AI Principles in Practice with a Learning Engineering Framework

Rachel Van Campenhout[a], Nick Brown[b] and Benny Johnson[c]
VitalSource, 227 Fayetteville Street, Raleigh, NC, U.S.A.

Keywords: Artificial Intelligence, AI Principles, Learning Engineering, Learning Engineering Process, Formative Practice, Automatic Question Generation.

Abstract: With the explosion of generative AI, rapid innovation in educational technology can lead to extraordinary advances for teaching and learning—as well AI tools that are ineffective or even harmful to learning. AI should be used responsibly, yet defining responsible AI principles in an educational technology context and how to put those principles into practice is an evolving challenge. Broad AI principles such as transparency, accountability, and human oversight should be paired with education-specific values. In this paper, we discuss the development of AI principles and how to put those principles in practice using learning engineering as a framework, providing examples of the application of responsible AI principles in the context of developing AI-generated questions and feedback. Frameworks to support the rapid development of innovative technology—and the responsible use of AI—are necessary to ground learning tools' efficacy and ensure their benefit for learners.

1 INTRODUCTION

The advent of a new, powerful age of AI tools and systems is not a time to abandon our core principles and commitment to students and learning. AI principles are needed to guide innovation in a safe, responsible, and ethical manner. As AI is evolving, so too should the principles that guide its application. More discussion of how to develop AI principles and how to apply them is needed, as AI's impact on educational technology has been—and will continue to be—significant.

In this conceptual paper we seek to engage in this discussion by examining our approach to developing and applying responsible AI practices. We discuss current examples of AI principles from governments, standards organizations, and corporate leaders and outline the process for creating our own set of AI principles. We then turn to the application of them by showcasing how the learning engineering process works as a framework for educational technology development to provide a structure for incorporating those principles. Examples from our own development of AI tools are provided. Our goal for this paper is to showcase how to shift from viewing AI principles as an abstract concept to an actionable guide, supported by learning engineering, for teams developing new learning tools using AI.

2 AI PRINCIPLES

2.1 Current Guidance

Significant work on AI principles has been done simultaneously by governments, standards organizations, and corporations alike. The European Union's proposal for AI regulation focuses on requirements for high-risk systems, but encourages all developers of low-risk AI systems to voluntarily adopt codes of conduct that align with regulations as closely as possible (European Union, 2024, Article 95). The U.S. government has provided foundational guidance to encourage responsible AI practices, particularly emphasizing the protection of individual rights and the ethical implications of AI across sectors. Central to this is the Blueprint for an AI Bill of Rights from the Office of Science and Technology Policy, (OSTP, 2022). The AI Bill of Rights identifies five principles: safe and effective systems,

[a] https://orcid.org/0000-0001-8404-6513
[b] https://orcid.org/0009-0006-4083-7579
[c] https://orcid.org/0000-0003-4267-9608

Van Campenhout, R., Brown, N. and Johnson, B.
AI Principles in Practice with a Learning Engineering Framework.
DOI: 10.5220/0013358600003932
In *Proceedings of the 17th International Conference on Computer Supported Education (CSEDU 2025) - Volume 2*, pages 312-318
ISBN: 978-989-758-746-7; ISSN: 2184-5026

algorithmic discrimination protections, data privacy, notice and explanation, and human alternatives, consideration, and fallback. While these principles are not specifically targeting education, they lay out fundamental protections to adhere to.

The U.S. Department of Education's report, Designing for Education with Artificial Intelligence (DoE, 2024), outlines five core recommendations for developers. The first recommendation, *Designing for Teaching and Learning*, urges developers to embed educational values in AI tools, focusing on "key ethical concepts such as transparency, justice and fairness, non-discrimination, non-maleficence/beneficence, privacy, pedagogical appropriateness, students' and teachers' rights, and well-being" to foster ethical, learner-centered environments (p. 12). The second recommendation, *Providing Evidence for Rationale and Impact*, calls on developers to establish clear, research-based rationales for AI designs or, if using new approaches, to transparently explain their underlying logic. Developers are encouraged to analyze data to make improvements and address risks, ensuring AI tools support diverse student outcomes and are rigorously evaluated. The report's third recommendation, *Advancing Equity and Protecting Civil Rights*, reminds developers to safeguard against bias and promote equitable access, while the fourth, *Ensuring Safety and Security*, calls for robust protections of student privacy and data security. Lastly, the fifth recommendation, *Promoting Transparency and Earning Trust*, emphasizes the importance of trust-building through open communication and clear information-sharing with educators.

