Scholedge International Journal of *Multidisciplinary & Allied Studies*
ISSN 2394-336X Vol. 12, Issue 02 (2025)
DOI: https://dx.doi.org/10.19085/sijmas120201

# The Future of Assessment: Rethinking Evaluation in an AI-Assisted Learning Environment

Minghui Gu
East China Normal University, Shanghai, China

**ABSTRACT**

Assessment has long served as the cornerstone of K-12 education, shaping how students learn, how teachers teach, and how systems are held accountable. The arrival of artificial intelligence in classrooms has not simply added a new tool to an old system; it has exposed the limits of that system in ways that can no longer be ignored. This paper examines how AI is changing the conditions under which students learn and, consequently, how evaluation must change to remain meaningful. Drawing on established frameworks in assessment theory and recent policy developments, the paper argues that the dominant model of standardized, summative testing is poorly suited to an environment where students have growing access to AI-assisted support. It presents four national cases (India, China, the United States, and Canada) to show how different educational systems are responding to the challenge, each at a different stage and with a different set of pressures. The paper identifies equity, data privacy, teacher preparedness, and algorithmic accountability as the four most pressing concerns in transitioning to AI-compatible assessment practices. It closes with a set of practical directions for policymakers, school administrators, and curriculum designers who want to build assessment systems that are both rigorous and relevant in the years ahead.

**Keywords:** K-12 assessment, artificial intelligence in education, formative assessment, competency-based evaluation, educational technology, assessment reform

## INTRODUCTION

Every few decades, something comes along that forces educators to ask a basic question: what exactly are we trying to measure? The printing press changed what knowledge looked like. The internet changed who could access it. AI is now changing what it means to produce it. When a student can use a generative AI tool to draft an essay, solve a multi-step math problem, or summarize a dense historical text, the traditional test, built on the assumption that knowledge production is an individual, timed, pen-to-paper act, starts to look like a measurement of the wrong thing.

This is not a small problem for K-12 education. Across the world, formal schooling still relies heavily on end-of-term examinations, standardized tests, and grades that are supposed to tell students, parents, employers, and universities something meaningful about what a young person knows and can do. But the conditions under which students now learn have shifted considerably, and the tools available to them are more powerful than anything a previous generation of test designers had to contend with.

The question this paper takes up is not whether AI will change assessment, since that much seems certain. The more useful question is how assessment should change, and what principles should guide that change. The goal is not to argue for abolishing formal evaluation. Structure, accountability, and measurement have their place. The goal is to ask what those things should look like when the learning environment itself has been fundamentally altered.

The paper focuses on K-12 education for a specific reason: this is where habits of learning are formed, where foundational skills are built, and where assessment has the most lasting effect on a student's relationship with knowledge. What gets tested in school shapes what students believe is worth knowing. That is a responsibility that deserves serious rethinking in light of what AI now makes possible.

Four national systems are considered as cases: India, China, the United States, and Canada. These are not presented as exemplars or cautionary tales but as different responses to a common pressure. Each illuminates something the others do not. Table 1 provides a comparative summary of assessment frameworks and reform trajectories across all four countries.

**Table 1:** Comparative Overview of K-12 Assessment Frameworks and AI Integration Across Four Countries

| Country | Primary Assessment Model | Recent Reforms | AI Integration Status | Key Challenge |
|---|---|---|---|---|
| India | Board examinations (Class 10 and 12); National Achievement Survey (NAS) | NEP 2020; NIPUN Bharat (2021); CBSE competency-based questions from 2020 | Early stage; significant urban-rural divide in infrastructure | Gap between policy ambition and on-the-ground implementation; unequal access |
| China | Gaokao (National College Entrance Examination); school-level internal assessments | Double Reduction Policy (2021); suzhi jiaoyu curriculum reforms | Significant government investment; AI monitoring tools deployed in urban schools | Examination culture resistant to structural change; surveillance concerns |
| United States | State standardized tests under ESSA (2015) framework; district-level assessments | CBE in select states; portfolio-based assessment in some schools | Widespread adaptive platform use; AI writing feedback tools in select districts | Decentralized system; highly uneven resource distribution and implementation |
| Canada (esp. BC) | Provincial assessments; curriculum-embedded formative evaluation | BC redesigned curriculum (2015); competency-focused frameworks in several provinces | Early to moderate; growing platform use in urban centers | Significant provincial variation; rural-urban digital access divide |

*Note.* NEP = National Education Policy; CBSE = Central Board of Secondary Education; ESSA = Every Student Succeeds Act; BC = British Columbia; CBE = competency-based education. AI integration status reflects the period 2020–2024.

