Cognitive Grounding Before AI Scaling: Why AI-in-Education Systems Need Learning Theory First

Syaamantak Das SYAAMANTAK.DAS@IITB.AC.IN
Centre for Educational Technology, Indian Institute of Technology Bombay, Mumbai, India

Abstract

AI-powered educational tools proliferate globally, reaching hundreds of millions of learners through national education platforms, EdTech products, and AI tutoring systems. A critical gap persists between the pace of AI deployment and the depth of cognitive grounding that underlies these systems. This paper argues that responsible AI for education requires a *cognitive grounding phase* before scaling: a systematic, empirically validated assessment of what existing educational systems actually demand cognitively, and whether AI tools are designed to enhance or merely replicate those demands. Drawing on a decade of empirical research classifying over 3,000 assessment items from national school board examinations using the revised Bloom's Taxonomy, we demonstrate that the dominant examination systems in a major Global South context overwhelmingly test the lowest cognitive levels, precisely the levels at which large language models perform most capably. We introduce the concept of digital rote to describe the phenomenon in which AI systems technologically reproduce pedagogically shallow learning at scale, and propose a three-stage Cognitive Grounding Framework — Cognitive Audit, Cognitive Specification, and Cognitive Validation — developed in the context of LLM-based assessment generation and extensible in principle to other AI-in-education applications. We demonstrate the framework's practical viability through a patented AI-based assessment system whose design was directly informed by cognitive-level analysis. Finally, we discuss policy implications for national-scale AI deployment in education and outline open questions for the responsible AI-in-education community.

Keywords: AI in education, cognitive grounding, Bloom's Taxonomy, assessment quality, responsible AI, learning theory, digital rote, Global South

1. Introduction

The integration of Artificial Intelligence into educational systems is accelerating globally, with particular momentum in the Global South where national-scale deployments aim to address access and quality challenges simultaneously. Government-supported digital platforms in several countries now deliver AI-augmented content — tutoring systems, automated assessment generators, and adaptive learning platforms — to tens or hundreds of millions of learners, at scales unimaginable a decade ago. These deployments are frequently motivated by compelling access arguments: AI can deliver educational content at a cost and scale that human instructors alone cannot match (Thakur and Mittal, 2025).

However, this acceleration creates a fundamental tension. Most AI-in-education systems are designed to optimize technical performance metrics such as accuracy of generated content, engagement duration, task completion rates, or user satisfaction scores, rather than pedagogical quality metrics grounded in learning science. The assumption, often implicit, is that technically proficient AI systems will produce educationally valuable outcomes. This

© 2026 S. Das.

assumption is not warranted, and the evidence suggests it may be actively harmful in contexts where the underlying educational system is itself pedagogically misaligned.

This paper articulates a specific diagnosis of this problem and proposes a framework for addressing it. Our central argument is that responsible AI deployment in education requires what we term a cognitive grounding phase — a systematic, empirically validated assessment of the cognitive demands of the educational system into which AI is being integrated, conducted before AI tools are deployed at scale. We develop this argument primarily through the lens of LLM-based assessment generation, which is the AI-in-education application most directly aligned with the cognitive-level concerns we raise, and we sketch in Section 7 how the framework would extend to other AI-in-education modalities. Without this grounding, AI risks becoming what we call *digital rote*: a technologically sophisticated mechanism for reproducing the same low-order cognitive demands that traditional educational systems already over-emphasize, now delivered at unprecedented scale and with the veneer of technological innovation.

2. Related Works

Our work connects three streams of research: cognitive taxonomies in educational assessment, responsible AI frameworks, and the emerging literature on AI in education.

2.1. Cognitive Taxonomies and Assessment Quality

The revised Bloom's Taxonomy (Anderson and Krathwohl, 2001) provides a six-level hierarchy of cognitive processes — Remember, Understand, Apply, Analyze, Evaluate, and Create — that has been widely used to classify educational objectives and assessment items. Webb's Depth of Knowledge (Webb, 1997) and the SOLO Taxonomy (Biggs and Collis, 2014) offer alternative frameworks with somewhat different emphases but a shared commitment to distinguishing levels of cognitive demand. These taxonomies have been applied extensively in curriculum design and assessment quality research, but their application to AI system design remains rare. Our contribution is to argue that cognitive taxonomies should serve not merely as analytical tools for evaluating existing assessments but as design constraints for AI systems that generate, curate, or evaluate educational content.

