



HSCred

Research Brief

HSCred white paper available separately.

Key Findings

AI broke the college application. The take-home written artifact, long the closest thing institutions had to a window on student thinking, no longer carries reliable evidentiary weight now that generative AI can produce an application essay on demand, and the answer cannot be better cheating detection. The deeper problem the essay's collapse exposes is that admissions has no trustworthy common standard for academic work, nothing that lets an institution compare students across schools.

- **Written authenticity has collapsed.** Generative AI produces fluent, on-prompt application essays at no cost, and admissions officers regard this as a serious problem, with half viewing the use of AI in admissions essays unfavorably and only fourteen percent favorably (Kaplan, 2025).
- **Institutions are deciding at record volume with no trusted common signal.** First-year applications passed ten million for the first time in 2024–25 and have risen 161 percent in a decade, accelerating after colleges dropped testing requirements, even as the offices reading them have not grown to match (Common App, 2025). Many have turned to artificial intelligence just to keep pace (Knox, 2023), and a senior official at the National Association for College Admission Counseling describes a “hard administrative wall,” with neither the people power nor the time to adopt anything radically different (Hawkins, as cited in Guha et al., 2018). Having gone test-optional, several of the most selective universities reversed course and reinstated the SAT, not because they trust it but because nothing else they collected carried common meaning (MIT Admissions, 2022; Dartmouth College, 2024).
- **Detection cannot fix it.** Institutions still relying on the college essay are hoping they can fight AI with AI. However, leading detectors are biased against non-native English writers (Liang et al., 2023), the maker of ChatGPT withdrew its own detector for low accuracy (OpenAI, 2023), and detector accuracy collapses under simple paraphrasing (Perkins et al., 2024).
- **Supervised, defended work is what survives.** A credential whose provenance is established through supervision and live revision, rather than inferred after submission, is the one generative AI cannot counterfeit (Dawson, 2021; Dawson et al., 2024; Eaton, 2023).
- **Richer evidence predicts at least as well, and it scales.** Selection rewards predictive validity, not the tidiness of a label, and a single test score is a limited, income-correlated predictor (Geiser & Santelices, 2007), while high-school performance predicts college completion at least as well (Allensworth & Clark, 2020), and a composite of multiple validated measures predicts readiness better than any single test (Gaertner & McClarty, 2015;

Sternberg & the Rainbow Project Collaborators, 2006). Because each institution can weight validated work into its own ranking, this evidence reaches admissions as a sortable signal rather than fifty thousand portfolios to watch.

- **Performance assessment predicts success, across many programs.** The evidence spans the City University of New York and the New York Performance Standards Consortium (Fine & Pryiomka, 2020), New Hampshire’s PACE system (Evans, 2019; Perez & Evans, 2023), a multistate deeper-learning study (Zeiser et al., 2014), and the International Baccalaureate’s supervised, defended extended essay (Inkelas et al., 2013).
- **Policy and demand are converging.** Students for Fair Admissions v. Harvard (2023) created a need for race-neutral talent identification, more than two thousand colleges are test-optional or test-free (FairTest, 2025), New York is removing the Regents exams as a standalone graduation requirement, with statewide testing continuing under federal law, while adopting a Portrait of a Graduate framework (New York State Education Department, 2025a, 2025b), and federal law explicitly authorizes performance-based assessment (Every Student Succeeds Act, 20 U.S.C. § 6364; Troppe et al., 2023).

The Missing Assessment

American admissions lacks a common unit of demonstrated academic work that retains a stable meaning across high schools. This is the structural deficiency beneath a set of problems usually treated separately.

Standardized scores are comparable across applicants but track socioeconomic advantage as much as capability. Family income exerts a substantial influence on test performance, largest for low-income and Black students (Dixon-Román et al., 2013), and the score gap between high- and low-income students reflects inequalities that begin years before the exam (Chetty et al., 2023; Georgetown University, 2023). Decades of measurement research warn, moreover, that scores under high-stakes accountability tend to inflate and to narrow what schools teach, so a number can rise without the underlying learning improving (Koretz, 2017). Grade point averages are not comparable across schools of differing rigor, and average high-school grades have risen for a decade even as test performance stagnated, a divergence documented by the test maker ACT itself (ACT, 2022). Personal essays and portfolios favor students with access to coaching, and the shift to test-optional admissions removed a common signal without supplying a replacement, obliging institutions to infer capability from noisier proxies (Syverson et al., 2018). The result is a measurement problem before it is an equity problem: institutions decide without a shared, verifiable record of what an applicant can actually do, and decision quality degrades accordingly (Carter & Welner, 2013).

These weaknesses now meet a second, compounding pressure: the sheer scale of the task. As early as 2018, David Hawkins, a senior policy official at the National Association for College Admission Counseling, named the constraint plainly, observing that admissions officers run up against a “hard administrative wall” because “the volume of applications is so large at so many institutions that there simply isn’t the people power or the time to implement something that is radically different from what

we see now” (as cited in Guha et al., 2018). Every pressure since has pushed in the same direction. First-year applications through the Common App passed ten million for the first time in 2024–25, an increase of eight percent in a single year and of 161 percent over the decade, with the steepest climb arriving after 2019–20 as institutions suspended their testing requirements (Common App, 2025). Students now file a record number of applications each, the staff who must read them have not grown in proportion, and many offices have turned to artificial intelligence simply to manage the load (Knox, 2023). An admissions system is thus asked to sort more students than ever, on weaker evidence than ever, with no more hands to do the reading. The result is a high-stakes sorting carried out, of necessity, in the dark.

Nothing exposes the gap more sharply than the testing reversals themselves. As of fall 2026, more than nine in ten ranked four-year colleges do not require the SAT or ACT (FairTest, 2025), even as a small group of highly selective institutions has reinstated a testing requirement, beginning with the Massachusetts Institute of Technology in 2022 (MIT Admissions, 2022; Wren, 2022) and continuing in 2024 with Dartmouth, Yale, Brown, and Harvard (Dartmouth College, 2024; Yale University, 2024; Brown University, 2024; Harvard University, 2024). Their reasoning deserves to be engaged rather than dismissed, because it is more careful than the headlines suggest: none of these institutions claimed the test is unbiased, and several conceded plainly that scores track income and access, resting their case instead on the narrower points that a score read against the resources of a student’s own school still predicted later performance and that a test-optional process had been overlooking high-achieving applicants from under-resourced schools (Dartmouth College, 2024; Brown University, 2024; Harvard University, 2024). That is an argument about the absence of a better signal, not a defense of the number itself; MIT said as much directly, noting that it would adopt a more equitable instrument as soon as a sufficient one existed (Wren, 2022). Having studied their own data, these institutions returned to an admittedly imperfect, income-correlated measure because nothing else they collected carried comparable meaning from one school to the next, which is exactly the deficiency described here. The standardized score persists by default, as the only common and auditable record the system now has, and the question it forces is not whether to keep or discard it but what trustworthy common standard could carry the weight a single number cannot.

The same deficiency is visible from the opposite direction in the schools. Just as selective colleges have moved back toward a number they concede is flawed, states are now mandating the kind of evidence that number cannot supply: New York is phasing out the Regents examinations as a graduation requirement and adopting a Portrait of a Graduate framework built on demonstrated competencies (New York State Education Department, 2025a, 2025b), without yet having a common, auditable way to certify such work across districts. One end of the system is retreating to an imperfect score while the other is being told to produce performance evidence it has no shared instrument to record. The field is reaching in both directions at once because a trusted common standard does not yet exist.

What admissions requires, and what no conventional metric supplies, is a record of authentic academic work that is at once comparable across contexts and resistant to the resource advantages that distort traditional measures. That record is the common standard this brief describes.

The Authenticity Collapse

The personal essay was the part of an application closest to the student's own mind. Generative AI has removed its evidentiary value, and the loss is not reparable by detection.

The capability is now universal and free. A student can produce a fluent, well-structured, on-prompt essay in seconds, and the people who read those essays know it. The 2025 Kaplan survey found that admissions officers overwhelmingly distrust AI-assisted essays, with half viewing the practice unfavorably and only fourteen percent favorably (Kaplan, 2025). The instinctive response, to catch and punish AI writing, fails on its own terms. Detectors are biased against non-native English writers, flagging much of their authentic prose as machine-generated (Liang et al., 2023). The maker of ChatGPT released and then withdrew its own classifier because its accuracy was too low to be useful (OpenAI, 2023). And detector accuracy collapses under paraphrasing so simple that any motivated student can apply it, which led the researchers who ran those tests to advise against using detectors to determine academic-integrity violations at all (Perkins et al., 2024).

