



Assessment Quality in the AI Era: A Conceptual Framework for Academic Management in International Schools

Federico Verri

British International School of Timisoara

*Corresponding Author, e-mail: federico.verri@britishschool-timisoara.ro

Abstract

Assessment quality in international schools is no longer only a technical issue of rubrics, moderation or examination security. Generative AI has made this problem more urgent because polished student products may conceal the learning process that produced them. The research gap addressed in this article is that current AI-assessment debates often focus on detection and integrity violations, whereas less attention is given to the systems through which schools manage assessment quality. This conceptual article synthesizes literature on academic integrity, educational leadership, organizational trust, stakeholder engagement, quality assurance and AI governance. It proposes a six-domain framework for academic management: policy coherence, assessment architecture, staff capability and calibration, AI governance, stakeholder role clarity and quality-assurance feedback loops. The framework is presented as a recursive management cycle rather than a checklist. Its main implication is that international-school leaders should not rely on surveillance or AI-detection tools alone, but should build defensible academic systems that make learning evidence visible, moderation decisions consistent and stakeholder expectations transparent.

Keywords: Academic Management, Assessment Quality, International Schools, Educational Leadership, Artificial Intelligence, Quality Assurance, Educational Administration

Introduction

The quality of assessment has emerged as one of the most challenging management challenges of schools. Many issues continue to be debated in technical terms, like task validity, marking reliability, or moderation processes. These questions are still significant, but they are not enough anymore. Assessment is now practiced in a space defined by heightened parental demands, accountability to universities, mediating digitally, and easy access to generative AI. In this case, the quality of assessment is not merely a classroom issue, but an organizational product. This article therefore treats assessment quality as a broader concept than academic integrity: integrity concerns honest production and rule compliance, whereas assessment quality also includes validity, reliability, moderation, curriculum alignment, evidence sufficiency, procedural fairness and institutional accountability.

This issue can be particularly noticeable at international schools. These schools often serve in more than a single curriculum tradition, have very mobile student bodies and operate in reputational markets where grades, promotion pathways and external examination results are very visible. These environments may be scholarly challenging, nevertheless they may generate stress to defend outcomes, uniformity, or to frame evaluation as a brand-sensitive procedure instead of a provider of credible information regarding learning. As it occurs, academic management is decisive. In practical terms, teachers may now face polished essays, portfolio reflections, presentation scripts or research drafts that have been shaped by AI, private tutoring, parental editing, translation software or language-support tools, making it difficult to see the student's actual learning process.

Generative AI has led to a more pressing need to address this challenge. Learning can be enhanced with the assistance of AI-aided drafting, summarizing, paraphrasing, translation, and ideation, yet it undermines the notion that the final products of work are always the results of independent student comprehension. Human-centred governance of generative AI in education is thus advocated by UNESCO (2023), whereas recent evidence of academic integrity demonstrates that even carefully designed authentic assessment can become less secure when schools are over-reliant on product-based evidence. The issue of leadership is not merely the way to identify AI abuse. It is the manner of controlling the quality of assessment in a manner that is educationally sound, institutionally viable, and professionally justifiable. At the same time, AI can also support academic management when used carefully for assessment-data review, moderation sampling, documentation of policy implementation and professional learning; the issue is therefore governance, not rejection of technology itself.

This paper addresses that challenge by suggesting a conceptual framework of academic management in international schools. It does not give new empirical information. Instead, it provides

a theoretical synthesis of the academic integrity, educational leadership, organizational trust, stakeholder engagement, quality assurance, assessment validity and AI in education literature. It argues that the quality of assessment in international schools should be handled as a whole school system. The article thus seeks to shift the discourse away from reactive control measures and toward a more robust model of policy coherence, assessment architecture, staff calibration, AI governance, stakeholder role clarity and quality assurance. The structure is as follows: conceptual foundations, synthesis method, assessment difficulties in international schools, the proposed framework, leadership implications and recommendations.

Conceptual Foundations: Academic Management and Assessment Quality

Academic management is a more general concept than a narrow bureaucratic understanding of academic administration. Academic management in this article means the coordinated leadership of the curriculum implementation, learning assessment, moderation, policy enactment, professional judgment and quality assurance routines that determine the credibility of the student outcomes. This perspective brings academic management to the cross of educational leadership, the quality of instruction, and the accountability of the institution.