## THE WEIGHT OF THE EXISTING SYSTEM

To understand why reform is so difficult, it helps to understand how deeply the current assessment model is embedded in educational systems. The dominant paradigm in K-12 assessment has been what Shepard (2000) called the "testing and accountability" model, a model built around discrete, measurable, standardized outcomes that can be compared across students, schools, districts, and nations. This model served its purposes. It created benchmarks, helped identify achievement gaps, and gave systems a shared vocabulary for talking about performance.

But it also built its entire logic around one particular assumption: that what matters can be captured in a time-limited, individual, written response to a pre-set question. Black and Wiliam (1998), in a review of over 250 research studies, found that formative assessment, the ongoing process of gathering and using evidence about learning during instruction rather than at its end, produced some of the largest gains in student achievement seen in any educational intervention. Yet most school systems continued to organize around summative tests: the board exam, the standardized national assessment, the end-of-year grade.

Pellegrino, Chudowsky, and Glaser (2001) made the point clearly in their report for the National Academies. Drawing on cognitive science research, they argued that assessment, cognition, and instruction form an interconnected triangle, and that when one element shifts, the others must respond. The model of cognition embedded in most standardized tests sees a student as a passive container of facts. That model was already outdated when their report, *Knowing What Students Know*, was published. It is more outdated still today.

Part of what makes the existing system so resistant to change is the function it serves beyond pedagogy. Stiggins (2002) observed that most formal assessment in schools exists to sort and rank students rather than to support their learning. Grades, rank lists, competitive examinations, and university entrance scores are not primarily teaching tools; they are gatekeeping tools. That function does not disappear because AI has arrived. If anything, it becomes more complicated, because AI makes it harder to trust individual student performance on tasks designed under older assumptions.

The persistence of the examination model is also, in part, a social and political phenomenon. A percentage score or a rank can be communicated across administrative levels in ways that a portfolio of learning evidence or a teacher's professional judgment often cannot. When reformers propose moving away from standardized testing, they often underestimate how much institutional trust is invested in the idea of the objective test, even when the evidence for its actual objectivity is limited.

**WHAT AI HAS CHANGED**

Describing AI as a single phenomenon would be a mistake. In K-12 education, AI appears in several different forms, each with different implications for assessment.

Adaptive learning systems, platforms that adjust the difficulty and sequence of content based on student responses, have been in use for at least two decades. These tools generate rich behavioral data about how a student learns, where they struggle, and what kinds of scaffolding help them move forward. That data is, in a meaningful sense, assessment data, continuous, contextual, and far more granular than anything a standardized test produces. The problem is that this data has not been well integrated into formal assessment structures in most school systems.

Automated essay scoring (AES) systems have been in development since the 1960s but have grown considerably more sophisticated with advances in natural language processing. Shute (2008), in a review of formative feedback research, identified automated feedback systems as among the more promising tools for scaling the kind of responsive, specific feedback that good teachers provide but cannot always deliver at the pace students need. The quality of AES systems varies considerably, and their tendency to reward surface features of writing over deeper argumentation and genuine reasoning has been a persistent concern among researchers.

Generative AI tools, particularly large language models available to students through widely accessible consumer products, represent the most recent and most disruptive shift. A student with access to a generative AI tool can now produce coherent writing, solve reasoning problems step by step, or get a detailed explanation of any concept taught in K-12 curricula within seconds. Holmes, Bialik, and Fadel (2019) argued in *Artificial Intelligence in Education* that AI promises to personalize learning at a scale

previously impossible and to generate new accountability questions that existing frameworks are not equipped to handle. The arrival of consumer-grade generative AI has made both of those observations considerably more urgent.

UNESCO's (2021) *AI and Education: Guidance for Policy-Makers* called on governments to rethink assessment frameworks alongside the introduction of AI tools in schools, not after the fact. The document pointed to a gap that has grown wider since: the speed of AI adoption in classrooms is significantly outpacing the speed of assessment reform in most national systems.