The more useful frame is validity rather than honesty. An assessment is valid when it measures what it claims to measure, and a take-home essay no longer measures the student's own thinking when the writing may not be the student's (Dawson et al., 2024). The collapse is not only an admissions problem but a measurement one. In a 2026 framework for assessment validation, Tim Fawns, David Boud, and Phillip Dawson treat validity as an ongoing judgment about whether the available evidence supports a claim about a student's learning, and they distinguish four proxies for that evidence: product, process, performance, and practice. A take-home written artifact supplies only the first, a single "product" proxy that on its own gives only a "partial and often opaque picture of learning," and the one most exposed to generative AI. It deepens that weakness, they argue, letting students "complete product-oriented tasks . . . without meaningful engagement with the intended learning outcomes" (Fawns et al., 2026). Hybrid human and machine authorship makes the separation of human from machine text in take-home work impractical in principle, not merely difficult in practice (Eaton, 2023). The discipline that replaces detection is assessment security: establishing the provenance of work through supervision and authentication rather than inferring it after the fact (Dawson, 2021).

This is why supervised, expert-validated, live-defended work is the credential that survives. When a student plans, drafts, revises, and defends a piece of work in front of expert evaluators, the work's authorship is established as it is produced. The artifact carries its own provenance. That property, and not any claim about the moral character of students, is what makes such work durable in conditions that dissolve the take-home essay.

A Quality Measure of Academic Achievement

The opportunity is therefore institutional rather than merely individual. The need is not one more way for a single student to stand out. It is shared infrastructure: a comparable, auditable, portable record of authentic academic work that many institutions can trust at once. This is the function that financial markets obtained when they acquired a common, auditable standard for reporting value, and that academic achievement still lacks.

The professional bodies that govern academic records are already moving toward such a standard, which signals demand rather than novelty. The Comprehensive Learner Record initiative of the American Association of Collegiate Registrars and Admissions Officers, with NASPA, defines richer, verifiable, portable records of student achievement (AACRAO & NASPA, 2018). The IEdTech Comprehensive Learner Record Standard makes those records machine-verifiable and interoperable across institutions (IEdTech Consortium, n.d.). The National Institute for Learning Outcomes Assessment argues that institutions should make evidence of student accomplishment transparent and comparable (National Institute for Learning Outcomes Assessment, n.d.). What these efforts standardize is the container. What they do not supply, and what remains the binding constraint, is a trustworthy and scalable way to validate the work that goes inside it.

That validated unit of work is what this brief calls Academic Capital: a piece of authentic student work that remains auditable to the underlying artifact and to the named experts who evaluated it. Unlike a standardized score, which compresses a student into a number whose derivation cannot be inspected, a unit of Academic Capital lets an institution weigh evidence directly rather than infer capability from an opaque proxy.

The unit is worth measuring because the economy that receives these students has changed. Half a century ago Herbert Simon observed that an abundance of information produces a scarcity of attention, since a wealth of information creates a poverty of attention (Simon, 1971), the insight that named what is now called the attention economy (Goldhaber, 1997). In a labor market reshaped by computing, routine tasks that follow explicit rules are increasingly automated, while the work computers complement rather than replace is non-routine problem-solving and complex communication (Autor, Levy, & Murnane, 2003), and demand has shifted further toward roles that combine analytical and social skill (Deming, 2017). What these trends reward is the capacity to focus attention, investigate, synthesize, and produce work that others find worth attending to. That is precisely what a unit of Academic Capital captures and what a standardized score does not.

Evidence, Not Labels

The reframe rests on a measurement claim that an academic reader will recognize, and it draws a line that is often blurred in practice. Performance-based assessment asks what a student can make, explain, defend, and revise, and it preserves the work. Competency-based and mastery-based systems often ask raters to infer, from that same work, whether a student has satisfied an item on a common checklist, and they preserve the label. The first produces evidence. The second produces a judgment whose reliability depends on the very cross-school comparability that the evidence shows is hard to achieve.

The asymmetry is well documented. The more ambitious a claim, the harder it is to validate, and a verdict of mastery is a more ambitious claim than the presentation of the work that supposedly demonstrates it (Kane, 2013). High-inference labels do not travel well between institutions. When complex student portfolios were scored across many schools, inter-rater agreement was too low to support comparison (Koretz et al., 1994). A century of grading research finds that locally assigned grades blend achievement with unrelated factors and carry no stable meaning from one school to another (Brookhart et al., 2016). Evaluations of competency-based pilots have been heterogeneous

and explicitly non-causal (Steele et al., 2014). There is, at present, no peer-reviewed evidence that mastery or competency transcripts improve the admissions signal, and the measurement literature implies they worsen comparability unless the underlying evidence travels with them.

The objection a psychometrician will raise is that selection does not aim to maximize information but to maximize prediction, and that a sparse signal can still predict well. That is the right test, and authentic evidence meets it. What a selection system should ask of any input is whether it improves the prediction of who will succeed, and on that question the single test score is weaker than its prominence suggests. High-school performance predicts college completion at least as well as admissions tests, and far more consistently from one school to the next (Allensworth & Clark, 2020; Geiser & Santelices, 2007), and it does so while carrying little of the family-income signal that test scores absorb, since admissions-test scores correlate with family income and parental education at roughly 0.25 to 0.4 while high-school grades correlate at close to zero (Geiser & Santelices, 2007). Richer still, a composite of multiple academic and behavioral indicators predicts readiness markedly better than any single measure (Gaertner & McClarty, 2015). In the Rainbow Project, adding validated assessments of analytical, creative, and practical performance to the SAT and the high-school record improved prediction while narrowing the differences between demographic groups (Sternberg & the Rainbow Project Collaborators, 2006). The honest claim is not that a portfolio out-predicts a test in a head-to-head, a study no one has run, but the stronger one that validated authentic work is the kind of richer evidence that improves prediction, which is what selection is for (Kane, 2013). And because the work is produced under supervision and defended live, its authorship is established as it is made. An unsupervised portfolio would indeed be easier to fake than a proctored test, which is exactly why the model does not rely on one: supervision and live defense give the work the costly-to-fake property a credible signal requires, while the evidence it preserves makes it a predictive one as well (Spence, 1973).

This also answers the objection that evidence cannot scale. An admissions dean reviewing fifty thousand applications will ask who is going to watch the work when a score takes three seconds, and the answer is that no one has to. The problem was never labels as such but labels that are opaque and unvalidated, a score whose derivation cannot be inspected and a grade that means different things in different schools. A ranking computed transparently from authentic work that has already been validated, and weighted to an institution's own definition of merit, is a label too, but a better one, predictive because it rests on richer evidence and trustworthy because the evidence stays in view. It scales like a score and audits like a portfolio. How that ranking is produced and consumed at the scale of real admissions is the subject of the framework below.

Keeping Teacher Effort Where It Counts

This division of labor also respects the limits of teacher time, which any workable system must. Teachers already work far beyond their contracted hours, about fifty-three hours per week against forty-four for comparable working adults, with much of the excess spent on documentation rather than instruction (Doan et al., 2024). Jurisdictions that required teachers to track and certify proficiency against standards found the added workload substantial. Maine educators named the need for more time as the most prevalent theme of implementation, and the state later made its proficiency-

based diploma mandate optional (Johnson & Stump, 2018). The honest version of this point is not that teaching is in free fall but that teacher time is chronically overcommitted to non-instructional work. Asking teachers to mentor and route authentic student work, rather than to manufacture inferential transcripts, keeps their effort on the part only they can do.

The Evidence Base for Performance Assessment

The case for examining authentic work does not rest on measurement logic alone. Performance-based assessment has an empirical record, and that record now draws on several independent programs rather than one.

The anchor is the collaboration between the City University of New York and the New York Performance Standards Consortium. Students admitted on the strength of authentic-work portfolios, despite lower average SAT scores than their peers, earned higher first-semester grades, completed more credits, and persisted into the second year at higher rates (Fine & Pryiomka, 2020). Independent, peer-reviewed evidence from outside New York corroborates the pattern and reduces reliance on any single program or sponsor. New Hampshire's PACE system, which substitutes local performance assessment for much standardized testing, has produced comparable or modestly improved state-test outcomes across several years, with larger gains for lower-achieving students (Evans, 2019; Perez & Evans, 2023). A multistate study of deeper-learning high schools that emphasize performance assessment found higher graduation rates, stronger test performance, and greater college enrollment than matched comparison schools (Zeiser et al., 2014). At the level of the single task, the International Baccalaureate's supervised and defended extended essay predicts later college grade point average (Inkelas et al., 2013).