There are three conceptual traditions that aid in explaining the importance of this perspective. The former is based on the academic integrity literature. Bretag (2016) and the International Center of Academic Integrity (ICAI, 2021) not only view integrity as a compliance with rules but also as an ethical culture that is based on honesty, trust, fairness, respect, responsibility, and courage. This framing is important to schools since the quality of assessment is determined by the extent to which the staff, students and families consider academic judgments as authentic. As soon as integrity is understood as an institutional condition, the question becomes not how do we catch misconduct? but how do we structure the school in a way that high-quality assessment has a better chance?

The second tradition is based on teaching leadership and trust in organizations. School research has indicated again and again that trust is not merely interpersonal, but structural. Trust is affected by leaders in terms of procedural fairness, consistency in communication, visible follow-through, and safeguarding of professional judgment. Once employees are convinced that policies will be implemented unequally, hard cases will be watered down or choices will be influenced by external influence, confidence in academic procedures is undermined. In this kind of environment, the quality of assessment becomes compromised even prior to any formal case of misconduct in sight since the quality of the standards is now under attack itself.

The third tradition is related to the management of the quality of education. Assessment quality does not only mean technical precision. It encompasses validity, transparency, consistency, correspondence to the intended learning, and confidence that the reported results represent work that students truly know and can justify. As demonstrated by OECD (2012), the failures in integrity in education usually arise as a result of poor governance, opacity, and misaligned incentives and not necessarily individual malpractice. This observation is all the more applicable in the context of modern online education: not only are assessment systems failing due to students abusing the tools, but because institutions still tend to reward smooth products without gathering enough data on the process, authorship, and judgment.

A fourth conceptual tradition is AI governance in education. AI is a source of assessment risk, but it can also become a quality-management support when it is placed under human-led governance. It may help leaders identify patterns in assessment data, review moderation anomalies, document recurring concerns and plan professional learning. These functions do not replace teacher judgment; they make the management of judgment more explicit.

Conceptual Synthesis Method

Because this article is conceptual rather than empirical, the framework was developed through an integrative conceptual synthesis. The purpose was not to conduct a statistical review or to claim representative findings across all international schools. The purpose was to connect established concepts from related literatures and translate them into a management framework for assessment quality in the AI era.

The synthesis comprised three steps of analysis. First, the problem space was identified at the intersection of assessment quality, AI-enabled student work, and international-school management. Second, literature was organised into five concept clusters: validity of assessments and moderation practice, academic integrity, school leadership and organisational trust, stakeholder communication in international schools, and governance of AI in education. Third, common themes in management functions cut across those clusters. The six domains were maintained because each domain was a necessary precondition for valid assessment evidence and since the domains were mutually dependent rather than siloed.

Table 1 Conceptual synthesis used to derive the framework domains.

Concept cluster	Contribution to framework	Domains informed
Assessment validity, moderation and feedback	Valid evidence, shared standards and feedback-informed improvement (Black & Wiliam, 1998; Sadler, 1989; OECD, 2013).	Assessment architecture; staff calibration; feedback loops
Academic integrity	Integrity as institutional culture, not only rule enforcement (Bretag, 2016; ICAI, 2021; UNESCO IIEP, 2020).	Policy coherence; AI governance; stakeholder role clarity
Leadership and organizational trust	Trust depends on procedural fairness, consistency and communication (Bryk & Schneider, 2002; Leithwood, 2021).	Policy coherence; staff calibration; feedback loops
International-school stakeholder context	Transnational curricula, mobile families, multilingual contexts and market sensitivity (Hayden, 2006; Bunnell et al., 2016; McIntosh & Hayden, 2022).	Stakeholder role clarity; policy coherence; assessment architecture
AI governance in education	AI changes authorship and evidence visibility but can support quality-assurance work under human governance (UNESCO, 2023; European Commission, 2022; Perkins, 2023; Kofinas et al., 2025).	AI governance; assessment architecture; feedback loops

This synthesis process clarifies the status of the framework. It is a theoretically derived model that still requires empirical validation. For this reason, Table 2 includes proposed implementation indicators that future research can test and refine.

Assessment Quality in International Schools in the Age of AI

The quality of assessment in international schools is especially a unique problem due to the fact that a number of pressures interact within the same institutional space. To begin with, they tend to be in very competitive markets. Exposure of performance results, college placements and satisfaction of parents can create high stakes to defend visible results. Such incentives are quite comprehensible in fee-paying environments but may corrupt academic judgment in case the outcomes become more significant than the soundness of the evidence that underlies them.