In addition to governmental guidelines, standards organizations and corporate leaders have outlined specific principles to support responsible AI practices. The AI Risk Management Framework by the National Institute of Standards and Technology (NIST) offers a structured approach to addressing AI risks. This framework articulates characteristics of trustworthy AI: valid and reliable, safe, secure and resilient, accountable and transparent, explainable and interpretable, privacy-enhanced, and fair with harmful bias managed (NIST, 2023). NIST's framework highlights the importance of transparency and reliability, stating that responsible AI "involves not only minimizing risk but maximizing benefit and accountability." The NIST framework provides detailed definitions and descriptions of each component that are helpful for guiding other organizations in their AI principles. Furthermore, some corporate leaders are aware there is more to do than simply define the principles—developers also need to put them into practice. Microsoft's Responsible AI Standard v2 (2022) operationalizes their principles into concrete and actionable guidance for their development teams. While not an education-specific document, it showcases the need to deeply consider how to apply AI principles during development.

This section does not provide an exhaustive review of the work being done in the area of AI principles and frameworks, but rather provide key examples across sectors that can provide examples and guidance. These works, among others, were consulted as we developed AI principles for our context.

2.2 Developing Our AI Principles

At VitalSource, our approach to responsible AI is rooted in a commitment to creating impactful, scalable educational tools grounded in rigorous learning science. The advent of powerful, open generative AI tools has significantly shifted the educational landscape, and we view this as a means for amplifying the reach of proven learning methods. We recognize the profound responsibility involved in using AI thoughtfully and with rigorous evaluation to improve educational experiences for learners worldwide. In developing our AI Principles, we started with the values that have long guided our work and aligned them with our core mission. From our existing development and research (including existing AI systems), we identified common themes such as transparency, accountability, and rigorous evaluation. We began the synthesis of our AI principles from our internal values because we agree with the sentiment that, "In the end, AI reflects the principles of the people who build it, the people who use it, and the data upon which it is built," from the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (White House, 2024). The AI principles developed would be both a reflection of our own values and a guide for future change by considering AI guidance from leading governmental and standards organizations. By distilling these resources into our educational technology context, we developed six principles (data privacy and corporate governance omitted for brevity):

1. Accountability: VitalSource is accountable for its use of AI, from decisions on how to apply AI to ensuring quality, validity, and reliability of the output. VitalSource maintains oversight of the output through human review, automated monitoring

systems, and analysing the performance of the AI tools in peer-reviewed publications.
2. Transparency and Explainability: The AI used to power learning tools in the VitalSource platforms will be identified and documented for all stakeholders. The AI rationale, approaches, and outputs used will be explainable for stakeholders and the AI methods for learning features will be described in efficacy research evaluating those features.
3. Efficacy: Leveraging AI in our learning platforms will be applied in ways that support student learning, with a strong basis in the learning sciences and rigorous research analyses on the efficacy of the AI tools used by learners.
4. Responsible and Ethical Use: VitalSource applies an ethical approach to the application of AI, considering fairness to users, avoiding bias, and applying a learner-centered approach to the design of AI tools. VitalSource will be responsible for ensuring our use of AI complies with legal requirements and regulations, as well as aligning to standards put forth by leading standards organizations.

In essence, our approach to AI is an authentic reflection of our core values. Through adherence to these principles, we aim to advance educational technology responsibly and ethically, ensuring that every application of AI supports meaningful, research-driven learning experiences. We believe these principles align with the recommendations from governing bodies and standards organizations.

3 LEARNING ENGINEERING AS A FRAMEWORK

The application of AI principles in real-world development processes is a challenge that all organizations developing AI tools must face. Learning engineering provides a framework for practicing responsible AI. Learning engineering is a systematic, interdisciplinary process that applies engineering principles to the design and evaluation of educational technologies. Learning engineering is “a process and practice that applies the learning sciences using human-centered engineering design methodologies and data-informed decision making to support learners and their development” (ICICLE, 2023). Learning engineering was inspired by Herbert Simon, a Nobel laureate and professor at Carnegie Mellon University (Simon, 1967), and the Open Learning Initiative carried Simon’s work forward, pioneering data-driven, iterative development processes for digital learning (an application of learning engineering that guided the learning science team responsible for this work). IEEE ICICLE was formed to formalize learning engineering as a discipline and provide a professional community of learning engineering practitioners.

