UEJTL: Universal Education Journal of Teaching and Learning
E-ISSN: 3047-8235
2026, Vol. 3, No. 3, pp. 337~342
DOI: https://doi.org/10.63081/uejtl.v3i3.283

Book Review

Artificial Intelligence for STEM Education Research: Advanced Methods and Applications

Zhai, X., & Lee, G. (Eds.). (2026). Springer Cham, xv + 258 pp., ISBN 978-3-032-06564-3 (hbk), ISBN 978-3-032-06567-4 (sbk), ISBN 978-3-032-06565-0 (eBook).

Aziz Şükrü Yeldan
Department of Computer and Instructional Technologies Education, Gazi University, Turkey

Article Info

Article history:

Received 10 31, 2026
Revised 11 15, 2026
Accepted 11 15, 2026

This is an open access article under the CC BY-SA license.

Corresponding Author:

Aziz Şükrü Yeldan
Department of Computer and Instructional Technologies Education, Gazi University
Gazi Üniversitesi Rektörlüğü, Emniyet, Bandırma Cad. No:6/1, 06560 Yenimahalle/Ankara, Türkiye
Email: azizsukruyeldan@gmail.com

1. INTRODUCTION

The growing use of artificial intelligence (AI) in science, technology, engineering, and mathematics (STEM) education raises a methodological question: how can computational systems extend educational inquiry without weakening the validity of its interpretations? Automated scoring, learner modelling, image analysis, and adaptive instructional systems can process evidence at scales difficult to achieve through manual analysis. Yet predictive performance does not, by itself, establish that an algorithm measures the intended educational construct or improves learning. Researchers must connect technical decisions to substantive learning theories, credible validation procedures, and ethical responsibilities.

Zhai and Lee (2026) bring together contributions to this methodological intersection in their edited volume, Artificial Intelligence for STEM Education Research: Advanced Methods and Applications. Rather than treating AI principally as a classroom product, the contributors approach it as a collection of research methods and design resources. This review examines the volume's conceptual integration, treatment of research evidence, practical utility, and capacity to distinguish promising demonstrations from educationally validated applications. Its central assessment is that the book offers a valuable methodological entry point, although the evidential strength and theoretical integration of its examples vary considerably. In particular, the review considers how effectively the chapters connect computational techniques with educational constructs, methodological validity, instructional decision-making, and responsible research practices. This perspective is important because the contribution of AI

to STEM education should be evaluated not simply through technical innovation but also through the quality and educational significance of the evidence it produces.

2. OVERVIEW OF THE BOOK

Published by Springer as the inaugural volume of Advances in Technology-Rich Science Education, this open-access edited collection comprises eleven chapters, with ten application-oriented chapters following an introduction. Its primary audience includes graduate researchers, educators, and scholars seeking to adapt computational approaches to STEM inquiry. The editors explicitly favour accessible explanations, annotated implementations, datasets, and modular reading over an exhaustive technical treatment of any single algorithm (pp. 1–13).

Taken together, the chapters cover a broad methodological progression, beginning with automatic scoring and AI-generated assessment items, followed by adaptive learning, computerized adaptive testing, computer vision, knowledge tracing, intelligent tutoring systems, recommender systems, sentiment analysis, and explainable AI. This progression reflects the different stages at which AI can contribute to educational research, from collecting and interpreting learning evidence to estimating learner characteristics and informing instructional decisions. Rather than presenting these methods as independent technological innovations, the volume implicitly illustrates their potential interdependence in the development of evidence-informed STEM education systems.

An important characteristic of the collection is its combination of conceptual explanations and practical methodological illustrations. Several chapters provide examples of algorithms, simulation procedures, datasets, and implementation resources. This makes the volume accessible to readers who are beginning to explore AI-based research methods while also offering more experienced researchers opportunities to examine alternative computational approaches. Nevertheless, the extent to which these resources can be translated into actual educational settings depends on the quality of the underlying data, the validity of the measurement procedures, and the appropriateness of the instructional context.