Luckin, Holmes, Griffiths, and Forcier (2016), in *Intelligence Unleashed*, drew a distinction worth preserving in this discussion: the difference between AI that supports a teacher's professional judgment and AI that replaces it. In assessment, this is a particularly consequential difference. An AI tool that helps a teacher identify which students have not yet grasped a concept, or that flags when a student's performance pattern suggests a need for intervention, is a different instrument from one that issues a grade or makes a high-stakes decision about a student's future without human review.

What the sum of these developments points to is a structural mismatch. The assessment systems used in K-12 education around the world were designed for a learning environment where knowledge was scarce, where producing a correct answer required effort, and where the ability to retrieve and present information under controlled conditions was a reasonable proxy for understanding. None of those assumptions holds in the same way today.

**RETHINKING THE FRAMEWORK**

What would an assessment system designed for an AI-assisted learning environment look like? Useful answers already exist in the research literature. They predate AI but AI makes them both more achievable and more necessary than before.

The most influential framework is the distinction between assessment of learning and assessment for learning. Sadler (1989) laid out the theoretical basis for what would later be widely adopted under the label of formative assessment: the idea that feedback is only educationally useful when a student can act on it. A grade at the end of a term tells a student how they have been ranked. It rarely tells them what to do differently. Formative feedback, delivered during the learning process, closes the gap between what a student currently understands and what they are working to understand.

This is precisely the kind of feedback that AI systems can, in principle, deliver at scale. An adaptive platform can detect a misconception in a student's reasoning and provide a targeted explanation within seconds. A language model, given appropriate context and constraints, can point to specific weaknesses in a piece of writing while explaining why those weaknesses matter. The challenge is that these capabilities exist largely outside formal assessment structures. They happen in the platform, not in the gradebook. Assessment reform needs to close that gap.

Anderson and Krathwohl (2001), revising Bloom's (1956) original taxonomy of educational objectives, described learning as a spectrum from remembering facts at one end to creating at the other. The taxonomy's upper levels, analyzing, evaluating, and creating, involve the kind of thinking that AI currently cannot replicate with real depth or reliability. This matters for assessment design. Tests that measure primarily recall and basic application are, in an AI-assisted environment, largely testing whether a student produced their own work rather than whether they learned anything meaningful. Assessment design needs to move toward tasks that require demonstrated reasoning at the higher levels of the taxonomy.

Gardner's (1983) *Frames of Mind* drew attention to the narrowness of assessment systems that recognize only the logical-linguistic intelligence rewarded by conventional standardized tests. An assessment

framework that is broader, one that recognizes the range of ways students can demonstrate understanding, fits more naturally into an AI-assisted environment where students follow different learning paths and develop different strengths.

Hattie (2009), synthesizing over 800 meta-analyses of educational influences in *Visible Learning*, identified formative assessment and teacher feedback as consistently among the highest-impact factors in student achievement. His findings also showed that student self-assessment and peer assessment, when structured carefully, produced substantial gains. Both practices fit naturally into an AI-assisted environment where students have tools to check their own understanding and receive feedback outside classroom hours. The implication is not to remove teacher judgment but to design environments that activate students as agents in their own learning, something AI can facilitate when the broader framework supports it.

The common thread running through these frameworks is a shift in who assessment is primarily for. The existing model is built primarily for institutions: it produces the data that systems need to sort, certify, and compare. A reformed model would be built primarily for learners: it would produce the feedback that students need to keep developing. AI, used well, makes the latter more achievable than at any previous point in the history of formal schooling.

## NATIONAL CASES

The following four cases show how different K-12 systems are responding to the challenge of assessment in an AI-assisted learning environment. Each system faces the same general pressure but brings different institutional starting points to the problem.

### INDIA

India's National Education Policy (NEP) 2020 represents the most ambitious formal rethinking of school assessment in the country since independence. The NEP explicitly critiques the existing culture of high-stakes examinations and calls for a shift toward holistic assessment that captures not just academic performance but skills, competencies, and personal development (Ministry of Education, Government of India, 2020). The document proposes replacing the model of high-pressure board examinations, taken at the end of Class 10 and Class 12, with modular, repeatable assessments designed to reduce the burden of a single high-stakes test.

The Central Board of Secondary Education (CBSE) began introducing competency-based questions into its board examinations from 2020 onward, questions designed to test application and analysis rather than the reproduction of memorized content. The NIPUN Bharat initiative, launched in 2021, introduced assessment frameworks for Grades 1 through 5 focused on specific, observable learning outcomes in foundational literacy and numeracy. Both represent a meaningful conceptual shift from examination-centered accountability toward outcome-based, classroom-embedded evaluation.