While learning engineering as a practice applies engineering and human-design methods with data-driven decision making to support learners (Goodell, 2022), the learning engineering process (LEP) is a model that provides structure for solving educational challenges (Kessler et al., 2022). The LEP is a cyclical process that focuses on a central educational challenge and iterates through creation, implementation, and investigation phases (Kessler et al., 2022). As part of this structured process, diverse teams collaborate to apply data-informed methods and theoretical principles to solve unique educational challenges (Goodell, 2022; Van Campenhout et al., 2023). As seen in the LEP model in Figure 1, the context, learners, and team all influence the LEP, and sub-cycles may be occurring concurrently (Kessler et al., 2022). This cyclic, iterative process ensures that educational technology is continually evaluated and improved.

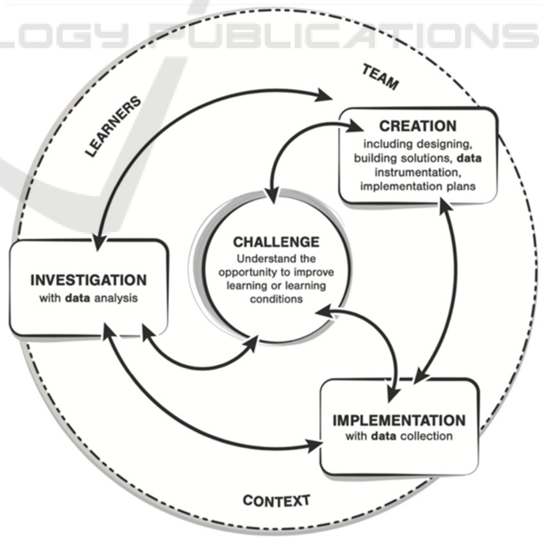

Figure 1: The LEP model (CC by Aaron Kessler).

The learning sciences provide essential foundations for learning engineering, shaping both the ideation and evaluation phases of the LEP cycle. Learning science theories inform the central challenge, design, implementation, and investigation of learning tools. In learning engineering, these theories guide each phase by providing a well-founded basis for hypotheses and informing the development of tools in a learner-centered approach. "[Learning engineering] leverages advances from different fields including learning sciences, design research, curriculum research, game design, data sciences, and computer science. It thus provides a social-technical infrastructure to support iterative learning engineering and practice-relevant theory for scaling learning sciences through design research, deep content analytics, and iterative product improvements" (Goodell and Thai, 2019, p. 563). It is clear how this deep integration of the learning sciences in the learning engineering process supports the DoE's second recommendation that developers establish clear, research-based rationales for AI designs. The investigation phase of the LEP similarly mirrors the recommendation that developers analyze data to make improvements and ensure AI tools are rigorously evaluated. By embedding learning science research into every phase of the LEP, learning engineering not only enhances the effectiveness of educational tools but also ensures their ethical alignment with key AI principles such as transparency, accountability, and efficacy.

4 AI PRINCIPLES AND LEARNING ENGINEERING: EXAMPLES FROM THE FIELD

4.1 Automatic Question Generation

Learning engineering as a practice and process has guided the development of the educational platforms and features developed by this learning science team over the past decade (Van Campenhout et al., 2023). Learning by doing—integrating formative practice with text content—was a foundation of courseware design, as doing practice was shown to be six times more effective for learning than reading alone (Koedinger et al., 2015) and shown to be causal in nature (Koedinger et al., 2016; 2018). The doer effect research guided the design of formative practice in courseware, which was then used to replicate the doer effect research with a different student population at a different university (Van Campenhout et al., 2021; 2022; 2023). Replicated and with generalizability established in natural learning contexts, this learning science research provided the basis for the decision to scale formative practice and increase the access of this learning method to more students. This became central challenge of an LEP, as seen in Figure 2. The team consisted of learning scientists, designers, engineers, and product managers and the learning environment was an ereader platform used globally by higher education institutions and learners. The solution to this central challenge was to use artificial intelligence to develop an automatic question generation (AQG) system. The creation phase consisted of many sub-cycles of ideation, development, and validation—all of which was shaped by the learning sciences, including linguistics, programming, psychometrics, educational psychology, and more. The AI-generated questions were first released in courseware used in college courses and data was collected by the platform as students interacted with the questions. The investigation phase analyzed this data and asked questions such as, "how did the AI questions perform compared to human-authored questions?" and "how did students perceive the questions?" These research topics were shared back to the educational community to contribute to the research base on AQG (Van Campenhout et al., 2021; 2022). A true "full circle" moment was reached when the doer effect, the learning science motivation for the AQG system in the first place, was found in university courses using the AI practice.