The book's coverage is consistent with the broader development of AI-supported STEM education research. In their systematic review of 63 empirical studies published between 2011 and 2021, Xu and Ouyang (2022) identified six categories of AI applications in STEM education and emphasized the importance of considering the relationships between AI technologies, subject matter, information, learning media, and educational environments. Their findings highlight that the integration of AI into STEM education requires more than the adoption of computational technologies; it also demands consideration of the complex pedagogical and contextual elements that shape teaching and learning. From this perspective, the volume provides a relevant methodological contribution by addressing AI applications across multiple dimensions of STEM education research, although the degree of integration among these dimensions varies across its chapters.

3. REVIEW OF THE BOOK'S CONTENT

Chapter 1 establishes the book's research-centred purpose and introduces its ten methodological domains. Chapter 2 then develops a scoring workflow encompassing data collection, preprocessing, algorithm selection, validation, and deployment. Its accounts of science-specific language-model pretraining, fine-tuning, demographic fairness, and knowledge distillation make clear that scoring quality depends on both the training corpus and the definition of achievement. Chapter 3 moves upstream from scoring responses to generating assessment items. Its culturally responsive science assessment generator illustrates the importance of educator input, while a medical physiology comparison introduces a valuable counterweight to technological optimism. Only 22.5% of the AI-generated questions met all three specified psychometric quality criteria, compared with 47.5% of faculty-generated questions (pp. 51–56). These findings support expert-mediated assessment development, not automated substitution for expert item writers.

Read together, Chapters 2 and 3 highlight an important distinction between automating assessment processes and establishing assessment validity. Whereas Chapter 2 emphasizes the computational procedures involved in evaluating student responses, Chapter 3 demonstrates that the generation of assessment items does not automatically guarantee their psychometric quality. The comparison between AI-generated and faculty-generated questions reinforces the need to examine item characteristics, construct alignment, and expert judgment before assessment instruments are used to make educational decisions. Consequently, these chapters offer a useful methodological reminder that efficiency and validity are complementary but distinct requirements in AI-supported assessment. This distinction is also consistent with the broader findings of Zawacki-Richter et al. (2019), whose systematic review identified assessment and evaluation, adaptive systems and personalization, profiling and prediction, and intelligent tutoring systems as major areas of AI applications in higher education. The authors also highlighted the need for stronger connections between AI applications and pedagogical theories. Accordingly, the book's emphasis on psychometric quality and expert involvement is particularly important because it reinforces the principle that technological efficiency must remain aligned with established educational and measurement principles.

Chapters 4 and 5 examine distinct forms of adaptation. Gao and Zhang's Chapter 4 uses simulated learners to test reinforcement-learning strategies, including Q-learning and an actor–critic approach. The simulations connect instructional sequencing to estimates of learner mastery while showing sensitivity to measurement error and assessment length (pp. 73–80). Chapter 5, by Wang, Huang, and Chang, locates computerized adaptive testing within item response theory and cognitive diagnostic modelling. Its simulation finds that a 24-item diagnostic adaptive test can achieve profile-classification accuracy comparable to a 32-item linear test, although differences in item exposure complicate design decisions (pp. 97–105). Together these chapters demonstrate why adaptivity requires an explicit measurement model rather than an algorithmic recommendation alone.