Second, the international schools often facilitate atypical complicated relationships of stakeholders. Harris (2021) states that leaders in schools need to involve stakeholders in a manner that facilitates learning and resilience of the system. However, McIntosh and Hayden (2022) demonstrate

that parental involvement in international schools tends to leave the traditional premises and be influenced by mobility, aspirational, and school choice. This is important since parent partnership can enhance learning, it can also blur the boundary between support and interference unless leaders understand how to define role boundaries. In assessment, this ambiguity can be manifested by deadline pressure, foreseen grades, accommodations, re-submission, or the handling of AI-assisted work.

Third, most international schools operate across several programmes, standards frameworks and reporting cultures simultaneously. This complexity may lead to inconsistency within the departments, year groups or curriculums. One team might focus on process evidence and oral defence; another can still be primarily dependent on polished products. One part can consider the unauthorized use of AI a learning problem; another can consider it a serious malpractice. In the absence of more robust academic management, these differences pose challenges to the quality of assessments since cases with similarities are handled with different sets of criteria.

Fourth, generative AI alters the evidentiary foundation of evaluation. UNESCO (2023) recognizes AI necessitates governance, capacity building and pedagogical redesign. Kofinas, Tsay and Pike (2025) also demonstrate that the integrity of authentic assessment is compromised when institutions use final products as proxies to education when using generative AI. To school leaders, the main implication is that quality can no longer be ensured through a mere review of products. Quality is now determined by the presence or absence of visibility of the processes of learning in the school assessment architecture to maintain credible judgment.

It's worth comparing to non-AI assessment management. Prior to genAI, there was no guarantee that a complete polished product was accurate/reliable, but it did often have better evidential value. Now fluency, structure, surface accuracy can all be externally or semi-externally produced. QA must therefore shift from product checking to evidence architecture: staged submissions, student-teacher dialogue, in-class itemized production points, source citation/explanation, draft submissions history, oral explanation.

AI may also support management and quality enhancement. Leaders can use AI-supported tools cautiously to review patterns in assessment data, identify moderation anomalies, organize evidence for quality assurance and plan staff development. The risk is automation bias; AI outputs or detection scores should never become the sole basis for academic judgment.

These circumstances imply that the quality of assessment in international schools ought to be the problem of management, not the problem of detection. Detection technologies and plagiarism software and post hoc investigation might also be involved, but they cannot replace coherent academic

management. When conflict exists between the work of a school, the standards of moderation, the training of personnel, the communication of stakeholders, and AI rules, then even when the misconduct is legally penalized the quality assurance is weak.

A Conceptual Framework for Academic Management

This article will offer a conceptual framework of academic management based on six domains on the basis of the literature. The framework does not form a procedural checklist. It is an organizational framework, demonstrating the process of building the quality of assessment by organizing policy, pedagogy, people and feedback systems rather than addressing each element as separate units. The six domains were selected because they connect the five concept clusters in Table 1 and because each domain performs a distinct management function in protecting assessment quality.

The initial area is policy coherence. Schools should have an academically accurate policy framework that relates the curriculum expectations and assessment guidelines, academic integrity standards, and guidance on AI-use. A disjointed policy environment beckons inconsistency. An internally consistent one minimizes discretionary drift and aids leaders to justify decisions in an open manner.

The second area is assessment architecture. The design of tasks should generate evidence, which is more than a final polished product. Independent learning is more visible with staged submission, monitored checkpoints, explanation of sources, brief viva-style discussions, logs of processes and in-class production points which contribute to the quality of judgments.

Staff capability and calibration is the third domain. Quality evaluation relies on collective professional meaning. Middle leaders and teachers need advice about borderline cases, common standards of AI-assisted work, routine of moderation, and the safe space to discuss it. Calibration should not be considered as an optional add-on training but it is a fundamental aspect of quality management.

The fourth sphere is AI governance. Schools must have regulations on what AI may or may not be used, but they must also have a rationale of what is taught about the use of AI. Hyperreaction may drive schools towards hypervigilantly and pedagogically anemic reactions. Underreaction exposes staff and puts standards in jeopardy. Good governance is made up of the disclosure standards, redesign of tasks, due process, and explicit education on ethical usage.