Honesty about the limits of this evidence strengthens rather than weakens it. A longer-run follow-up of the deeper-learning cohorts found that several college-enrollment and persistence advantages were no longer statistically significant in the postsecondary years (American Institutes for Research, 2022). The durable claim is that well-implemented performance assessment matches or modestly exceeds conventional measures while widening who can demonstrate capability, not that it guarantees downstream degree completion. None of these programs is a study of HSCred itself. They establish the research foundation on which it is built, and HSCred intends to hold its own results to the same standard by publishing outcome data as the platform operates.

A separate body of work supports the instruction that performance assessment encourages, and it should be kept distinct from the admissions-prediction evidence above. Actively retrieving and applying knowledge improves long-term retention more than passive review (Roediger & Karpicke, 2006), project-based learning produces moderate-to-large achievement gains relative to traditional instruction (Zhang & Ma, 2023; Chen & Yang, 2019), and project-based methods are a well-established route to deep learning in the learning sciences (Krajcik & Blumenfeld, 2006). Finally, broader analyses of college readiness that combine multiple forms of evidence improve prediction of college success beyond the SAT (Gaertner & McClarty, 2015), a pattern also reported in the Rainbow and Kaleidoscope projects (Sternberg, 2010).

Four Institutions That Need to Compare Learning Outcomes

A common source for academic achievement data has four natural adopters. Colleges and admissions offices need a race-neutral way to identify talent after *Students for Fair Admissions v. Harvard* (2023), which prohibited race-conscious admissions while leaving institutions committed to diverse and capable classes, and federal guidance has pressed them to avoid demographic proxies (U.S. Department of Education, Office for Civil Rights, 2023). At the same moment they have lost trust in the essay (Kaplan, 2025). Authentic work answers both pressures: it predicts success on its own terms (Fine & Pryiomka, 2020; Gaertner & McClarty, 2015), and selective institutions are already building channels to collect it. The University of Michigan's Ross School of Business requires an authentic-work portfolio of every first-year applicant (Willis & Martinez, 2022); the Massachusetts Institute of Technology invites a Maker Portfolio of self-designed engineering and build projects, reviewed by faculty and alumni with maker expertise (Massachusetts Institute of Technology, 2026); Olin College of Engineering requires every advancing applicant to attend a Candidates' Weekend whose design challenge and exercises are scored by the Olin community (Franklin W. Olin College of Engineering, 2026); and Bard College admits students outright on a set of long analytical essays graded by its faculty like coursework (Bard College, 2025). No single one of these is yet the norm, and most are optional or alternative routes rather than the main door, which is the honest reading of them. Taken together, though, they show the appetite is not a quirk of one business program but a pattern across business, engineering, and the liberal arts, and the structural pressures behind it are larger than any school: the end of race-conscious admissions, state graduation mandates, federal assessment authority, and a national admissions-counseling coalition now organized around exactly this kind of evidence (Learning Policy Institute, n.d.).

K-12 districts face state mandates that require exactly this infrastructure. New York presented a plan to transform its graduation requirements in 2024 (New York State Education Department, 2024), adopted the Portrait of a Graduate framework in July 2025 (New York State Education Department, 2025a), and is phasing out Regents examinations as graduation requirements beginning in the 2027-28 school year (New York State Education Department, 2025b). Districts now need a graduation-aligned, defensible way to certify performance, and the research base for doing so is established (Evans, 2019; Guha et al., 2018).

The federal Department of Education has already written performance assessment into law. The Innovative Assessment Demonstration Authority enacted under the Every Student Succeeds Act explicitly enumerates performance-based assessments as a permissible state system (20 U.S.C. § 6364). The authority is real, funded, and evaluated: the official federal evaluation of its early implementation was conducted by the Institute of Education Sciences, not by any absent oversight report (Troppe et al., 2023), and the program continues to invite states into innovative assessment (U.S. Department of Education, Office of Elementary and Secondary Education, 2025).

Foundations and philanthropy have funded this transition for more than a decade and seek scalable vehicles for it. The William and Flora Hewlett Foundation's deeper-learning strategy set out to transform what educational success looks like and how it is measured (William and Flora Hewlett Foundation, n.d.), and it underwrote much of the supporting evidence (Zeiser et al., 2014; Fine & Pryiomka, 2020). The field has also organized: the Reimagining College Access initiative, launched by

the Learning Policy Institute with EducationCounsel in 2017 and since 2023 administered by the National Association for College Admission Counseling, has built a national coalition dedicated to using authentic high-school performance evidence in admissions (Learning Policy Institute, n.d.). That the field's principal professional body now stewards this work marks it as an institutional priority rather than a fringe experiment.

Students and Under-Resourced Schools

The same standard that serves institutions answers a problem students face directly. The advanced, distinction-bearing coursework that signals capability to selective colleges is unevenly available. Among public high schools that offer Advanced Placement at all, roughly seventy percent of low-poverty schools offer more than ten AP courses, against about thirty percent of high-poverty schools, a gap the Government Accountability Office ties to differences in resources and staffing (U.S. Government Accountability Office, 2018). Federal civil-rights data show the same pattern by race and school composition, with calculus, physics, and advanced coursework markedly less likely to be offered in schools serving mostly Black and Hispanic students (U.S. Department of Education, Office for Civil Rights, 2024), and the College Board's own figures show that the share of students whose high schools offer five or more AP courses ranges from 49 percent of Native American students to 93 percent of Asian students (College Board, 2025). A capable student at a school that offers no Advanced Placement, no International Baccalaureate, and few advanced courses has, through no fault of their own, fewer ways to demonstrate what they can do.

A portable, expert-validated credential closes that gap without asking every school to stand up an AP or IB program. A student can produce and defend authentic work, and a college can see it, regardless of what the building happens to offer. For a high school, the same platform is a way to give its strongest students advanced academic distinction, and to meet new graduation expectations, without constructing a parallel apparatus from scratch.

Scalable, Expert-Validated Assessment

The product that meets this demand is a shared platform that lets distributed experts validate authentic student work against common rubrics. Its assessment design proceeds in three phases, building on three decades of Consortium practice.

In the first phase, students demonstrate foundational content mastery before beginning project work, which ensures they reach advanced work prepared and is consistent with cognitive-load theory (Sweller, 1988) and with reviews of mastery learning (Winget & Persky, 2022). In the second phase, students design original projects within established rubric frameworks, working with supervising educators who act as mentors rather than directors. Structured choice of this kind raises motivation and engagement while rubrics preserve rigor (Ryan & Deci, 2000; Evans & Boucher, 2015; Cullen & Oppenheimer, 2024). In the third phase, students produce and publicly defend a polished piece of work, which is evaluated by the supervising teacher together with several independent expert evaluators through cycles of feedback and revision.

The first phase deserves emphasis, because it answers the objection most often raised against performance assessment, that portfolios are a way around rigor and that only a test is demanding enough. The model takes the opposite position. It makes mastery of the content layer a precondition for project work rather than an afterthought: before a student may begin, the first phase requires demonstrated command of a subject's core vocabulary and concepts, verified by assessment. A student cannot report or defend learning in a field without first holding its basic knowledge, and most portfolio systems impose no such gate. The requirement rests on a settled finding in cognitive science. Higher-order thinking is not a content-free skill exercised in the abstract; critical thinking is bound to knowledge of the domain in which one is thinking (Willingham, 2007), the generic skills schools prize are in practice applications of domain-specific knowledge and cannot be taught or shown without it (Tricot & Sweller, 2014), and novices set loose on inquiry before they hold the relevant schemas are overloaded rather than enabled (Kirschner, Sweller, & Clark, 2006; Sweller, 1988). What separates expert from novice performance is organized domain knowledge, not general ability (Chi, Feltovich, & Glaser, 1981). Requiring mastery first is therefore the condition under which project work becomes genuinely rigorous rather than merely active.

The requirement is also, on its own, one of the better-evidenced practices in education. Mastery learning, in which students must demonstrate command of prerequisite material before advancing, raises achievement by roughly half a standard deviation, with the largest gains for students who begin behind (Bloom, 1968; Kulik, Kulik, & Bangert-Drowns, 1990). What the model changes is not whether students are tested but where the test sits. In a conventional unit the test is the terminal event, the moment learning is presumed to stop; here it is the entry condition, after which the harder work of applying, producing, and defending knowledge begins. The effect is to raise the bar, not lower it, by extending rigor past the point most curricula treat as the finish line.

