. Q. Chung, Y. Zhang, J. Abedi, and M. Warschauer. 2020. The Effectiveness and Features of Formative Assessment in US K-12 Education: A Systematic Review. Applied Measurement in Education 33(2): 124–140.

[4] Thompson, S., C. J. Johnstone, and M. L. Thurlow. 2002. Universal design applied to large scale assessments (Synthesis Report 44). University of Minnesota, National Center on Educational Outcomes.

[5] Cipriano, C., M. J. Strambler, L. H. Naples, C. Ha, M. Kirk, M. Wood, K. Sehgal, A. K. Zieher, A. Eveleigh, M. McCarthy, M. Funaro, A. Ponnock, J. C. Chow, and J. Durlak. 2023. The state of evidence for social and emotional learning: A contemporary meta-analysis of universal school-based SEL interventions. Child Development 94: 1181–1204.

[6] Council of Chief State School Officers. 2018. Revising the definition of formative assessment (p. 2).

[7] Black and Wiliam (1998).

[8] Kingston and Nash (2011); Lee et al. (2020).

[9] Hattie, J., and H. Timperley. 2007. The power of feedback. Review of Educational Research 77(1): 81–112.

[10] Shepard, L. A. 2021. Ambitions teaching and equitable assessment: A vision for prioritizing learning not testing. American Educator 6(7): 1–33.

[11] Furtak, E. M., K. Kiemer, R. K. Circi, R. Swanson, V. de León, D. Morrison, and S. C. Heredia. 2016. Teachers' formative assessment abilities and their relationship to student learning: Findings from a four-year intervention study. Instructional Science 44(3): 267–291.

[12] Perie, M., S. Marion, and B. Gong. 2009. Moving toward a comprehensive assessment system: A framework for considering interim assessments. Educational Measurement: Issues and Practice 28(3): 5–13.

[13] Personal communication with Chris Minich, CEO of NWEA, on June 24, 2024.

[14] See for example the home pages of the three major interim assessment providers: NWEA, Renaissance Learning, and Curriculum Associates.

[15] Konstantopoulos, S., S. R. Miller, A. van der Ploeg, and W. Li. 2016. Effects of interim assessments on student achievement: Evidence from a large-scale experiment. Journal of Research on Educational Effectiveness 9(sup1): 188–208.

[16] Slavin, R. E., A. Cheung, G. Holmes, N. A. Madden, and A. Chamberlain. 2013. Effects of a data-driven district reform model on state assessment outcomes. American Educational Research Journal 50(2): 371–396.

[17] Goertz, M. E., L. N. Oláh , and M. Riggan. 2009. Can Interim Assessments Be Used for Instructional Change? Policy Briefs RB-51, Consortium for Policy Research in Education; Oláh, L. N., N. R. Lawrence, and M. Riggan. 2010. Learning to learn from benchmark assessment data: How teachers analyze results. Peabody Journal of Education 85(2): 226–245.

[18] Oláh, Lawrence, and Riggan (2010).

[19] See, for example, Lewis, K. and M. Kuhfeld. 2023. Education's long COVID: 2022–23 achievement data reveal stalled progress toward pandemic recovery. NWEA.

[20] National Research Council. 2004. Keeping Score for All: The Effects of Inclusion and Accommodation Policies on Large-Scale Educational Assessments. Committee on Participation of English Language Learners and Students with Disabilities in NAEP and Other Large-Scale Assessments. Koenig, Judith A., and Lyle F. Bachman, Editors. Board on Testing and Assessment, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

[21] National Academies of Sciences, Engineering, and Medicine. 2002. Reporting Test Results for Students with Disabilities and English-Language Learners: Summary of a Workshop. Washington, DC: The National Academies Press.

[22] Tindal, G., B. Heath, K. Hollenbeck, P. Almond, and M. Harniss. 1998. Accommodating students with disabilities on large-scale tests: An experimental study. Exceptional Children 64(IV): 439–450.

[23] Abedi, J., and C. Lord. 2001. The language factors in mathematics tests. Applied Measurement in Education 14(3): 219–234.

[24] Sireci, S. G., S. E. Scarpati, and S. Li. 2005. Test accommodations for students with disabilities: An analysis of the interaction hypothesis. Review of Educational Research 75(4): 457–490.

[25] Thompson, Johnstone, and Thurlow (2002).

[26] CAST. 2018. Universal Design for Learning Guidelines version 2.2.

[27] Liu, K. K., M. L. Thurlow, M. Quanbeck, J. A. Bowman, and A. Riegelman. 2024. Universal design and K-12 academic assessments: A scoping review of the literature (NCEO Report 442). National Center on Educational Outcomes.

[28] Russell, M. 2018. Recent Advances in the Accessibility of Digitally Delivered Educational Assessments. In: Elliott, S., Kettler, R., Beddow, P., Kurz, A. (eds) Handbook of Accessible Instruction and Testing Practices. Springer, Cham.

[29] 1EdTech Consortium. 2022. Question and Test Interoperability Standards 3.0.

[30] Duckworth, A. 2016. Grit: The Power of Passion and Perseverance. Scribner.

[31] Dweck, C. S. 2006. Mindset: The New Psychology of Success. Random House.

[32] Dwyer, C.P., A. Boswell, and M. A. Elliott. 2015. An evaluation of critical thinking competencies in business settings. Journal of Education for Business 90(5): 260–269.

[33] National Research Council. (2012). Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century. Committee on Defining Deeper Learning and 21st Century Skills, Pellegrino, J. W., and M.L. Hilton, Editors. Board on Testing and Assessment and Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

[34] Durlak, J.A., R. P. Weissberg, A. B. Dymnicki, R. D. Taylor, and K. B. Schellinger. 2011. The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions. Child Development 82: 405–432.

[35] Cipriano et al. (2023).

[36] Harvard Graduate School of Education Ecological Approaches to Social Emotional Learning (EASEL). n.d. Navigate the Complex Field of Social Emotional Learning.

[37] Greenberg, M. T. 2023. Evidence for social and emotional learning in schools. Learning Policy Institute.

[38] McKown, Clark. 2015. Challenges and Opportunities in the Direct Assessment of Children's Social Emotional Comprehension. In Handbook of Social and Emotional Learning, ed. Joseph A. Durlak. New York: Guilford Press: 320–35.

[39] Evans, C., J. Thompson, and C. Brandt. 2020. Instructing and Assessing 21st Century Skills: Key Measurement and Assessment Considerations. National Center for the Improvement of Educational Assessment.

[40] Melnick, H., C. M. Cook-Harvey, and L. Darling-Hammond. 2017. Encouraging social and emotional learning in the context of new accountability (brief). Learning Policy Institute.

Suggested citation

Scott Marion and Chris Domaleski (2025). "Student Assessment," in Live Handbook of Education Policy Research, in Douglas Harris (ed.), Association for Education Finance and Policy, viewed MM/DD/YYYY, <WEB LINK>.