2.2. Responsible AI Frameworks

The responsible AI literature has developed extensive frameworks for fairness auditing (Mehrabi et al., 2021), bias detection (Suresh and Guttag, 2021), and impact assessment in AI systems. The EU AI Act[1] classifies educational AI as "high-risk," requiring conformity assessments and human oversight. However, these frameworks focus primarily on statistical fairness (e.g., ensuring that AI systems do not discriminate across demographic groups) and procedural accountability (e.g., requiring human-in-the-loop review). What they do not address is pedagogical quality, i.e., the question of whether AI-generated educational content promotes the cognitive processes that learning science indicates are most valuable.

1. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Our cognitive grounding framework fills this gap by providing a learning-science-specific dimension of responsible AI evaluation, complementary to existing fairness and accountability mechanisms.

2.3. AI in Education

The AI-in-education community has produced substantial work on intelligent tutoring systems (VanLehn, 2011), automated question generation (Kurdi et al., 2020), and adaptive learning platforms. Recent work has focused on the integration of large language models into educational tools, with both optimistic assessments of their potential (Kasneci et al., 2023) and concerns about their limitations (Becker et al., 2023). However, much of this literature evaluates AI systems against technical benchmarks (e.g., fluency of generated text, accuracy of answers) rather than cognitive-level criteria. A notable exception is the literature on knowledge tracing and item response theory (Corbett and Anderson, 1994), which models learner cognition explicitly. Our contribution extends this tradition by arguing that the cognitive profile of the educational system (not just the learner) must be understood and specified before AI tools are integrated into it.

3. The Cognitive Grounding Gap

3.1. What Educational Systems Actually Test: Empirical Evidence

The premise of many AI-in-education interventions is that they improve upon the existing system. But improving upon a system requires understanding what that system currently does — not in aspirational terms, but in empirically verified terms. This understanding is often absent. To illustrate this gap, we draw on a programme of research that conducted large-scale cognitive-level analysis of educational assessments in a major Global South context. Over 3,000 assessment items from multiple national-level school board examinations in STEM subjects (Physics, Chemistry, Biology, and Mathematics) were manually classified using the revised Bloom's Taxonomy. The classification covered examinations administered between 2011 and 2020, spanning three national school boards that together serve the majority of the country's secondary school students (Das et al., 2022; Chowdhury et al., 2023).

The cognitive-level distribution revealed a stark pattern. Across boards and subjects, the two lowest cognitive levels — *Remember* and *Understand* — accounted for the dominant share of assessment items. Higher-order cognitive levels — *Analyze*, *Evaluate*, and *Create* — were rare, typically representing fewer than 15-20% of total items, and in some subjects and boards, fewer than 10%. The distribution was not uniform across subjects: Mathematics tended to include somewhat more Apply-level items (procedural problem-solving), while Biology and Chemistry were heavily skewed toward Remember (factual recall of terminology, definitions, and processes). The pattern was also not uniform across boards: some boards showed marginally higher representation of Analyze-level items, but no board consistently tested at the upper cognitive levels across subjects (Chowdhury et al., 2023).

These findings have been validated through complementary methodological approaches, including multi-class text classification of assessment questions (Das et al., 2020) and empirical classification of the action verbs used in learning objectives and examination questions (Das et al., 2022). The convergence of findings across methods strengthens the conclu-

sion: the dominant assessment systems in this context are empirically calibrated to test lower-order cognition.

3.2. When AI Meets a Cognitively Shallow System

The empirical findings above acquire urgent significance in the context of generative AI. Recent research has demonstrated that large language models can perform remarkably well on standardized educational assessments. In a study evaluating GPT-3.5, GPT-4, and Google Bard on a major national premedical entrance examination, one of the highest-stakes gateway exams in the country, taken by over two million aspirants annually, GPT-4 scored 42.9% of total marks, well above the General-category qualifying threshold of approximately 19%, while GPT-3.5 (20.7%) just cleared it and Bard (16.4%) did not (Farhat et al., 2024). While this performance does not represent mastery (the model did not score in the top percentiles), it demonstrates that existing LLMs can already clear the qualifying bar of a high-stakes examination designed for human test-takers.