In this LEP, the role of the learning sciences is clearly anchored throughout the process; the learning sciences shape the central challenge and motivation for the project, from the creation to the investigation phases. The LEP normalizes this integration with research for a diverse team who each have different responsibilities for the project. Adhering to a process grounded in learning science eliminates the risk of building a feature simply for the sake of using a new technology. This becomes especially important when considering AI. AI should be treated as a tool that can be used in service of developing learning science-based features and environments. This same sentiment was reflected in the DoE's (2024) report: "Notably, the 2024 NETP is not directly about AI. That is because a valid educational purpose and important unmet need should be the starting point for development, not excitement about what a particular technology can do," (p. 12).

Our existing beliefs on the use of AI during the development of our automatic question generation system made clear several AI principles deeply held in our team, shown at key stages in the LEP in Figure 2. Accountability was expressly involved when we determined the type of AI we chose to use for the AQG system and how we maintained oversight. Transparency and explainability were cornerstones to

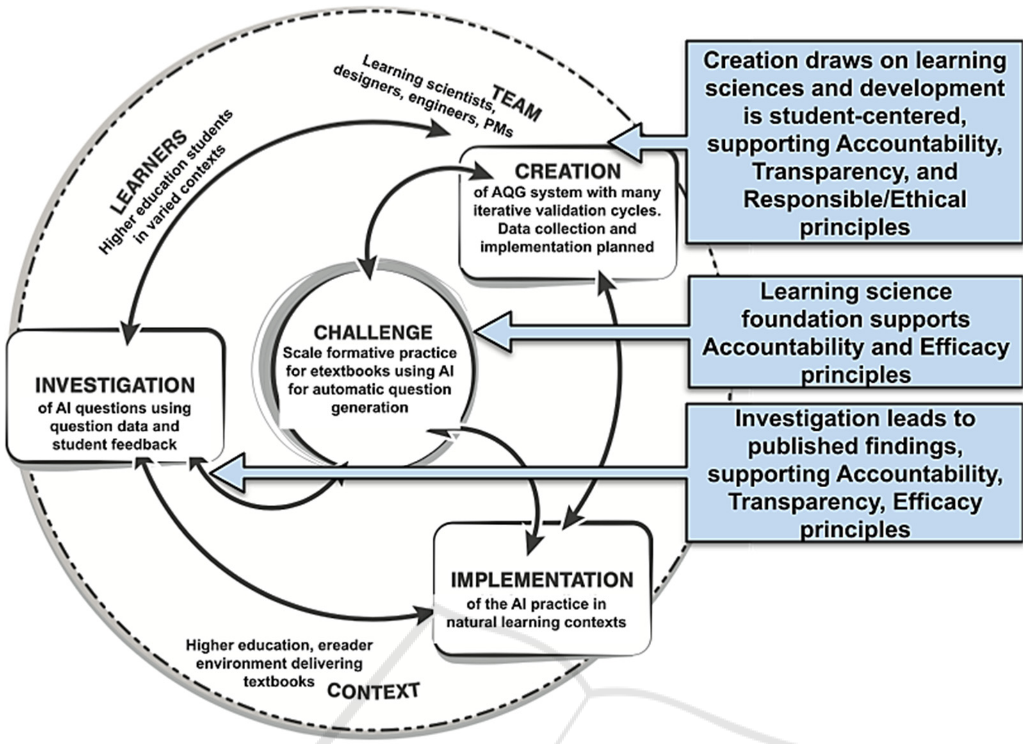

Figure 2: The LEP model for automatic question generation with notations for AI principle considerations.

our development process (Van Campenhout et al., 2023). We believed we should be able to explain exactly how our AI worked and committed to outlining the process in published research. Lastly, rigorous evaluation of the performance of the questions outputted from our AQG system was critical, as ensuring the efficacy of these formative practice questions for students was of the utmost importance. AI principles are not abstract concepts, but rather principles-in-practice that are applied throughout the LEP.