Where this content-first, project-second sequence has been tested directly, it holds, and on conventional measures. Project-based instruction has been examined in large randomized trials whose outcomes were scored on standardized instruments. In a multi-district trial, students in a project-based version of two Advanced Placement courses earned credit-qualifying examination scores at higher rates than peers in lecture-based sections, including students from low-income households (Saavedra et al., 2022). A cluster-randomized trial of second-grade units raised social-studies and informational-reading achievement in low-income schools (Duke, Halvorsen, Strachan, Kim, & Konstantopoulos, 2021), and a randomized trial of a project-based science program raised standardized science scores by about 0.28 standard deviations (Krajcik et al., 2023). The recurring worry that project work trades away rigor is, on this evidence, not supported: gated by the content mastery just described, project work preserves and often raises measurable achievement.

The concern is sharpest in mathematics, a cumulative and precise subject that seems to demand drilling rather than projects. Yet the discipline's own standards locate rigor in exactly the depth that follows fluency. The National Council of Teachers of Mathematics names productive struggle and high-cognitive-demand tasks as core practices of effective teaching rather than enrichment (National Council of Teachers of Mathematics, 2014); reviews of classroom research conclude that struggling with important mathematics is a condition for understanding it (Hiebert & Grouws, 2007); and controlled studies of productive failure show that students who grapple with a novel problem before being taught the canonical method develop deeper conceptual understanding (Kapur, 2014). Tasks

built with a low floor and a high ceiling let students at different levels work the same rich problem (Boaler, 2016), and the heterogeneous, flexibly grouped classrooms such tasks invite raise achievement at every level while rigid between-class tracking does not (Steenbergen-Hu, Makel, & Olszewski-Kubilius, 2016). None of this is hypothetical: New York's Consortium has for three decades assessed mathematical problem-solving through performance tasks that students defend orally before evaluators. The subject thought least hospitable to performance assessment is, on the evidence, one it already serves.

Reliability is the question administrators raise first, and it is the one on which performance assessment has historically failed: when complex work is scored across many sites without strong supports, agreement among raters can fall too low to sustain comparison (Koretz et al., 1994). The lesson of the measurement literature, however, is that inter-rater reliability is not a fixed property of a method but a quantity that rubrics, rater training, calibration, and structured moderation raise (Jonsson & Svingby, 2007; Reddy & Andrade, 2010; Brennan, 2000). The New York Performance Standards Consortium demonstrates this in practice: external evaluators applying shared rubrics with calibration and moderation reach consistent judgments of graduation-level work across thirty-eight schools over three decades, with students defending their work orally before a panel in the manner of a thesis defense (Cook & Tashlik, 2017; New York Performance Standards Consortium, n.d.). These are the supports that established large-scale assessments, Advanced Placement among them, rely on for reliable scoring. The claim is not that expert judgment is more reliable than a multiple-choice key, but that it can meet the standard to which high-stakes assessments are held, and that the model makes meeting it a measured and continuous obligation rather than a one-time assertion.

The platform operationalizes that obligation in two ways. First, each work product is scored independently and blind by several expert evaluators, and the pattern of their scores is monitored over time. A persistent rater effect, an evaluator who is consistently more lenient or more severe than the colleagues they are randomly paired with, is exactly the kind of signal that measurement models are built to detect (Myford & Wolfe, 2003); evaluators who diverge are flagged for retraining, and those who do not converge with their peers are removed from scoring. Because evaluation is a compensated role, raters have a standing incentive to calibrate themselves in order to keep it, so reliability is maintained by the same mechanism that pays for it. Second, because each evaluation channel may design its own rubric and training, and because the trust admissions offices place in a channel, and therefore the endorsements, applicants, and compensation that follow, depends on the reliability it can show, channels have a market incentive to improve their methods. Reliability becomes something the system continuously competes to raise. Establishing the level it reaches at scale remains an open empirical question.

The design borrows the quality mechanisms that make established assessments credible while removing what they conceal. Advanced Placement earns trust through trained graders, calibration, and sustained practice, yet it keeps the student work itself out of view, reaches only students whose schools offer it, and recognizes only what its own curriculum covers. A validated work portfolio can keep the mechanisms, several expert evaluators, shared rubrics, calibration, and revision, while making the underlying work visible and available to any student in any school. Rather than a central au-

thority pushing a single instrument onto schools, any university team can define the criteria for work it values and invite students to meet them, so evaluation becomes a distributed network of experts rather than a single gatekeeper.

A second concern follows naturally: if the work is a produced artifact, will a slick presentation beat substantive thinking? The answer lies in the rubric, the instrument that decides what a score rewards. A well-constructed analytic rubric scores the quality of thinking separately from the polish of its delivery, and rubrics of this kind not only raise the reliability of expert judgment but direct it toward substance, the criteria the rubric names rather than the surface it does not (Jonsson & Svingby, 2007; Reddy & Andrade, 2010). The University of Michigan's Ross School of Business, whose admissions portfolio is scored with rubrics adapted from the Association of American Colleges and Universities VALUE rubrics, illustrates the point. Its readers score for problem definition, problem-solving, content knowledge, and reflection, and they report that polished signals can mislead rather than impress, since "the use of technical jargon or detailed descriptions of elite internships may even distract from demonstrating the critical thinking and inquiry skills at the heart of the question" (Willis & Martinez, 2022, p. 17). In one case a reader rewarded an applicant precisely for submitting a video of a practice session rather than a polished final performance, because it showed genuine effort and a willingness to improve (Willis & Martinez, 2022, p. 21). A rubric built to value thinking is what makes performance-based admission resistant to gaming by production values, and it is a design property that high-quality performance assessments share (Badrinarayan, 2022).

The platform turns rubric quality into a competitive property rather than a fixed input. Admissions offices decide how much weight to give each evaluation channel, and that weighting, which produces the effective ranking of students and shapes which channels students seek out, rewards channels whose rubrics reliably identify capable students and penalizes those that reward polish over depth. A channel that certifies impressive-looking but shallow work loses the trust of the institutions it is trying to reach, and with it the applicants and the compensation that follow. The incentive runs toward rubrics that measure thinking, the same direction the measurement evidence points, so that the property academics most want from an assessment, that it cannot be gamed by surface, is the property the platform's economics reward.

What an admissions office receives from this is not a stack of submissions to watch but a ranked view it controls. Because every work product is validated and scored by its evaluation channel before it is ever surfaced, and because each institution composes its own weights across the channels it trusts into a single ranking, the platform presents incoming students as a sorted dashboard tuned to that institution's priorities, with the underlying work and its named evaluators one click beneath each name. A reader can trust the ranking the way a reader trusts any score, and then open the evidence behind any candidate in a way no score allows. This is how examining authentic work survives admissions at scale: the evaluation is done before the reader arrives, the ranking is the institution's own, and the work is there to audit rather than to wade through.

The On-Ramp

A reader may notice that the brief has so far made a general case for authentic academic work in every subject, and may fairly ask why the discussion now turns to journalism in particular. The reason is practical rather than ideological. The platform is discipline-agnostic, and the strongest authentic work will come from science, engineering, history, and mathematics as readily as from reporting. But a credential is only as good as the work that flows into it, and the case for performance-based assessment is widely accepted in principle and rarely adopted in practice. The obstacle is not evidence but entry. Building a portfolio system from nothing asks a school to invent tasks, train evaluators, and sustain moderation before it can produce a single comparable record, and local portfolios that never leave the building do not scale (Koretz et al., 1994). What schools need is not another mandate but a first step that is inexpensive, motivating, and already familiar. The school newsroom is an unusually good one, not because journalism is the destination but because it is the most universal on-ramp, since nearly every school can start one tomorrow, in any subject, at almost no cost. Journalism across the curriculum treats reporting and publishing as the entry point through which any subject can generate authentic, reviewable student work.

From writing across the curriculum to journalism across the curriculum

The idea has a direct ancestor. Writing across the curriculum began as a theory that writing is not merely a way to record knowledge but a mode of learning in its own right, a means of generating and organizing thought that belongs in every subject rather than only in English (Emig, 1977; Britton et al., 1975). The empirical case followed: different writing tasks engage different kinds of thinking and shape what students come to understand in content classrooms (Langer & Applebee, 1987), a meta-analysis of writing-to-learn interventions found consistent achievement gains that were larger when tasks asked students to reflect on their own understanding (Bangert-Drowns et al., 2004), and disciplinary-literacy research shows that the reading and writing that matter are specific to each field and best taught inside it (Shanahan & Shanahan, 2008).