At the same time, India faces structural challenges that complicate any vision of AI-assisted assessment. The country operates one of the world's largest school systems, serving hundreds of millions of students across a vast diversity of infrastructure, language, and socioeconomic contexts. Digital access remains deeply unequal: urban private schools have increasingly integrated technology into instruction, but rural government schools in many states lack the basic infrastructure to support it reliably. A framework that works well in a school in Delhi or Bengaluru may be entirely inapplicable in a government school in rural Bihar or Odisha.

The NEP's vision is progressive in its direction. But the distance between policy intention and implementation in India's education system is historically significant, and the institutional infrastructure,

including teacher training, assessment design capacity, and monitoring, needed to support the shift is not yet in place at scale. India illustrates, perhaps more clearly than any of the other cases here, that the challenge of assessment reform is not primarily technological. It is a governance challenge.

**CHINA**

China's K-12 assessment system has long been defined by one instrument: the *gaokao*, the National College Entrance Examination taken by students at the end of high school. The *gaokao* carries enormous weight in determining university admission, and its influence reaches back through the entire school experience, shaping curriculum, instruction, and student behavior from early secondary school onward. Dello-Iacovo (2009) documented how China's "quality education" (*suzhi jiaoyu*) movement, which gained momentum in the 1990s, represented an official attempt to shift away from test-centered schooling toward broader educational development. That movement influenced curriculum design but largely failed to dislodge the examination culture, because the *gaokao* itself has not changed structurally enough to make test preparation a less rational strategy for students and families.

In 2021, the Chinese government introduced the "Double Reduction" policy (*shuang jian*), which restricted homework hours assigned to students in compulsory education and banned for-profit academic tutoring for core subjects in Grades 1 through 9. The policy was a direct response to the intense academic pressure created by the exam-driven system. Its practical effects have been more complicated: many families shifted to informal tutoring arrangements, and demand for test preparation did not disappear so much as migrate to different channels.

China has also made substantial investment in AI in education at the national level. The 2017 New Generation AI Development Plan explicitly identified education as a sector for AI deployment, and AI tools claiming to assess student performance, monitor engagement, and personalize instruction have been piloted in Chinese schools. UNESCO (2021) highlighted the use of AI behavioral monitoring systems in some Chinese classrooms as examples that raise serious questions about surveillance, consent, and the appropriate boundaries of assessment data.

The Chinese case illustrates a tension visible in different forms across all four countries: AI can generate far more data about students than existing assessment systems know how to use constructively. When that data is harnessed to intensify the same test-preparation logic that already dominates schooling, it adds pressure without changing purpose. Whether China can develop AI-assisted tools that work against the grain of *gaokao*-driven anxiety rather than with it remains a central policy question.

**UNITED STATES**

The United States operates a decentralized education system in which curriculum and assessment decisions are primarily made at the state and district level, with federal legislation setting broad accountability requirements. The Every Student Succeeds Act (ESSA) of 2015 shifted some authority back to states after the more prescriptive requirements of No Child Left Behind, giving states greater flexibility to design their own accountability systems, including how they measure and report student performance.

Within this framework, several states and districts have been experimenting with assessment models that go beyond standardized testing. Competency-based education, in which students demonstrate mastery of specific skills before advancing, has been adopted as formal state policy in a small but growing number of states. Portfolio-based assessment has been used in some schools as a complement to or partial replacement for standardized examinations.

AI has entered U.S. K-12 classrooms through multiple channels. Adaptive learning platforms are widely used across grade levels, and several districts have piloted AI-assisted writing feedback tools.

Darling-Hammond and Adamson (2013) argued, before generative AI became a consumer reality, that assessments of "deeper learning" require fundamentally different designs than standardized tests, and that the costs of developing those designs are more than offset by the educational and social benefits. That argument has become more pressing now that AI makes it easy for students to produce work that satisfies the surface criteria of traditional assessments without necessarily doing the thinking those assessments were built to evaluate.

The United States also illustrates the limits of decentralization in assessment reform. Because education policy is highly distributed across fifty states and thousands of districts, innovations that work in one context do not automatically travel. A competency-based assessment approach that succeeds in a small New England state may have almost no influence on how students are assessed across the country. This limits the pace at which any system-wide change can happen.

**CANADA**

Canada, like the United States, has a decentralized education system in which provinces hold primary authority over schooling. This has produced meaningful variation in how assessment is conceptualized and practiced across the country, with some provinces considerably more advanced than others in reforming their approaches.