The relationship between these chapters is particularly relevant to the methodological development of personalized STEM learning. Chapter 4 focuses on how reinforcement-learning algorithms can recommend learning materials according to estimated learner knowledge states, whereas Chapter 5 examines how adaptive assessment can efficiently classify learners' knowledge profiles. Both demonstrate the importance of measurement accuracy in adaptive systems. However, their simulation-based findings should be interpreted as evidence of algorithmic performance under specified conditions rather than direct evidence of improved learning outcomes in authentic classrooms. This distinction is essential when researchers consider transferring simulated models into educational practice. This interpretation is also supported by the broader literature on artificial intelligence in education. Chiu et al. (2023), through their systematic literature review, examined the opportunities, challenges, and future research directions associated with AI applications across educational domains and learning outcomes. Their findings demonstrate that the potential of AI in education should be considered alongside the challenges of implementation and the need to establish meaningful educational outcomes. In this regard, the simulation-based results presented in Chapters 4 and 5 are valuable for examining algorithmic performance, but their implications for actual learning improvement require further empirical investigation in authentic educational environments.

Chapter 6 expands the evidential repertoire to drawings, classroom images, and laboratory scenes. Lee and Zhai compare convolutional neural networks with vision–language models and demonstrate image-based classroom interpretation, laboratory safety assessment, and rubric-guided analysis of a draw-a-scientist response. Their discussion of overfitting and the distinction between automation and triangulation usefully cautions against treating plausible visual interpretations as validated observations (pp. 120–137). Chapter 7 connects knowledge tracing to ACT-R and mastery learning, then compares Bayesian knowledge tracing, performance factors analysis, and deep knowledge tracing using educational interaction datasets. The reported cross-validation results favour the simpler Bayesian model in the illustrated settings, an instructive reminder that model complexity does not guarantee superior prediction (pp. 161–165).

These chapters extend the discussion from assessment results to the interpretation of more complex forms of learning evidence. Chapter 6 demonstrates the potential of visual data for examining classroom activities and student representations, while Chapter 7 focuses on the development of computational models that estimate changes in learners' knowledge over time. Their combined contribution lies in illustrating how different forms of educational data can inform research. At the same time, they raise an important methodological consideration: the accuracy of computational interpretation must be established in relation to the educational constructs being investigated, rather than inferred solely from a model's predictive performance.

Chapters 8 and 9 translate learner modelling into instructional decisions. Chapter 8 examines intelligent tutoring systems through the domain, student, pedagogical, and interface components, with examples from ALEKS, PhysicsAssistant, and Inq-ITS. These cases display different combinations of feedback, laboratory assistance, and inquiry scaffolding. Chapter 9 differentiates content-based, collaborative-filtering, and large-language-model recommender systems. Its small worked examples, including video selection and linear-algebra module sequencing, clarify implementation but should not be mistaken for evidence that recommendations improve long-term learning (pp. 200–204). The relationship among Chapters 4, 7, 8, and 9 is especially important: estimates of knowledge become educationally consequential only when valid instructional decisions follow from them.

More broadly, these chapters reveal a connection between learner modelling and pedagogical intervention. Knowledge estimates generated through adaptive learning and knowledge-tracing approaches can inform the selection of learning materials, the provision of feedback, and the design of instructional support. However, an accurate prediction of learner performance does not necessarily establish that a particular instructional recommendation is pedagogically appropriate. The effectiveness of intelligent tutoring and recommender systems therefore depends on the integration of computational predictions with sound instructional design, disciplinary knowledge, and educators' professional judgment. This is one of the most important cross-chapter insights offered by the volume.

Chapter 10 introduces sentiment analysis through lexicon-based, conventional machine-learning, deep-learning, and language-model approaches. Its examples of student feedback and educators' views of ChatGPT demonstrate how sentiment classification can inform inquiry into engagement and technology adoption, while raising familiar problems of interpretation, bias, and data provenance. Chapter 11 concludes with intrinsically

interpretable models, counterfactual explanations, feature attributions, and prompts. The chapter acknowledges that plausible language-model outputs can remain incorrect and that explanation techniques have distinct limitations (pp. 241–257). Its placement at the end emphasizes that intelligibility and researcher oversight should accompany the book's earlier applications.