Journalism across the curriculum is the digital-era successor to that idea, and it adds one decisive property. Where writing across the curriculum aimed to make thinking happen, journalism across the curriculum makes thinking visible. It requires a student to externalize reasoning into a sourced, published, defensible artifact addressed to a real audience, which is precisely the move that cognitive apprenticeship identifies as the engine of learning, making expert thinking observable so it can be examined and improved (Collins, Brown, & Holum, 1991; Ritchhart, Church, & Morrison, 2011). Industrial schooling embedded literacy across the curriculum because the economy required readers and writers in every field. The digital economy rewards people who can investigate, synthesize, and publish, so the contemporary analog embeds reporting and publishing in every subject. A science investigation becomes a filmed explainer, a history inquiry becomes a documentary segment, a mathematics problem becomes a worked and narrated solution. In each case the student's thinking stops being private and becomes a reviewable record, which is exactly what a portfolio credential and the low-inference standard described earlier in this brief require.

That visibility serves more than assessment. Producing the explanation is itself a learning act. Generative-learning research finds that strategies requiring a student to explain or teach what they have studied, rather than restudy it, deepen understanding (Fiorella & Mayer, 2016), and the benefit of

prompting self-explanation holds across dozens of studies (Bisra et al., 2018). The medium bears directly on the choice to host video rather than written work: when students explained material to an imagined peer, those who recorded a spoken explanation on video learned more than those who restudied, while those who wrote their explanation did not (Hoogerheide et al., 2016), and creating a short instructional video as a homework task produced the same advantage over restudy (Hoogerheide et al., 2019). The audience matters as much as the medium. Simply expecting to teach material to others, rather than to be tested on it, leads students to study it more fully and to recall it in a better-organized way (Nestojko, Bui, Kornell, & Bjork, 2014), and that benefit becomes durable when they go on to actually teach or present it rather than only prepare to do so (Fiorella & Mayer, 2013). A credential built on work that students present and defend to real evaluators, rather than submit and forget, is therefore organized around the conditions under which learning lasts. Asking a student to narrate and defend reasoning on camera therefore serves two ends at once: it produces the visible, reviewable artifact the credential needs, and it advances the learning that artifact records.

Why a newsroom is the natural first step

A newsroom supplies the three conditions that make authentic assessment work and that conventional classrooms usually lack: a real audience, a real deadline, and a reason to verify. Writing changes when someone beyond the teacher will read it. Audience awareness is bound up with both the cognitive work of writing and the motivation to do it well (Magnifico, 2010), and design studies show that giving students a genuine audience and a stake in their topic raises both engagement and the authenticity students perceive in the task (Behizadeh, 2019). A deadline turns intention into product. The obligation to verify, which is native to journalism, turns assertion into evidence. Together these conditions convert passive participation into knowledge production, which is the defining feature of authentic intellectual work, inquiry that yields products with value beyond school (Newmann & Wehlage, 1993).

What the research shows about newsroom outcomes

Among school-based practices, journalism has an unusually strong outcomes record, though the evidence is correlational and must be read with care. The most credible work confronts that limitation directly. A first study using the nationally representative Education Longitudinal Study of 2002 established that journalism attracts students who are already stronger writers, more engaged, and more advantaged, which means raw comparisons overstate its effects (Bobkowski, Cavanah, & Miller, 2017). A reassessment of the same dataset that controlled for those prior characteristics still found that students who took more high school journalism scored higher on standardized English measures and, if they majored in journalism, earned higher college English grades (Bobkowski & Cavanah, 2019). Read together, these two studies are stronger than either alone, because the second answers the selection problem the first identified.

The civic findings are similar in quality. Using the same national dataset, high school journalism is associated with higher rates of voting and community volunteering in early adulthood, with the association comparable to debate or student government and notably larger for lower-income students, for whom journalism appears to narrow a participation gap (Bobkowski & Miller, 2016). A later survey of high school journalists and their advisers traced a mechanism, a sense of civic self-efficacy that grows when students have genuine editorial freedom rather than tightly controlled output (Bobkowski

& Rosenthal, 2022). A large descriptive analysis linked to college-entrance data corroborates the achievement pattern across more than thirty thousand students, while honestly showing that the advantages concentrate in verbal rather than mathematical measures (Dvorak, 2008). And the capacity journalism is meant to build, the disciplined evaluation of sources and claims, is a measurable construct in its own right (Ashley, Maksl, & Craft, 2013). No randomized or quasi-experimental study of scholastic journalism exists, so the honest claim is not that a newsroom causes these gains in isolation but that, across nationally representative data and with selection effects accounted for, journalism is among the better-supported routes to the academic and civic outcomes schools say they want.

Journalism and the Portrait of a Graduate

Those are precisely the outcomes that the new graduation frameworks now ask schools to produce. New York's Portrait of a Graduate, adopted in July 2025, defines six attributes a graduate should demonstrate, namely Academically Prepared, Creative Innovator, Critical Thinker, Effective Communicator, Global Citizen, and Reflective and Future Focused, all viewed through a culturally responsive lens (New York State Education Department, 2025c). The state intends these to be shown through multiple measures and authentic demonstration rather than a single new test, which is the performance-based model this brief describes (New York State Education Department, 2025b). The question for any district is which classroom practice produces evidence of all six at once, and journalism is an unusually direct answer because the research maps onto the attributes rather than merely asserting them.

The fit is closest on the attribute most frameworks struggle to operationalize. An Effective Communicator, in the state's words, articulates ideas through speaking, writing, and different media for various purposes while engaging diverse audiences, which is a description of journalistic practice itself, and the strongest controlled evidence for journalism concerns exactly these communication and writing outcomes (Bobkowski & Cavanah, 2019). The Critical Thinker who evaluates evidence and connects information across domains is enacting the reporter's discipline of verification, the same capacity captured by news-literacy measures (Ashley et al., 2013). The Global Citizen who acts on civic knowledge in local and digital communities is the attribute best supported of all, given journalism's documented association with voting, volunteering, and civic self-efficacy, and its larger effect for lower-income students speaks to the equity aim at the center of the Portrait (Bobkowski & Miller, 2016; Bobkowski & Rosenthal, 2022). The Creative Innovator produces original work for an audience, which is what publishing is. The Reflective and Future Focused student revises against feedback and an approaching deadline, which is the newsroom's daily cycle. And Academically Prepared is the through-line of the achievement evidence above. No single subject covers the Portrait so completely while generating, as a byproduct, the reviewable artifacts a portfolio credential needs.

This is not an argument that journalism is the only path. Other disciplines, science and engineering most of all, produce authentic and demanding work, and a strong STEM investigation can anchor a portfolio as well as any feature story. The distinction is narrower and more important. A portfolio system, and the Effective Communicator attribute specifically, requires students to present their work to a real audience and defend it, and that presentational competence is journalism's core subject rather than an add-on. STEM supplies the substance; journalism supplies the practice of making that

substance public, sourced, and defensible. Journalism across the curriculum is therefore best understood not as a rival to other subjects but as the connective discipline that takes the thinking happening in any of them and renders it visible, which is the form in which it can be assessed.

Journalism as a low-inference stance

This visibility is what ties the newsroom back to the brief's central measurement argument. Authentic assessment means examining real student performance on worthy tasks rather than proxies for it (Wiggins, 1990), and a published, sourced piece of student work is a stronger signal than a score or a narrative claim of mastery, for the reasons set out earlier. The competency systems with which performance assessment is often confused ask raters to infer a label from the work, and the label's reliability depends on a cross-school comparability that the evidence shows is hard to achieve (Steele et al., 2014). The newsroom keeps the evidence itself in view and leaves the higher-inference judgments to the universities that must compare applicants. It does so, moreover, without adding to the documentation burden that already consumes teacher time (Doan et al., 2024; Johnson & Stump, 2018), because the artifact is the record.

From a single newsroom to a trusted pipeline

A newsroom begins as a classroom practice, but the same artifacts can feed a credential that institutions trust. A defended, published, expert-validated artifact resists generative AI in a way a take-home essay does not, because its provenance is established through supervision and live revision rather than inferred after submission. It also answers the post-decision admissions problem directly: a record of authentic published work lets a college identify and invite talent on the strength of what a student has actually produced, rather than through demographic proxies or the socioeconomic shadow that standardized scores carry (Students for Fair Admissions v. Harvard, 2023). The newsroom, then, is not only the easiest place for a school to begin. It is the point at which this brief's larger argument, that authentic work is the signal that survives AI and travels across institutions, becomes something a single teacher and a single student can start now.