The fifth area is the clarity of the stakeholder roles. The families are supposed to learn the expectations of the school regarding assistance, autonomy, translation, tutoring, editing, and the use of AI. The quality of assessment suffers when parents, students and staff act under different assumptions

of what is viewed as legitimate assistance. The management of academics thus needs to be not only internally policy driven but also externally communicated.

Quality assurance feedback loops is the sixth domain. Isolated incidents should not be used by schools to determine the effectiveness of their assessment systems. They require regular examination of patterns of moderation, integrity issues, the results of the redesign of assessments, employee issues, and new vulnerabilities related to AI. Quality assurance is further enhanced when the leaders utilize this information in revision of policy and practice on a continuous basis.

Collectively, these six domains constitute a management logic: policy influences practice; practice must produce legitimate evidence; personnel require common interpretive norms; AI policies must be educationally plausible; stakeholders require position in the role; and management requires a feedback mechanism to improve the system. Table 1 is a summary of the framework. The relationship is recursive rather than linear: policy enables design, design produces evidence, staff calibration interprets evidence, AI governance sets boundaries, stakeholder clarity protects the assessment environment, and quality-assurance feedback loops return information to policy and practice.

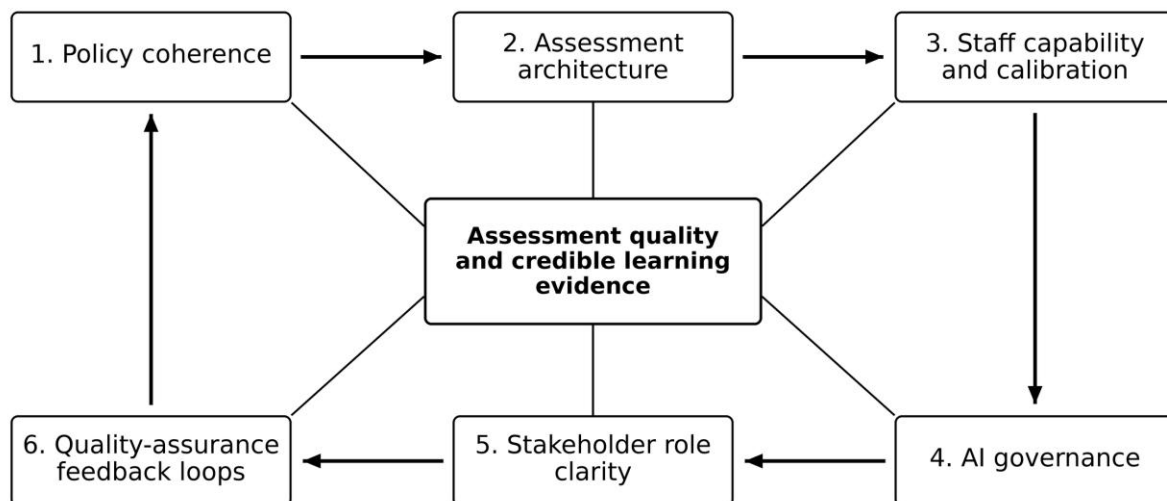


Figure 1 Relational logic of the framework

Table 2 Conceptual framework for academic management to enhance assessment quality.

Framework Domain	Academic Management Action	Expected Contribution to Assessment Quality	Proposed indicators of success
Policy coherence	Align curriculum rules, assessment procedures, academic integrity standards, and AI-use guidance.	Reduces inconsistency, clarifies disclosure expectations, and improves defensibility of decisions.	A policy map exists; AI-use categories are defined; disclosure and appeal procedures are documented; similar cases are treated through comparable criteria.
Assessment architecture	Use staged, evidence-rich, and process-sensitive tasks rather than relying only on polished final products.	Improves validity and makes student learning more visible.	High-risk tasks include checkpoints, in-class evidence, source explanation, draft history or brief oral defence.
Staff capability and calibration	Build moderation routines, case discussion, and shared interpretation of borderline or AI-related cases.	Strengthens consistency of professional judgment across programmes and departments.	Departments hold scheduled calibration meetings; difficult cases are recorded; moderation decisions show reduced inconsistency.
AI governance	Define permitted, limited, and prohibited uses of AI and connect rules to pedagogy, due process, and student disclosure.	Protects fairness while shifting the response from prohibition alone to educational governance.	Students receive AI-use guidance; AI disclosure is used where relevant; detection tools are not used as sole evidence.
Stakeholder role clarity	Clarify the boundaries of support, editing, tutoring, translation, and AI-mediated assistance for families and students.	Protects the independence of assessment and reduces conflict around outcomes.	Parent and tutor guidance is communicated; students understand support boundaries; disputes about assistance are reduced.
Quality assurance feedback loops	Review patterns in moderation, redesign, staff concerns, and integrity incidents as part of routine quality assurance.	Supports continuous improvement and earlier identification of emerging vulnerabilities.	Leaders review trend data; policy revisions are evidence-informed; assessment redesign is evaluated after implementation.