The final two chapters provide an important dimension to the volume's overall argument by addressing the interpretation of AI-generated evidence and the transparency of computational decisions. Sentiment analysis illustrates how unstructured feedback can be transformed into research evidence, whereas explainable AI addresses the need to understand and communicate the reasoning behind model outputs. Their inclusion suggests that the responsible application of AI in STEM education requires more than accurate predictions: researchers must also consider how data are interpreted, how model limitations are communicated, and how human oversight is maintained throughout the research process.

The emphasis on explainability is particularly relevant to the development of trustworthy AI-supported educational research. Khosravi et al. (2022) proposed the Explainable Artificial Intelligence in Education (XAI-ED) framework, which identifies six important dimensions of explainability, including stakeholders, potential benefits, explanation approaches, AI model types, human-centred interface design, and the potential limitations of explanations. This framework provides a useful perspective for evaluating Chapter 11 because it demonstrates that explainability should not be reduced to making algorithmic outputs understandable. Rather, explanations must also be meaningful to educational stakeholders and appropriate for the decisions being made. Consequently, the chapter's discussion of interpretable models and counterfactual explanations represents an important contribution, although a stronger connection to stakeholder needs and the educational consequences of AI-generated explanations would further strengthen its practical relevance.

4. CRITICAL EVALUATION

The volume's strongest contribution is its translation of computational techniques into recognisable educational research tasks. Rather than presenting model architectures in isolation, it repeatedly discusses rubrics, knowledge components, learner profiles, item exposure, and instructional feedback. Chapter 5 is particularly convincing because it specifies the measurement construct, item-bank design, comparative conditions, and multiple evaluation criteria. Chapter 7 likewise makes its predictive comparisons interpretable in light of data requirements and model complexity. These choices help researchers judge whether a technique is appropriate for their question, not merely whether it is technically available.

The book also deserves credit for including evidence that qualifies its own technological promise. Chapter 3 reports weaker psychometric characteristics for generated medical assessment items, including less effective distractors, despite demonstrating efficient item production. Chapter 6 acknowledges that a computer-vision model's apparently strong training performance can deteriorate under more demanding cross-validation. Such examples strengthen the reviewer's confidence in the volume's methodological intentions because they distinguish technical feasibility from defensible assessment quality.

Nevertheless, the status of evidence shifts across chapters without an equally explicit synthesis of those differences. Chapter 4 evaluates simulated learning environments, Chapter 5 evaluates simulated assessment designs, Chapter 7 uses observed educational interaction datasets, and Chapter 9 offers illustrative recommendation outputs. These are legitimate approaches for a methodological handbook, but they establish different kinds of claims. A simulation can test algorithmic behaviour under stated assumptions; it cannot independently establish classroom effectiveness, equity, or durable conceptual learning. Readers would benefit from an editorial framework that labels each case according to its validation stage and specifies the additional evidence required before educational deployment.

One reporting detail further illustrates the need for care. Chapter 4 states that the first of four average rewards under varying measurement-error conditions was 0.81 while describing all conditions as outperforming a random baseline of 1.58 (pp. 74–75). The accompanying Figure 4.6 appears to show a value above five for that condition. This discrepancy is likely a reporting error, but it merits correction because the chapter uses those numbers to support a robustness claim. It is a localized issue rather than grounds for discounting the simulation as a whole.

A second limitation concerns conceptual integration. Chapters 4, 7, 8, and 9 discuss interconnected learner models, adaptive engines, tutoring systems, and recommendations, but the book offers limited cross-chapter guidance on how their assumptions and evaluation criteria should align in an integrated research design. The intentional modular organization explains this choice, yet a comparative synthesis would help readers distinguish prediction accuracy, instructional appropriateness, and demonstrated learning benefit. Similarly, Chapter 11's account of transparent prompts could more clearly separate the readability of an instruction from an explanation of the model's internal decision process. A comprehensible prompt does not itself validate the causal interpretation of a generated answer.