This finding is often framed as a threat to academic integrity — the worry that students will use AI to cheat on exams. But the more fundamental implication is about what the examination is measuring. If an LLM, which processes language statistically rather than understanding concepts cognitively, can clear a qualifying threshold, then the examination must be testing cognitive tasks that are achievable through statistical language processing, primarily recall-level retrieval and pattern-based comprehension. The AI is not demonstrating understanding; it is demonstrating that the examination does not require understanding.

3.3. Digital Rote: A Systemic Risk

The combination of these two findings — (i) educational systems predominantly test lower-order cognition, and (ii) AI systems are optimized to perform at these same cognitive levels — creates what we term *digital rote*.

Digital rote is the phenomenon in which AI systems technologically reproduce pedagogically shallow educational experiences at scale. An AI question generator trained on recall-level questions will produce more recall-level questions; an AI tutor trained on factual rather than analytical explanations will generate informative but cognitively shallow content; and an adaptive platform optimizing for task completion will route learners toward the quickest, lowest-demand tasks.

Digital rote is distinct from, and arguably more concerning than, other recognized risks of AI in education such as bias, hallucination, or privacy violations. These risks are failures of the AI system; they occur when the system does something wrong. Digital rote, by contrast, occurs when the AI system does exactly what it is designed to do: faithfully reproduce the cognitive profile of its training data, which in turn reflects the cognitive profile of the educational system it was trained on. The system is not malfunctioning; it is functioning perfectly, within a pedagogically impoverished specification. This makes digital rote difficult to detect through standard technical evaluation methods: the AI's outputs are accurate, fluent, and curriculum-aligned, but they are also cognitively shallow. Only an evaluation grounded in cognitive taxonomy, an evaluation that asks not just "*is this correct?*" but "*at what cognitive level does this operate?*", can identify the problem.

4. A Cognitive Grounding Framework

We propose a three-stage Cognitive Grounding Framework for AI-in-education systems. The framework was developed in the context of LLM-based assessment generation — where the cognitive-level question is most directly tractable — but its three-stage logic of audit, specification, and validation is, in principle, extensible to other contexts where AI tools generate, curate, or evaluate educational content, including adaptive tutoring systems, AI-supported reading tools, and simulation-based learning environments. We discuss this extensibility in Section 7. The framework is designed to complement existing responsible AI practices (fairness auditing, bias testing, impact assessment) by adding a learning-science-specific dimension that addresses the digital rote risk.

4.1. Stage 1: Cognitive Audit

Before deploying AI in any educational context, conduct an empirical audit of the cognitive demands of the existing system. This involves classifying a representative sample of assessment items, learning objectives, and content materials using an established cognitive taxonomy. The output of the audit is a cognitive profile, i.e., a quantitative map describing what cognitive levels the system currently emphasizes and what it neglects. Prior work has demonstrated that such audits are methodologically feasible at large scale: the classification of 3,000+ items described in Section 3.1 was conducted over a multi-year period using a codebook-based approach with validated inter-rater reliability (Das et al., 2020, 2022). The cognitive profile reveals systemic patterns that are invisible to individual curriculum designers, examiners, or teachers working within the system; patterns that emerge only when the full distribution of cognitive demands across subjects, years, and boards is made visible.

The Cognitive Audit is analogous to a fairness audit in the responsible AI literature: just as a fairness audit examines a system's behavior across demographic groups to identify disparate impact, a Cognitive Audit examines an educational system's demands across cognitive levels to identify taxonomic skew, establishing the cognitive baseline that AI tools will either reinforce or reshape.

4.2. Stage 2: Cognitive Specification

Based on the audit, specify the cognitive levels that the AI system should target. This specification should be informed by learning science and curricular goals, not merely by the existing cognitive profile.