4.2 Generative AI for Personalized Feedback

Another example of how learning engineering supports responsible AI practices is a current LEP that is utilizing large language models (LLMs). LLMs exploded in quality and accessibility, becoming the driving force of a new AI era that impacted students, faculty, and companies seemingly overnight. AI tools were suddenly appearing everywhere, but were they effective for learning? The learning engineering process and its focus on the learning sciences can help guide the use of LLMs appropriately for learners. Within the central challenge we ask, what is the problem we are trying to solve, and then can evaluate if AI is the right solution. When determining how to scale formative practice in the previous AQG system example, we evaluated the potential of incorporating LLM technology. While current LLMs are far more sophisticated than those available during the development of the AQG system, we have maintained our decision not to rely solely on LLMs for generating questions at scale. This decision is grounded in ethical considerations. LLMs are capable of generating sets of questions for textbook content, and with thorough human review, these questions could potentially be used. However, ensuring the accuracy of every question generated by an LLM is impractical, particularly when scaling to millions of questions where individual human review is unfeasible. Allowing misinformation to enter formative practice at such a scale would be irresponsible and would violate key AI principles, including accountability, efficacy, and the responsible and ethical use of AI. Instead, we can explore ways to incorporate LLMs in more controlled and targeted aspects of the question generation process, ensuring their use aligns with our principles and maintains the integrity of our educational tools.

However, LLMs could be used to solve other educational challenges. One challenge that had not yet been solved by the team was how to provide feedback to students' open ended question responses at scale. Cognitive science gives a theoretical foundation for this feedback component of the learning process through working memory and cognitive load research (Sweller et al., 2011; Sweller, 2020), and VanLehn (2011) argues for the need to support students in persisting after incorrect responses as an important aspect of the learning process, of which feedback is a critical component. While there had been advances in natural language processing methods for evaluating text responses, our team had not identified a solution that could be scaled and provide feedback to our satisfaction. LLMs could help solve this long-standing challenge.

A new LEP began with a clear central challenge, a research foundation setting requirements for effective feedback, and a new idea for a solution to be tested. In this case, the application of LLM technology could be applied in a way that upheld our AI principles. Given significant content constraints (i.e. only accessing the textbook) and careful prompting, the LLM can be given the student response, the relevant textbook content, and be directed to give constructive personalized feedback. This LEP is currently in the creation phase (as shown in Figure 3), focused on aligning AI use with our principles and applying learning science research effectively. This creation phase is also focused on planning for the implementation and investigation phases of the LEP—how to evaluate that the LLM feedback is valid and effective for students. The creation phase is lengthy and involves significant work, but by grounding the development of new AI tools in the LEP and the learning sciences, the AI principles of accountability, transparency, efficacy, and responsible and ethical use are easily achieved.

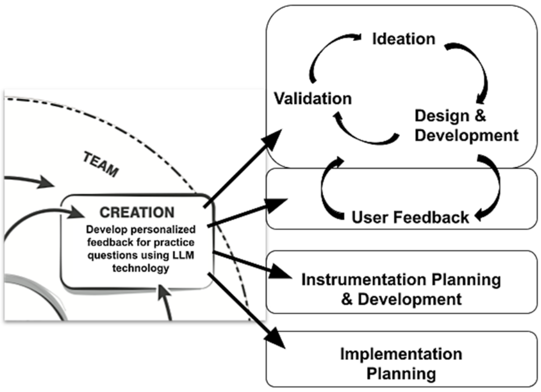

Figure 3: The creation phase of the LEP for LLM feedback development, showcasing the various tasks including cyclical development processes.

5 CONCLUSION

All organizations developing tools and environments using AI should have AI principles clearly defined and made publicly available. These principles should align with standards put forth by governing agencies and organizations, be appropriate to an educational context, and align with the core values of the people building it.

In addition to defining AI principles, organizations need a way to apply them in practice. We advocate for engaging in a learning engineering process that can provide a framework for applying both AI principles and the learning sciences for educational technology development. The contextualization, research foundation, and continuous feedback and evaluation embedded in learning engineering offer an accountable framework that is essential for applying AI in educational contexts. This approach allows for the creation of AI-driven educational tools that are both student-centered and ethically developed, providing transparency, accountability, and effectiveness in line with responsible AI principles.

REFERENCES

Department of Education. (2024). Designing for education with artificial intelligence: An essential guide for developers. Retrieved November 18, 2024, from https://tech.ed.gov/files/2024/07/Designing-for-Education-with-Artificial-Intelligence-An-Essential-Guide-for-Developers.pdf

European Union. (2024). Artificial Intelligence Act. Official Journal of the European Union. Retrieved November 18, 2024, from https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689

Goodell, J. (2022). What is learning engineering? In J. Goodell & J. Kolodner (Eds.), Learning engineering toolkit: Evidence-based practices from the learning sciences, instructional design, and beyond. New York: Routledge.