Conclusion

This brief identifies three gaps in contemporary academic assessment and shows that they are one gap seen from three sides. The predictive-validity gap is, stated precisely, an evidence gap: institutions decide without a shared, high-fidelity record of demonstrated capability, and prediction improves when such a record replaces proxies that track test-taking facility and institutional advantage. The equity gap is the same deficiency felt by capable students at under-resourced schools who have no comparable way to be seen. The scalability gap is why the remedy, long recognized in elite and grant-funded settings, has never reached ordinary institutions.

Generative AI has made the gap urgent by dissolving the take-home written artifact and the detection regimes built to police it, leaving supervised, expert-validated, live-defended work as the credential that still means what it says. What that work needs in order to function across institutions is a common standard: a comparable, auditable, portable record of authentic achievement, validated by a distributed network of experts rather than asserted by a single gatekeeper. The measurement case favors preserving the evidence over compressing it into a label, the empirical record across several

independent programs shows the approach works, four institutions are positioned to adopt it, and the school newsroom offers a place to begin tomorrow. The opportunity is to build the standard, not to wait for one to arrive.

References

- IEdTech Consortium. (n.d.). *Comprehensive Learner Record Standard*. Retrieved June 23, 2026, from <https://www.iedtech.org/initiatives/digital-credentials/clr>
- ACT. (2022). *Grade inflation continues to grow in the past decade* (Report No. R2134). ACT. <https://www.act.org/content/dam/act/unsecured/documents/2022/R2134-Grade-Inflation-Continues-to-Grow-in-the-Past-Decade-Final-Accessible.pdf>
- Allensworth, E. M., & Clark, K. (2020). High school GPAs and ACT scores as predictors of college completion: Examining assumptions about consistency across high schools. *Educational Researcher*, 49(3), 198–211. <https://doi.org/10.3102/0013189X20902110>
- American Association of Collegiate Registrars and Admissions Officers, & NASPA. (2018). *Comprehensive Learner Record*. <https://www.aacrao.org/consulting/our-services/comprehensive-learner-record-learning-and-employment-record>
- American Institutes for Research. (2022). *The study of deeper learning: College enrollment, persistence, and completion* (Report No. 7). <https://www.air.org/sites/default/files/2022-07/Deeper-Learning-College-Outcomes-Report-7-June-2022-508.pdf>
- Ashley, S., Maksl, A., & Craft, S. (2013). Developing a news media literacy scale. *Journalism & Mass Communication Educator*, 68(1), 7–21. <https://doi.org/10.1177/1077695812469802>
- Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, 118(4), 1279–1333. <https://doi.org/10.1162/003355303322552801>
- Badrinarayan, A. (2022). *Performance assessments in college admission: Designing an effective and equitable process*. Learning Policy Institute. <https://doi.org/10.54300/150.937>
- Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research*, 74(1), 29–58. <https://doi.org/10.3102/00346543074001029>
- Bard College. (2025). *The Bard Entrance Examination*. <https://www.bard.edu/bardexam/>
- Behizadeh, N. (2019). Aiming for authenticity: Successes and struggles of an attempt to increase authenticity in writing. *Journal of Adolescent & Adult Literacy*, 62(4), 411–419. <https://doi.org/10.1002/jaal.911>
- Bisra, K., Liu, Q., Nesbit, J. C., Salimi, F., & Winne, P. H. (2018). Inducing self-explanation: A meta-analysis. *Educational Psychology Review*, 30(3), 703–725. <https://doi.org/10.1007/s10648-018-9434-x>
- Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment*, 1(2), 1–12.
- Boaler, J. (2016). *Mathematical mindsets: Unleashing students' potential through creative math, inspiring messages, and innovative teaching*. Jossey-Bass.
- Bobkowski, P. S., & Cavanah, S. B. (2019). When “journalism kids” do better: A reassessment of secondary and postsecondary achievement and activities. *Journalism & Mass Communication Educator*, 74(4), 438–451. <https://doi.org/10.1177/1077695819841186>
- Bobkowski, P. S., Cavanah, S. B., & Miller, P. R. (2017). Who are the “journalism kids”? Academic predictors of journalism participation in secondary schools. *Journalism & Mass Communication Educator*, 72(1), 68–82. <https://doi.org/10.1177/1077695815622770>
- Bobkowski, P. S., & Miller, P. R. (2016). Civic implications of secondary school journalism: Associations with voting propensity and community volunteering. *Journalism & Mass Communication Quarterly*, 93(3), 530–550. <https://doi.org/10.1177/1077699016628821>
- Bobkowski, P. S., & Rosenthal, H. M. (2022). Journalism civic self-efficacy: Predicting political participation among secondary-school journalism students. *Journalism Practice*, 16(10), 2166–2184. <https://doi.org/10.1080/17512786.2021.1897475>
- Brennan, R. L. (2000). Performance assessments from the perspective of generalizability theory. *Applied Psychological Measurement*, 24(4), 339–353. <https://doi.org/10.1177/01466210022031796>
- Britton, J., Burgess, T., Martin, N., McLeod, A., & Rosen, H. (1975). *The development of writing abilities (11–18)*. Macmillan Education.
- Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research*, 86(4), 803–848. <https://doi.org/10.3102/0034654316672069>
- Brown University. (2024, March 5). *Brown University to reinstate test requirement, retain Early Decision, further study legacy preferences*. <https://www.brown.edu/news/2024-03-05/admissions>
- 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.
- Chen, C.-H., & Yang, Y.-C. (2019). Revisiting the effects of project-based learning on students' academic achievement: A meta-analysis investigating moderators. *Educational Research Review*, 26, 71–81. <https://doi.org/10.1016/j.edurev.2018.11.001>