Implications for Educational Leadership and Quality Management

There are a number of implications of the framework to peer-reviewed discussion in educational management. First, it does not focus on misconduct as an event but on quality assessment as a system. This is a significant step as most times international schools react to integrity issues once issues are made apparent. A more robust academic management practice operates at a more designly, expectancy, coordinating, and institutional learning level.

Second, the model contends that international school educational leadership must be evaluated in part based on the credibility of the academic processes as opposed to based on headline performance results only. Leaders can attain short-term relaxations through informal exceptions, ambiguous AI practices, or not consistent responses across programmes. Nevertheless, these decisions undermine the defensibility of the school judgements over time.

Third, the framework asserts quality of assessment as an intermediary between educational innovation and educational accountability. AI does not merely pose risk, but also poses pressure to redesign assessment in a more intelligent manner. Well responding schools will also tend to shift towards evidence based, dialogic and process sensitive forms of assessment. This change can only be interpreted as academic management and quality improvement rather than emergency control.

Fourth, the article brings an explicitly international school viewpoint to the larger educational management literature. A lot of the research on AI and assessment has centered on universities. However, international schools encounter a unique set of transnational curriculums, transnational families, mobile and multilingual students, and market sensitivity. The framework thus carries on the current discussions by defining assessment quality as a sector of the school that academic management is particularly prominent and politically sensitive.

Implementation can also be challenging. Teachers can pushback against new assessment practices if they are already burnt out from a heavy workload. Schools will have inconsistent policies. There can be lack of access to technology or different cultural expectations about family involvement. There is also rapid pressure to shift mindsets for senior teachers who have not adapted to digital changes.

Middle leadership is therefore critical. Assessment quality is often moderated through heads of department, programme coordinators, grade-level leaders and curriculum leaders in international schools rather than (just) through the principal. It is these intermediate leaders who turn whole-school policy into assessment tasks, moderation routines and dialogue between teachers. Without proper

training and empowerment, the framework is nothing more than a statement from senior leadership, rather than a functioning quality-assurance system.

Schools will also need to hire and train teachers with stronger AI competencies. We don't want schools to just teach kids how to write good prompts. Teachers will need to understand how AI alters the evidence of learning for particular subjects and how it can supplement learning in ways that don't substitute for students' understanding.

Trade-offs are present in the framework as well. Gathering more process evidence may enhance validity, but will likely increase administrative burden. Developing clearer policies may enhance consistency, but may decrease teacher autonomy if codified too strictly. AI-supported quality assurance may help leaders see patterns, but may create a false sense of confidence if leaders view analytic outputs as objective. These tensions require governance, not avoidance.

Lastly, the framework welcomes empirical follow up. The future studies might be able to investigate the manifestation of the six domains in other international school settings, how middle leaders understand the quality of assessment in an AI setting, and whether schools of better calibration and policy coherence can have more stable judgments and fewer cases of integrity challenges. The idea of the conceptual article in that regard is not aimed at closure but rather as an entry point to a further enquiry.

Policy, Practice and Research Recommendations

At policy level, international schools should review assessment, academic-integrity and AI-use documents together rather than as separate policies. The review should clarify permitted, limited and prohibited forms of support; define disclosure expectations; protect due process; and specify how decisions are recorded.

At practice level, schools should redesign selected high-stakes tasks so that final products are supported by process evidence. This does not mean that every assessment must become longer. It means that tasks most vulnerable to outsourced or AI-assisted production should include checkpoints, in-class components, source explanation or brief oral defence.

At leadership level, schools should assign explicit responsibility to middle leaders for assessment-quality monitoring. Heads of department and programme coordinators should be supported to review moderation patterns, staff concerns, student disclosure practices and parent communication.