Finally, ethical reflection is recurrent but unevenly operationalized. Fairness metrics in Chapter 2 and data-protection cautions in Chapters 6, 7, and 9 are useful starting points. What remains less developed is a consistent procedure for deciding when demographic performance differences, privacy risks, or insufficient external validation should prevent deployment. This is a reasonable avenue for further development, especially because the book explicitly seeks responsible and inclusive applications. Its open-access availability and accompanying implementation resources increase accessibility, although the adaptability of individual examples will depend on technical support, local data quality, and institutional safeguards.

This limitation is particularly important in light of the ethical challenges identified by Holmes et al. (2022), who emphasize that responsible AI in education requires explicit consideration of fairness, accountability, transparency, bias, autonomy, learner agency, and inclusion. Their proposed community-wide perspective suggests that ethical responsibility should extend beyond identifying potential risks to establishing practical mechanisms for addressing them throughout the design, development, and deployment of educational AI systems. From this standpoint, the volume would benefit from more systematic guidance on how researchers can translate ethical principles into operational procedures, particularly regarding data governance, informed consent, algorithmic bias assessment, human oversight, and continuous monitoring of AI-supported educational interventions.

5. CONCLUSION

Artificial Intelligence for STEM Education Research makes a substantive contribution as an accessible, research-oriented guide to computational methods in STEM education. Its strongest chapters show how educational constructs, assessment design, algorithm selection, and evaluation must inform one another. Its principal limitation is not the absence of additional technologies but the uneven distinction between illustrative feasibility, predictive validation, and demonstrated educational impact, compounded by limited synthesis across related applications.

The volume will be most useful to graduate students, STEM education researchers, and interdisciplinary teams designing exploratory studies or extending established research workflows. Readers should treat its code and examples as starting points for context-sensitive inquiry rather than ready-made evidence of instructional effectiveness. With that qualification, the collection provides a credible foundation for technically informed and critically accountable AI research in STEM education.

ACKNOWLEDGEMENTS

The author would like to express sincere gratitude to the Chair and faculty members of the Department of Computer and Instructional Technologies Education, Gazi University, Türkiye, for their valuable suggestions, constructive feedback, and academic insights throughout the preparation of this book review. Their guidance and thoughtful recommendations have contributed significantly to strengthening the critical analysis and improving the overall quality of this manuscript. The author greatly appreciates their time, support, and academic contributions, which have been instrumental in the completion of this work.

DECLARATION OF INTEREST

The author declare no conflicts of interest.

RESEARCH FUNDING

This research did not receive any financial support.

ETHICAL STATEMENT

Not applicable. This book review is based solely on the critical analysis of a published book and does not involve human participants, human data, or experimental interventions.

AI USE STATEMENT

The author used ChatGPT (OpenAI) to support brainstorming and the development of preliminary ideas during the preparation of this book review. Grammarly was also used to assist with proofreading, grammar checking, and improving writing clarity. The author reviewed and revised all AI-assisted suggestions and takes full responsibility for the content, critical analysis, interpretation, and final version of the manuscript.

REFERENCES

Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 4*, Article 100118. https://doi.org/10.1016/j.caeai.2022.100118

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Buckingham Shum, S., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education, 32*(3), 504–526. https://doi.org/10.1007/s40593-021-00239-1

Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence, 3*, Article 100074. https://doi.org/10.1016/j.caeai.2022.100074

Xu, W., & Ouyang, F. (2022). The application of AI technologies in STEM education: A systematic review from 2011 to 2021. *International Journal of STEM Education, 9,* Article 59. https://doi.org/10.1186/s40594-022-00377-5

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education, 16,* Article 39. https://doi.org/10.1186/s41239-019-0171-0

Zhai, X., & Lee, G. (Eds.). (2026). *Artificial intelligence for STEM education research: Advanced methods and applications (Advances in Technology-Rich Science Education, Vol. 1)*. Springer. https://doi.org/10.1007/978-3-032-06565-0