- Chetty, R., Deming, D. J., & Friedman, J. N. (2023). *Diversifying society's leaders? The determinants and causal effects of admission to highly selective private colleges* (Working Paper No. 31492). National Bureau of Economic Research. <https://doi.org/10.3386/w31492>
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5(2), 121-152. https://doi.org/10.1207/s15516709cog0502_2
- College Board. (2025). *AP national and state data*. AP Central. <https://apcentral.collegeboard.org/about-ap/ap-data-research/national-state-data>
- Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator*, 15(3), 6-11, 38-46.
- Common App. (2025). *End-of-season report, 2024-2025: First-year application trends*. https://www.commonapp.org/files/DAR/deadline-updates/Common-App-End-of-Season-Report_24-25.pdf
- Cook, A., & Tashlik, P. (2017). Building a system of assessment: Examining the New York Performance Standards Consortium. In D. Meier & M. Knoester (Eds.), *Beyond testing: Seven assessments of students and schools more effective than standardized tests* (pp. 89-112). Teachers College Press.
- Cullen, S., & Oppenheimer, D. M. (2024). Choosing to learn: The importance of student autonomy in higher education. *Science Advances*, 10(29), Article eado6759. <https://doi.org/10.1126/sciadv.ado6759>
- Dartmouth College. (2024, February 5). *Reactivating the SAT/ACT requirement for Dartmouth undergraduate admissions*. Office of the President. <https://president.dartmouth.edu/news/2024/02/reactivating-satact-requirement-dartmouth-undergraduate-admissions>
- Dawson, P. (2021). *Defending assessment security in a digital world: Preventing e-cheating and supporting academic integrity in higher education*. Routledge. <https://doi.org/10.4324/9780429324178>
- Dawson, P., Bearman, M., Dollinger, M., & Boud, D. (2024). Validity matters more than cheating. *Assessment & Evaluation in Higher Education*, 49(7), 1005-1016. <https://doi.org/10.1080/02602938.2024.2386662>
- Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, 132(4), 1593-1640. <https://doi.org/10.1093/qje/qjx022>
- Dixon-Román, E. J., Everson, H. T., & McArdle, J. J. (2013). Race, poverty and SAT scores: Modeling the influences of family income on Black and White high school students' SAT performance. *Teachers College Record*, 115(4), 1-33. <https://doi.org/10.1177/016146811311500406>
- Doan, S., Steiner, E. D., & Pandey, R. (2024). *Teacher well-being and intentions to leave in 2024: Findings from the 2024 State of the American Teacher Survey* (RR-A1108-12). RAND Corporation. <https://doi.org/10.7249/RR-A1108-12>
- Duke, N. K., Halvorsen, A.-L., Strachan, S. L., Kim, J., & Konstantopoulos, S. (2021). Putting PjBL to the test: The impact of project-based learning on second graders' social studies and literacy learning and motivation in low-SES school settings. *American Educational Research Journal*, 58(1), 160-200. <https://doi.org/10.3102/0002831220929638>
- Dvorak, J. (2008). *High school journalism matters*. Newspaper Association of America Foundation.
- Eaton, S. E. (2023). Postplagiarism: Transdisciplinary ethics and integrity in the age of artificial intelligence and neurotechnology. *International Journal for Educational Integrity*, 19, Article 23. <https://doi.org/10.1007/s40979-023-00144-1>
- Emig, J. (1977). Writing as a mode of learning. *College Composition and Communication*, 28(2), 122-128. <https://doi.org/10.58680/cc197716382>
- Evans, C. M. (2019). Effects of New Hampshire's innovative assessment and accountability system on student achievement outcomes after three years. *Education Policy Analysis Archives*, 27, Article 10. <https://doi.org/10.14507/epaa.27.4014>
- Evans, M., & Boucher, A. R. (2015). Optimizing the power of choice: Supporting student autonomy to foster motivation and engagement in learning. *Mind, Brain, and Education*, 9(2), 87-91. <https://doi.org/10.1111/mbe.12073>
- Every Student Succeeds Act of 2015, 20 U.S.C. § 6364.
- FairTest. (2025, September 4). *Overwhelming majority of U.S. colleges and universities remain ACT/SAT-optional or test-blind/score-free for fall 2026* [Press release]. National Center for Fair & Open Testing. <https://fairtest.org/overwhelming-majority-of-schools-remain-test-optional-or-test-free-for-fall-2026-admission/>
- Fawns, T., Boud, D., & Dawson, P. (2026). Identifying what our students have learned: A framework for practical assessment validation. *Assessment & Evaluation in Higher Education*. Advance online publication. <https://doi.org/10.1080/02602938.2026.2620053>
- Fine, M., & Priyomka, K. (2020). *Assessing college readiness through authentic student work: How the City University of New York and the New York Performance Standards Consortium are collaborating toward equity*. Learning Policy Institute. <https://learningpolicyinstitute.org/product/assessing-college-readiness-authentic-student-work-report>
- Fiorella, L., & Mayer, R. E. (2013). The relative benefits of learning by teaching and teaching expectancy. *Contemporary Educational Psychology*, 38(4), 281-288. <https://doi.org/10.1016/j.cedpsych.2013.06.001>
- Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, 28(4), 717-741. <https://doi.org/10.1007/s10648-015-9348-9>

- Franklin W. Olin College of Engineering. (2026). *Admission process*. <https://www.olin.edu/admission/apply/admission-process>
- Gaertner, M. N., & McClarty, K. L. (2015). Performance, perseverance, and the full picture of college readiness. *Educational Measurement: Issues and Practice*, 34(2), 20–33. <https://doi.org/10.1111/emip.12066>
- Geiser, S., & Santelices, M. V. (2007). *Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes* (Research & Occasional Paper Series CSHE.6.07). Center for Studies in Higher Education, University of California, Berkeley. <https://cshe.berkeley.edu/publications/validity-high-school-grades-predicting-student-success-beyond-freshman-yearhigh-school>
- Georgetown University. (2023, October 27). *SAT score gaps reveal deeper inequality in education, opportunity*. The Feed. <https://feed.georgetown.edu/access-affordability/sat-score-gaps-show-the-consequences-of-widening-economic-disparities/>
- Goldhaber, M. H. (1997). The attention economy and the Net. *First Monday*, 2(4). <https://firstmonday.org/ojs/index.php/fm/article/view/519>
- Guha, R., Wagner, T., Darling-Hammond, L., Taylor, T., & Curtis, D. (2018). *The promise of performance assessments: Innovations in high school learning and college admission*. Learning Policy Institute. <https://learningpolicyinstitute.org/product/promise-performance-assessments-report>
- Harvard University. (2024, April 11). *Harvard announces return to required testing*. Harvard Gazette. <https://news.harvard.edu/gazette/story/2024/04/harvard-announces-return-to-required-testing/>
- Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371–404). Information Age Publishing.
- Hoogerheide, V., Deijkers, L., Loyens, S. M. M., Heijltjes, A., & van Gog, T. (2016). Gaining from explaining: Learning improves from explaining to fictitious others on video, not from writing to them. *Contemporary Educational Psychology*, 44–45, 95–106. <https://doi.org/10.1016/j.cedpsych.2016.02.005>
- Hoogerheide, V., Visee, J., Lachner, A., & van Gog, T. (2019). Generating an instructional video as homework activity is both effective and enjoyable. *Learning and Instruction*, 64, Article 101226. <https://doi.org/10.1016/j.learninstruc.2019.101226>
- Inkelas, K. K., Swan, A. K., Pretlow, J., & Jones, J. N. (2013). *Exploring the benefits of the International Baccalaureate extended essay for university studies at the University of Virginia*. International Baccalaureate Organization. <https://www.ibo.org/research/outcomes-research/diploma-studies/>
- Johnson, A., & Stump, E. (2018). *Proficiency-based high school diploma systems in Maine: Educator perceptions of implementation*. Maine Education Policy Research Institute. <https://files.eric.ed.gov/fulltext/ED586185.pdf>
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130–144. <https://doi.org/10.1016/j.edurev.2007.05.002>
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73. <https://doi.org/10.1111/jedm.12000>
- Kaplan. (2025). *College admissions officers survey: Applicants' use of AI in essays*. Kaplan. <https://kaplan.com/about/press-media/college-admissions-officers-survey-2025-essays-ai>
- Kapur, M. (2014). Productive failure in learning math. *Cognitive Science*, 38(5), 1008–1022. <https://doi.org/10.1111/cogs.12107>
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1
- Knox, L. (2023, October 9). *Admissions offices turn to AI for application reviews*. Inside Higher Ed. <https://www.insidehighered.com/news/admissions/traditional-age/2023/10/09/admissions-offices-turn-ai-application-reviews>
- Koretz, D. (2017). *The testing charade: Pretending to make schools better*. University of Chicago Press.
- Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994). The Vermont portfolio assessment program: Findings and implications. *Educational Measurement: Issues and Practice*, 13(3), 5–16. <https://doi.org/10.1111/j.1745-3992.1994.tb00443.x>
- Krajcik, J. S., & Blumenfeld, P. C. (2006). Project-based learning. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 317–333). Cambridge University Press.
- Krajcik, J., Schneider, B., Miller, E. A., Chen, I.-C., Bradford, L., Baker, Q., Bartz, K., Miller, C., Li, T., Codere, S., & Peek-Brown, D. (2023). Assessing the effect of project-based learning on science learning in elementary schools. *American Educational Research Journal*, 60(1), 70–102. <https://doi.org/10.3102/00028312221129247>
- Kulik, C.-L. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research*, 60(2), 265–299. <https://doi.org/10.3102/00346543060002265>
- Langer, J. A., & Applebee, A. N. (1987). *How writing shapes thinking: A study of teaching and learning* (NCTE Research Report No. 22). National Council of Teachers of English.