This is a crucial distinction. If the Cognitive Audit reveals that 80% of existing assessment items test Remember and Understand, the Cognitive Specification should not instruct the AI system to generate content at the same distribution. To do so would be to encode digital rote by design. Instead, the specification should define a target cognitive distribution aligned with the educational objectives of the context, e.g., at least 40% of AI-generated questions target Apply, Analyze, Evaluate, or Create levels, even if the existing system does not. The Cognitive Specification serves as a design constraint on the AI system. It is analogous to fairness constraints in responsible AI (e.g., requiring that a model's false positive rate not differ by more than a specified threshold across demographic groups). In this case,

the constraint is pedagogical: the AI system's output space is bounded by cognitive-level requirements that ensure educational quality, not just technical accuracy. This specification should be developed collaboratively — involving curriculum designers, learning scientists, and domain educators rather than being set by AI engineers alone, because the question of what cognitive levels education should target is fundamentally a pedagogical and curricular question, not a technical one.

4.3. Stage 3: Cognitive Validation

After deployment, continuously validate whether the AI system's outputs conform to the Cognitive Specification. This requires applying the same classification methodology used in the audit to a sample of AI-generated content at regular intervals, comparing the cognitive profile of AI outputs against the specification, identifying drift or deviation, and iterating on the system design.

Cognitive Validation implements the "continuous evaluation" paradigm, the shift from one-time assessment to ongoing monitoring that keeps pace with rapidly changing AI systems. But it grounds this paradigm specifically in learning-theoretic criteria rather than purely technical performance metrics. A system might maintain high accuracy and engagement scores while drifting toward lower cognitive levels in its generated content, a drift that only cognitive-level monitoring would detect.

The three stages are designed to be cyclical: as curricula evolve, as educational goals change, and as AI systems are updated, the Cognitive Audit is repeated, the Specification is revised, and the Validation continues. This cyclical design ensures that cognitive grounding is not a one-time checkbox but an ongoing process embedded in the AI system's lifecycle.

5. From Framework to System: Cognitive Grounding in Practice

The Cognitive Grounding Framework is not merely theoretical. We demonstrate its practical application through the design of an AI-based assessment system developed over several years of research.

The system, Q-GENius, was designed to address a specific limitation of conventional AI question generators: their inability to diagnose the misconceptions behind errors. The system was developed for secondary-school STEM students preparing for national board examinations in India, and has been granted Indian Patent No. 543855 for its method of generating cognitive-level-diagnostic multiple-choice items. The system uses a large language model to generate modified multiple-choice questions in which the distractor options (incorrect answers) are not random but are systematically designed to correspond to specific cognitive-level deficiencies. When a learner selects a particular incorrect option, the system can diagnose not only that the learner's answer is wrong, but at what cognitive level the learner's understanding breaks down, e.g., whether the learner can recall a fact but cannot apply it, or can apply a procedure but cannot analyze the underlying principles.

The design of this system was directly informed by the Cognitive Audit methodology. Years of empirical work classifying assessment items by cognitive level, analyzing the action verbs used in educational objectives at different cognitive levels (Das et al., 2022), and identifying the textual and structural features that distinguish questions at different cognitive

levels (Das et al., 2020) provided the cognitive grounding that shaped the system's architecture. The system's question generation module is constrained by a Cognitive Specification that requires generated questions to target specified cognitive levels. And the system includes a validation component that classifies generated questions against the specification and flags deviations.

The patent grant confirms that the approach has sufficient novelty, non-obviousness, and practical applicability to warrant intellectual property protection. Initial evaluation at AIED 2023 demonstrated the system's ability to generate cognitively targeted questions that align with the specified cognitive levels (Prakash et al., 2023). The key lesson from this practical experience is that cognitive grounding is not an abstract ideal but an engineering discipline — one that adds complexity to AI development but is the price of ensuring that AI-in-education systems serve learning rather than merely serve content.

6. Policy Implications: Scaling Without Grounding

AI-in-education systems are increasingly deployed at national scale, reaching tens or hundreds of millions of learners through government-supported platforms. These deployments typically lack any mechanism for ensuring the cognitive quality of the content they deliver. There are no regulatory frameworks in most countries that require cognitive-level assessment of AI-generated educational content before deployment. No equivalent of a fairness audit exists for pedagogical quality. Content is evaluated, if at all, for factual accuracy and curriculum alignment, necessary but insufficient criteria that do not address the cognitive-level question.