Goodell, J., & Thai, K.-P. (2020). A learning engineering model for learner-centered adaptive systems. In C. Stephanidis et al. (Eds.), HCII 2020. LNCS (Vol. 12425, pp. 557–573). Cham: Springer. https://doi.org/10.1007/978-3-030-60128-7

ICICLE. (2020). What is learning engineering? Retrieved from https://sagroups.ieee.org/icicle/

Kessler, A., Craig, S., Goodell, J., Kurzweil, D., & Greenwald, S. (2022). Learning engineering is a process. In J. Goodell & J. Kolodner (Eds.), Learning engineering toolkit: Evidence-based practices from the learning sciences, instructional design, and beyond. New York: Routledge.

Koedinger, K., Kim, J., Jia, J., McLaughlin, E., & Bier, N. (2015). Learning is not a spectator sport: Doing is better than watching for learning from a MOOC. In Proceedings of the Second ACM Conference on Learning@Scale (pp. 111–120). https://doi.org/10.1145/2724660.2724681

Koedinger, K. R., McLaughlin, E. A., Jia, J. Z., & Bier, N. L. (2016). Is the doer effect a causal relationship? How can we tell and why it's important. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, 388–397. https://doi.org/10.1145/2883851.2883957

Koedinger, K. R., Scheines, R., & Schaldenbrand, P. (2018). Is the doer effect robust across multiple data sets? In Proceedings of the 11th International Conference on Educational Data Mining, 369–375.

Office of Science and Technology Policy. (2024). Blueprint for an AI Bill of Rights. Retrieved November 18, 2024, from https://www.whitehouse.gov/ostp/ai-bill-of-rights/

Microsoft. (2022). Microsoft Responsible AI Standards, v2. Retrieved November 18, 2024, from https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2022/06/Microsoft-Responsible-AI-Standard-v2-General-Requirements-3.pdf

National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0). https://doi.org/10.6028/NIST.AI.100-1

Simon, H. A. (1967). The job of a college president. Educational Record, 48, 68–78.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory.* New York, NY: Springer-Verlag.

Sweller, J. (2020). Cognitive load theory and educational technology. Educational Technology Research and Development, 68(1), 1–16. https://doi.org/10.1007/s11423-019-09701-3

Van Campenhout, R., Johnson, B. G., & Olsen, J. A. (2021). The doer effect: Replicating findings that doing causes learning. Presented at *eLmL 2021: The Thirteenth International Conference on Mobile, Hybrid, and Online Learning,* 1–6. https://www.thinkmind.org/index.php?view=article&articleid=elml_2021_1_10_58001

Van Campenhout, R., Jerome, B., & Johnson, B. G. (2023). Engaging in student-centered educational data science through learning engineering. In A. Peña-Ayala (Ed.), *Educational data science: Essentials, approaches, and tendencies,* 1–40. Singapore: Springer. https://doi.org/10.1007/978-981-99-0026-8_1

Van Campenhout, R., Johnson, B. G., & Olsen, J. A. (2022). The doer effect: Replication and comparison of correlational and causal analyses of learning. *International Journal on Advances in Systems and Measurements, 15*(1–2), 48–59. https://www.iariajournals.org/systems_and_measurements/sysmea_v15_n12_2022_paged.pdf

Van Campenhout, R., Jerome, B., Dittel, J. S., & Johnson, B. G. (2023). The doer effect at scale: Investigating correlation and causation across seven courses. In *13th International Learning Analytics and Knowledge Conference (LAK 2023),* 357–365. https://doi.org/10.1145/3576050.3576103

Van Campenhout, R., Dittel, J. S., Jerome, B., & Johnson, B. G. (2021). Transforming textbooks into learning by doing environments: An evaluation of textbook-based automatic question generation. *Third Workshop on Intelligent Textbooks at the 22nd International Conference on Artificial Intelligence in Education CEUR Workshop Proceedings*, 1–12. https://ceur-ws.org/Vol-2895/paper06.pdf

Van Campenhout, R., Hubertz, M., & Johnson, B. G. (2022). Evaluating AI-generated questions: A mixed-methods analysis using question data and student perceptions. In M. M. Rodrigo, N. Matsuda, A. I. Cristea, V. Dimitrova (Eds.), *Artificial Intelligence in Education. AIED 2022. Lecture Notes in Computer Science, vol 13355,* 344–353. Springer, Cham. https://doi.org/10.1007/978-3-031-11644-5_28

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educational Psychologist, 46(4), 197–221.

White House. (2023). Executive order on the safe, secure, and trustworthy development and use of artificial intelligence. Retrieved November 18, 2024, from https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/