Recommendations for future research. The indicators summarized in Table 2 need to be tested. Do schools that report higher levels of policy coherence around promotion criteria, stronger calibration

routines and more frequent systematic feedback loops also report more consistent judgments and fewer unresolved integrity disagreements? Cross-national research could also explore how the framework functions in international schools across Asia, Europe, Africa and beyond, and across different curriculum systems.

Conclusion

The quality of assessment in international schools is no longer a technical challenge and has to be anchored in a combination of moderation paperwork or post-facto policing. Assessment credibility in the era of artificial intelligence lies in the effectiveness of the relationship between policy, task design, professional judgment, stakeholder expectations and quality assurance as handled in schools. This is especially important in post-generative-AI educational systems, where surface fluency is no longer enough to demonstrate independent learning.

This paper has stated that academic management offers the right perspective of interpreting the challenge. It introduced a six-domain conceptual framework, uniting the literature on academic integrity, organizational trust, stakeholder engagement, educational leadership, and AI governance, which can assist schools in enhancing the quality of their assessment: policy coherence, assessment architecture, staff capability and calibration, AI governance, stakeholder role clarity, and quality assurance feedback loops.

The essence of the argument is simple. The quality of assessment is greatest when the schools promote the visibility of learning, consistency of expectations and defensive decisions. In the case of international schools, it is not a fringe administrative issue. Institutional credibility revolves around it. Academic management in this sense is not merely the management of assessment in an efficient way; it is also the maintenance of the credibility of judgments of education about which the schools rely. The point is ultimately institutional: credible assessment sustains educational legitimacy, institutional credibility and public trust.

References

- 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>
- Bretag, T. (Ed.). (2016). *Handbook of academic integrity*. Springer. <https://doi.org/10.1007/978-981-287-098-8>

- Bryk, A. S., & Schneider, B. (2002). *Trust in schools: A core resource for improvement*. Russell Sage Foundation.
- Bunnell, T., Fertig, M., & James, C. (2016). What is international about International Schools? An institutional legitimacy view. *Oxford Review of Education*, 42(4), 408-423. <https://doi.org/10.1080/03054985.2016.1195735>
- European Commission. (2022). *Ethical guidelines on the use of artificial intelligence and data in teaching and learning for educators*. Publications Office of the European Union. <https://doi.org/10.2766/153756>
- Harris, A. (2021). The relationship between school leaders and education stakeholders. Regional Center for Educational Planning, UNESCO. <https://rcepunesco.ae/en/KnowledgeCorner/WorkingPapers/WorkingPapers/For%20Online%20-%20Harris%20-%20The%20Relationship%20Between%20School%20Leaders%20and%20Education%20Stakeholders%20%281%29.pdf>
- Hayden, M. (2006). *Introduction to international education: International schools and their communities*. SAGE.
- International Center for Academic Integrity. (2021). *The fundamental values of academic integrity* (3rd ed.). <https://www.academicintegrity.org/aws/ICAI/pt/sp/values>
- Kofinas, A. K., Tsay, C. H.-H., & Pike, D. (2025). The impact of generative AI on academic integrity of authentic assessments within a higher education context. *British Journal of Educational Technology*, 56(6), 2522-2549. <https://doi.org/10.1111/bjet.13585>
- Leithwood, K. (2021). *A review of evidence about equitable school leadership*. *Education Sciences*, 11(8), Article 377. <https://doi.org/10.3390/educsci11080377>
- McIntosh, S., & Hayden, M. (2022). Disrupting conventional conceptions of parental engagement: Insights from international schools. *Research in Comparative and International Education*, 17(1), 51-70. <https://doi.org/10.1177/17454999211038423>
- OECD. (2012). *Strengthening integrity and fighting corruption in education*. OECD Publishing. https://www.oecd.org/en/publications/2012/09/strengthening-integrity-and-fighting-corruption-in-education_g1g1fc09.html
- OECD. (2013). *Synergies for better learning: An international perspective on evaluation and assessment*. OECD Publishing. <https://doi.org/10.1787/9789264190658-en>
- Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching & Learning Practice*, 20(2). <https://doi.org/10.53761/1.20.02.07>

- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144. <https://doi.org/10.1007/BF00117714>
- UNESCO. (2023). Guidance for generative AI in education and research. UNESCO. <https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research>
- UNESCO IIEP. (2020). Building an institutional culture of academic integrity. ETICO. <https://etico.iiep.unesco.org/en/building-institutional-culture-academic-integrity>