The result is a governance vacuum. AI-powered educational tools are making pedagogical decisions, determining what counts as a good explanation, what constitutes an appropriate difficulty level, how student responses should be evaluated, without any cognitive grounding constraining those decisions. These are curricular decisions being made, in effect, by language model training distributions rather than by educators informed by learning science. Research has documented the risks this creates, including unresolved concerns about accuracy, intellectual property, and educational malpractice in AI-powered learning tools (Ahuja and Babita, 2023).

The Cognitive Grounding Framework proposed in this paper offers a concrete, actionable mechanism for responsible AI governance in education: require a Cognitive Audit before deployment, mandate Cognitive Specifications as part of the AI system design process, and institute continuous Cognitive Validation as a regulatory standard. These requirements are implementable within existing educational governance structures; ministries of education, curriculum bodies, and examination boards already possess the domain expertise needed to specify target cognitive levels. What is missing is the institutional expectation that AI systems should be held to cognitive-level standards, not just accuracy standards.

7. Limitations and Open Questions

We acknowledge several limitations and identify open questions for the community.

- **Taxonomic scope** Our empirical work and framework are grounded primarily in the revised Bloom's Taxonomy, which is the most widely used cognitive taxonomy

in educational assessment. However, Bloom's Taxonomy has known limitations, including ambiguity in the boundaries between cognitive levels (particularly between Understand and Apply) and challenges in application to non-STEM disciplines where cognitive processes may not map cleanly onto the six-level hierarchy. Future work should investigate whether the Cognitive Grounding Framework can be adapted to alternative taxonomies such as Webb's Depth of Knowledge or the SOLO Taxonomy, and whether hybrid approaches that draw on multiple taxonomies could provide more robust cognitive profiles.

- **Scalability of cognitive classification** The Cognitive Audit stage, as implemented in our research, relied on manual expert classification of assessment items, a process that is rigorous but labor-intensive. For the framework to be practical at the national scale, automated or semi-automated cognitive-level classification is necessary. Prior work has explored the use of natural language processing techniques for automated classification of educational texts by cognitive level (Das et al., 2020), with promising but imperfect results. Advances in large language models may enable more accurate automated classification, but this creates a recursive challenge: using AI to audit the cognitive quality of AI-generated educational content requires confidence that the auditing AI itself can reliably distinguish cognitive levels. This is an important direction for future research.

- **Generalizability across contexts** Our empirical evidence comes from a single national context, and the cognitive profiles of educational systems vary across countries, curricula, and levels of education. While we believe the framework is generalizable in principle, any educational context can benefit from an empirical understanding of its cognitive demands before AI integration; the specific findings (the dominance of recall-level assessment) may not apply universally. Comparative cognitive-level audits across educational systems would strengthen the empirical foundation of the framework and reveal the extent to which digital rote is a global rather than context-specific risk.

- **Generalizability across AI-in-education modalities** The framework is most fully developed in the context of LLM-based assessment generation, where the cognitive-level question is closely aligned with the structure of the AI's output. Extending the framework to other modalities raises modality-specific questions. For adaptive tutoring systems, the Cognitive Audit would need to characterize the cognitive-level profile of recommended learning paths rather than discrete items, and the Cognitive Specification would constrain the routing logic to ensure that adaptive sequences include higher-order cognitive challenges rather than optimizing solely for completion. For AI-supported reading tools, the Cognitive Audit would need to characterize the cognitive demands of the comprehension scaffolds and questions the system generates, with the Cognitive Specification ensuring that scaffolds promote analytical engagement rather than surface-level extraction. Systematic demonstration of the framework's applicability across these modalities is an important direction for future work.

- **The Cognitive Specification as a contested space** We have argued that the Cognitive Specification should be informed by learning science rather than by the

existing cognitive profile. But what the "right" cognitive distribution is for any given educational context is not a purely scientific question; it is a curricular and political question, involving trade-offs between competing educational goals (e.g., breadth of knowledge coverage versus depth of analytical skill), resource constraints, and stakeholder interests. The framework identifies the Cognitive Specification as a necessary step but does not resolve the normative question of what it should contain. This is by design: we believe this question should be resolved through deliberative, participatory processes involving educators, policymakers, and communities, not through technical prescription.