British Columbia's redesigned curriculum, introduced in 2015 and implemented progressively through the following years, represents one of the more thorough attempts among developed education systems to move K-12 learning frameworks away from content coverage and toward competency development. The BC curriculum organizes learning around intellectual, personal, and social competencies alongside content knowledge. Assessment in this framework is expected to be ongoing, holistic, and focused on the development of the whole student (British Columbia Ministry of Education, 2015).

BC's approach draws on principles that Vygotsky (1978) articulated in *Mind in Society*: that development is a social process, and that the zone of proximal development, the gap between what a learner can do independently and what they can do with support, is where productive learning happens. In an AI-assisted learning environment, this has practical implications. If AI is a legitimate form of support in learning, analogous to a more experienced peer or a patient tutor, then measuring a student's ability to perform without any support at all may not be measuring the right thing.

Ontario has also implemented student assessment frameworks that emphasize growth over time and include self-assessment as a formal component of the learning process. The province's curriculum documents require teachers to gather assessment evidence from multiple sources and to use that evidence to adjust instruction, a description that maps closely onto what Black and Wiliam (1998) identified as the core of effective formative assessment.

Canada's AI integration in classrooms is still in relatively early stages, with significant variation across provinces and between urban and rural schools. But the policy groundwork in provinces like British Columbia, which already emphasizes competency, process, and teacher professional judgment over standardized test scores, may make the transition to AI-informed assessment frameworks less disruptive than in systems built more firmly around examination accountability.

**CHALLENGES AND CONCERNS**

Any serious discussion of AI-assisted assessment must reckon with the problems that come with it. Four deserve direct attention.

**Equity and Access**

The most immediate concern is that AI-assisted learning tools are not distributed equally. Schools in wealthier areas, in all four countries discussed here, are better positioned to integrate technology, provide teacher training, and support students in using AI tools appropriately. If AI-assisted assessment becomes standard practice while access to AI remains uneven, assessment systems risk measuring not just what students have learned but what resources their schools and families could access. AI, without deliberate policy intervention, has the potential to widen existing gaps significantly.

**Data Privacy**

AI-assisted assessment generates large volumes of data about individual students: how long they spent on a task, where they made errors, what kinds of hints they needed, how their performance changed over time. This data can be educationally useful, but it also creates risks. Questions about who owns student data, how it can be used, how long it is retained, and what protections exist against misuse are not yet settled in most jurisdictions. UNESCO's (2021) *AI and Education* called for international frameworks to govern the use of AI-generated data in educational contexts, but national-level regulation in this area remains fragmented and has not kept pace with the speed of adoption in most countries.

**Teacher Preparedness**

Assessment is a professional practice, not just an administrative requirement. Teachers need to understand what they are measuring, why, and how to interpret results in ways that actually help students learn. The introduction of AI tools into assessment creates new professional demands: teachers must evaluate AI-generated feedback critically, explain to students when AI output is and is not reliable, and make the judgment calls that no algorithm can fully substitute for. Investment in professional development, not just technical training but genuine learning about assessment theory and practice, is a prerequisite for any successful transition. Most school systems have not yet made that investment in any systematic way.

**Algorithmic Accountability**

AI systems that contribute to assessment decisions are not neutral. They are trained on data that reflects historical patterns, including the biases embedded in that data. An automated essay scoring system trained primarily on writing from well-resourced schools may systematically undervalue writing that reflects different cultural and linguistic traditions. A recommendation algorithm that predicts which students are "at risk" may encode the same demographic assumptions that have historically driven inequitable outcomes. Selwyn (2019), in *Should Robots Replace Teachers?*, examined several AI applications in education and found that data-driven systems can reproduce and amplify existing inequities when their design and training data are not carefully scrutinized. Transparency about how AI assessment tools make decisions, and clear mechanisms for contesting those decisions, need to be built into implementation frameworks before adoption, not added afterward.

**A Way Forward**

The direction that emerges from the evidence and the cases reviewed here is not a binary choice between traditional testing and AI-generated assessment. It is a considered hybrid: a system that uses the strengths of AI, continuous data generation, personalized feedback, and the ability to track growth over time, while preserving what requires human judgment, the interpretation of context, the recognition of genuine intellectual effort, and the capacity to have a real conversation with a student about what they understand and why.

Several practical directions follow from this.