- Learning Policy Institute. (n.d.). *Reimagining college access*. Retrieved June 23, 2026, from <https://learningpolicyinstitute.org/project/reimagining-college-access>
- Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns*, 4(7), Article 100779. <https://doi.org/10.1016/j.patter.2023.100779>
- Magnifico, A. M. (2010). Writing for whom? Cognition, motivation, and a writer's audience. *Educational Psychologist*, 45(3), 167-184. <https://doi.org/10.1080/00461520.2010.493470>
- Massachusetts Institute of Technology. (2026). *First-year applicants: Creative portfolios*. MIT Admissions. <https://mitadmissions.org/apply/firstyear/portfolios-additional-material/>
- MIT Admissions. (2022, March 28). *We are reinstating our SAT/ACT requirement for future admissions cycles*. <https://mitadmissions.org/blogs/entry/we-are-reinstating-our-sat-act-requirement-for-future-admissions-cycles/>
- Myford, C. M., & Wolfe, E. W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of Applied Measurement*, 4(4), 386-421.
- National Council of Teachers of Mathematics. (2014). *Principles to actions: Ensuring mathematical success for all*. National Council of Teachers of Mathematics.
- National Institute for Learning Outcomes Assessment. (n.d.). *Transparency framework*. Retrieved June 23, 2026, from <https://www.learningoutcomesassessment.org/ourwork/transparency-framework/>
- Nestojko, J. F., Bui, D. C., Kornell, N., & Bjork, E. L. (2014). Expecting to teach enhances learning and organization of knowledge in free recall of text passages. *Memory & Cognition*, 42(7), 1038-1048. <https://doi.org/10.3758/s13421-014-0416-z>
- Newmann, F. M., & Wehlage, G. G. (1993). Five standards of authentic instruction. *Educational Leadership*, 50(7), 8-12.
- New York Performance Standards Consortium. (n.d.). *History of the Consortium*. Retrieved June 24, 2026, from <http://www.performanceassessment.org/history-22>
- New York State Education Department. (2024). *State Education Department presents vision to transform New York State's graduation requirements*. <https://www.nysed.gov/news/2024/state-education-department-presents-vision-transform-new-york-states-graduation>
- New York State Education Department. (2025a). *Board of Regents and Department adopt New York State Portrait of a Graduate*. <https://www.nysed.gov/news/2025/board-regents-and-department-adopt-new-york-state-portrait-graduate>
- New York State Education Department. (2025b). *Frequently asked questions related to New York State graduation measures initiative*. <https://www.nysed.gov/grad-measures/frequently-asked-questions-related-new-york-state-graduation-measures-initiative>
- New York State Education Department. (2025c). *New York State Portrait of a Graduate* [PDF]. <https://www.nysed.gov/sites/default/files/programs/grad-measures/portrait-of-a-graduate.pdf>
- OpenAI. (2023, January 31). *New AI classifier for indicating AI-written text*. <https://openai.com/index/new-ai-classifier-for-indicating-ai-written-text/>
- Perez, A. L., & Evans, C. (2023). Keeping up the PACE: Evaluating grade 8 student achievement outcomes for New Hampshire's innovative assessment system. *Applied Measurement in Education*, 36(2), 137-156. <https://doi.org/10.1080/08957347.2023.2201700>
- Perkins, M., Roe, J., Vu, B. H., Postma, D., Hickerson, D., McGaughan, J., & Khuat, H. Q. (2024). Simple techniques to bypass GenAI text detectors: Implications for inclusive education. *International Journal of Educational Technology in Higher Education*, 21, Article 53. <https://doi.org/10.1186/s41239-024-00487-w>
- Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435-448. <https://doi.org/10.1080/02602930902862859>
- Ritchhart, R., Church, M., & Morrison, K. (2011). *Making thinking visible: How to promote engagement, understanding, and independence for all learners*. Jossey-Bass.
- Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249-255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68-78. <https://doi.org/10.1037/0003-066X.55.1.68>
- Saavedra, A. R., Lock Morgan, K., Liu, Y., Garland, M. W., Rapaport, A., Hu, A., Hoepfner, D., & Haderlein, S. K. (2022). The impact of project-based learning on AP exam performance. *Educational Evaluation and Policy Analysis*, 44(4), 638-666. <https://doi.org/10.3102/01623737221084355>
- Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review*, 78(1), 40-59. <https://doi.org/10.17763/haer.78.1.v62444321p602101>
- Simon, H. A. (1971). Designing organizations for an information-rich world. In M. Greenberger (Ed.), *Computers, communications, and the public interest* (pp. 37-72). The Johns Hopkins Press.
- Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, 87(3), 355-374. <https://doi.org/10.2307/1882010>

- Steele, J. L., Lewis, M. W., Santibañez, L., Faxon-Mills, S., Rudnick, M., Stecher, B. M., & Hamilton, L. S. (2014). *Competency-based education in three pilot programs: Examining implementation and outcomes* (RR-732-BMGF). RAND Corporation. <https://doi.org/10.7249/RR732>
- Steenbergen-Hu, S., Makel, M. C., & Olszewski-Kubilius, P. (2016). What one hundred years of research says about the effects of ability grouping and acceleration on K-12 students' academic achievement: Findings of two second-order meta-analyses. *Review of Educational Research*, 86(4), 849-899. <https://doi.org/10.3102/0034654316675417>
- Sternberg, R. J. (2010). *College admissions for the 21st century*. Harvard University Press.
- Sternberg, R. J., & The Rainbow Project Collaborators. (2006). The Rainbow Project: Enhancing the SAT through assessments of analytical, practical, and creative skills. *Intelligence*, 34(4), 321-350. <https://doi.org/10.1016/j.intell.2006.01.002>
- Students for Fair Admissions, Inc. v. President and Fellows of Harvard College, 600 U.S. 181 (2023).
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257-285. https://doi.org/10.1207/s15516709cog1202_4
- Syverson, S. T., Franks, V. W., & Hiss, W. C. (2018). *Defining access: How test-optional works*. National Association for College Admission Counseling. <https://www.nacacnet.org/globalassets/documents/publications/research/defining-access-report-2018.pdf>
- Tricot, A., & Sweller, J. (2014). Domain-specific knowledge and why teaching generic skills does not work. *Educational Psychology Review*, 26(2), 265-283. <https://doi.org/10.1007/s10648-013-9243-1>
- Troppe, P., Osowski, M., Wolfson, M., Ristow, L., Lomax, E., Thacker, A., & Schultz, S. (2023). *Evaluating the federal Innovative Assessment Demonstration Authority: Early implementation and progress* (NCEE 2023004). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences. <https://ies.ed.gov/use-work/resource-library/report/evaluation-report/evaluating-federal-innovative-assessment-demonstration-authority-early-implementation-and-progress>
- U.S. Department of Education, Office for Civil Rights. (2023). *Dear colleague letter: The Supreme Court's decision in Students for Fair Admissions and considerations for student recruitment, admissions, and retention*. <https://www2.ed.gov/about/offices/list/ocr/letters/colleague-20230814.pdf>
- U.S. Department of Education, Office for Civil Rights. (2024). *Student access to and enrollment in mathematics, science, and computer science courses and academic programs in U.S. public schools: 2020-21 Civil Rights Data Collection*. <https://civilrightsdata.ed.gov/>
- U.S. Department of Education, Office of Elementary and Secondary Education. (2025). *Innovative Assessment Demonstration Authority*. <https://oese.ed.gov/offices/office-of-formula-grants/school-support-and-accountability/innovative-assessment-demonstration-authority/>
- U.S. Government Accountability Office. (2018). *K-12 education: Public high schools with more students in poverty and smaller schools provide fewer academic offerings to prepare for college* (GAO-19-8). <https://www.gao.gov/products/gao-19-8>
- Wiggins, G. (1990). The case for authentic assessment. *Practical Assessment, Research, and Evaluation*, 2, Article 2. <https://doi.org/10.7275/ffb1-mm19>
- William and Flora Hewlett Foundation. (n.d.). *Deeper learning*. Retrieved June 23, 2026, from <https://hewlett.org/strategy/deeper-learning/>
- Willingham, D. T. (2007). Critical thinking: Why is it so hard to teach? *American Educator*, 31(2), 8-19.
- Willis, L., & Martinez, M. R. (2022). *Authentic student work in college admissions: Lessons from the Ross School of Business*. Learning Policy Institute. <https://doi.org/10.54300/756.774>
- Winget, M., & Persky, A. M. (2022). A practical review of mastery learning. *American Journal of Pharmaceutical Education*, 86(10), Article 8906. <https://doi.org/10.5688/ajpe8906>
- Wren, K. (2022, March 28). Q&A: Stuart Schmill on MIT's decision to reinstate the SAT/ACT requirement. MIT News. <https://news.mit.edu/2022/stuart-schmill-sat-act-requirement-0328>
- Yale University. (2024, February 22). *Yale announces new test-flexible admissions policy*. Yale News. <https://news.yale.edu/2024/02/22/yale-announces-new-test-flexible-admissions-policy>
- Zeiser, K. L., Taylor, J., Rickles, J., Garet, M. S., & Segeritz, M. (2014). *Evidence of deeper learning outcomes* (Report No. 3, Study of Deeper Learning: Opportunities and Outcomes). American Institutes for Research. <https://www.air.org/resource/report/evidence-deeper-learning-outcomes>
- Zhang, L., & Ma, Y. (2023). A study of the impact of project-based learning on student learning effects: A meta-analysis study. *Frontiers in Psychology*, 14, Article 1202728. <https://doi.org/10.3389/fpsyg.2023.1202728>