- **Interaction between cognitive grounding and other responsible AI dimensions** Our framework focuses specifically on cognitive quality, but deployed AI-in-education systems must simultaneously satisfy fairness, privacy, safety, and accessibility requirements. The interaction between cognitive grounding and these other dimensions is largely unexplored. For instance, *does optimizing for higher cognitive levels inadvertently create accessibility barriers for learners with certain disabilities? Does cognitive-level targeting interact with demographic fairness, i.e., are some cognitive levels more difficult to achieve equitably across student populations?* These intersections represent an important frontier for responsible AI-in-education research.

8. Conclusion

Responsible AI deployment in education requires cognitive grounding: an empirically validated understanding of educational demands and a way to ensure AI enhances rather than reproduces them. To name the risk of deploying AI without this foundation, we introduce the term digital rote, and we propose a three-stage Cognitive Grounding Framework — Audit, Specification, and Validation — to precede AI deployment, demonstrating its practicality through a patented AI-based assessment system and examining its implications for national-scale AI governance. The central point is simple: applying AI to a flawed educational system without fixing its flaws simply automates those flaws. We therefore argue that cognitive grounding should become standard practice in AI-in-education research and deployment, informing system design from the start. The key question is not whether AI can scale educational delivery, but whether the education it delivers is genuinely valuable.

References

Amit Ahuja and Babit Babita. Fairness by awareness: Evaluating AI-powered learning resources and legal repercussions in India. *Educational Quest*, 14(3):175–180, 2023.

Lorin W Anderson and David R Krathwohl. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: complete edition*. Addison Wesley Longman, Inc., 2001.

Brett A Becker, Paul Denny, James Finnie-Ansley, Andrew Luxton-Reilly, James Prather, and Eddie Antonio Santos. Programming is hard-or at least it used to be: Educational opportunities and challenges of ai code generation. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*, pages 500–506, 2023.

John B Biggs and Kevin F Collis. *Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome)*. Academic press, 2014.

Anirban Roy Chowdhury, Vijay Prakash, Syaamantak Das, et al. A comparative analysis of the cognitive levels of science and mathematics secondary school board examination questions in India. In *Proceedings of the 16th International Conference on Educational Data Mining*, pages 509–514, 2023.

Albert T Corbett and John R Anderson. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, 4(4):253–278, 1994.

Syaamantak Das, Shyamal Kumar Das Mandal, and Anupam Basu. Identification of cognitive learning complexity of assessment questions using multi-class text classification. *Contemporary Educational Technology*, 12(2):ep275, 2020.

Syaamantak Das, Shyamal Kumar Das Mandal, and Anupam Basu. Classification of action verbs of Bloom's Taxonomy cognitive domain: An empirical study. *Journal of Education*, 202(4):554–566, 2022.

Faiza Farhat, Beenish Moalla Chaudhry, Mohammad Nadeem, Shahab Saquib Sohail, and Dag Øivind Madsen. Evaluating large language models for the national premedical exam in India: comparative analysis of GPT-3.5, GPT-4, and Bard. *JMIR Medical Education*, 10:e51523, 2024.

Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, et al. ChatGPT for good? on opportunities and challenges of large language models for education. *Learning and individual differences*, 103:102274, 2023.

Ghader Kurdi, Jared Leo, Bijan Parsia, Uli Sattler, and Salam Al-Emari. A systematic review of automatic question generation for educational purposes. *International journal of artificial intelligence in education*, 30(1):121–204, 2020.

Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)*, 54 (6):1–35, 2021.

Vijay Prakash, Kartikay Agrawal, and Syaamantak Das. Q-GENius: A GPT based modified mcq generator for identifying learner deficiency. In *International Conference on Artificial Intelligence in Education*, pages 632–638. Springer, 2023.

Harini Suresh and John Guttag. A framework for understanding sources of harm throughout the machine learning life cycle. In *Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, pages 1–9, 2021.

Bharti Thakur and Neeru Mittal. Role of the united nations educational, scientific and cultural organization (UNESCO), the organization for economic cooperation and development (OECD) and global edtech standards in shaping artificial intelligence and cyber law curriculum policies: A comparative study of Finland, Singapore and India. *International Annals of Criminology*, 63(4):754–767, 2025.

Kurt VanLehn. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational psychologist*, 46(4):197–221, 2011.

Norman L Webb. Criteria for alignment of expectations and assessments in mathematics and science education. research monograph no. 6. 1997.