The assessment systems should be redesigned around learning rather than ranking. This means investing in formative assessment processes, regular and low-stakes feedback that students can act on, and treating summative assessments as one data point among several rather than the definitive verdict on what a student is capable of. The frameworks developed by Black and Wiliam (1998) and Sadler (1989) provide a well-evidenced foundation for this kind of redesign.

AI-assisted tools should enter the assessment process as support for teachers, not as substitutes for professional judgment. A platform that helps a teacher see which students have not yet grasped a concept is a useful tool. A platform that issues a grade without teacher review is something altogether different and more problematic.

Assessment frameworks should include tasks that AI cannot assist with in ways that undermine the assessment's purpose. This does not mean returning to timed, closed-book examinations as the only valid format. It means designing tasks that require students to demonstrate understanding through authentic reasoning: oral presentations with follow-up questions, collaborative projects with individual reflection components, or iterative writing with documented revision processes where the evidence of learning is visible in the process as much as in the product.

The access must be treated as a prerequisite rather than an afterthought. Policymakers who want to introduce AI-assisted learning and assessment tools need to address infrastructure, connectivity, and device access before expecting those tools to produce equitable outcomes. A framework designed for well-resourced schools and adopted without adaptation elsewhere will reproduce the same inequitable results as the existing model.

Student data governance frameworks need to be developed and enforced. Schools and districts that use AI tools in assessment should be required to disclose what data those tools collect, how it is used, how long it is retained, and how students and families can access or contest it. This is a regulatory matter, and it will require regulatory action.

## CONCLUSION

Assessment has always been a reflection of what a society believes education is for. When the dominant purpose is sorting students for a hierarchical labor market, high-stakes standardized testing makes a kind of institutional sense. When the purpose shifts toward developing capable, thinking, adaptable human beings, the assessment system has to shift with it.

AI has not created that shift. But it has made it harder to avoid. Students who have access to powerful AI tools are not learning in the same environment their teachers were trained to teach in, and the assessment systems designed for that older environment are increasingly measuring something that is difficult to define with confidence.

The four countries examined here are each responding to this in their own way. India has articulated a progressive reform vision that has not yet translated into widespread practice. China has invested heavily in AI tools within an examination culture that has so far proven more durable than reformers hoped. The United States has the policy flexibility to experiment but lacks the coordination to bring successful experiments to scale. Canada, particularly British Columbia, has made the most systematic structural changes to curriculum and assessment philosophy, offering a practical model worth examining, though AI integration there too remains uneven.

What these cases share is recognition at the policy level that the existing assessment model cannot remain unchanged indefinitely. What they also share is the difficulty of changing systems with deep

institutional roots, connected to university admissions and labor market signals, and requiring not just new tools but genuinely different values about what schooling is for.

The technology is available. The frameworks exist in the research literature. The case for reform is well supported by evidence. What AI has done is take a question that educators could once treat as theoretical and make it one that now demands a practical answer.

**REFERENCES**


Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. Longmans Green.

British Columbia Ministry of Education. (2015). *BC's new curriculum*. Government of British Columbia. https://curriculum.gov.bc.ca

Darling-Hammond, L., & Adamson, F. (2013). *Developing assessments of deeper learning: The costs and benefits of using tests that help students learn*. Stanford Center for Opportunity Policy in Education.

Dello-Iacovo, B. (2009). Curriculum reform and 'quality education' in China: An overview. *International Journal of Educational Development*, *29*(3), 241–249. https://doi.org/10.1016/j.ijedudev.2008.02.008

Every Student Succeeds Act, Pub. L. No. 114-95, 129 Stat. 1802 (2015). https://www.congress.gov/114/plaws/publ95/PLAW-114publ95.pdf

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. Basic Books.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson.

Ministry of Education, Government of India. (2020). *National Education Policy 2020*. Government of India. https://www.education.gov.in/sites/upload_files/mhrd/files/NEP_Final_English_0.pdf

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. National Academies Press. https://doi.org/10.17226/10019

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119–144. https://doi.org/10.1007/BF00117714

Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity Press.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, *29*(7), 4–14. https://doi.org/10.3102/0013189X029007004

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, *83*(10), 758–765. https://doi.org/10.1177/003172170208301010

UNESCO. (2019). *Beijing consensus on artificial intelligence and education*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000368303

UNESCO. (2021). *AI and education: Guidance for policy-makers*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000376709

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.