Stealth Assessments in Digital Learning Environments: Current Trends, New Directions, and Ethical Considerations

Seyedahmad Rahimi, Valerie J. Shute, & Russell G. Almond

Advances in cognitive psychology deepen our understanding of how students gain and use knowledge. Advances in technology make it possible to capture more complex performances in assessment settings, by including, for example, simulation, interactivity, collaboration, and constructed response. The challenge is in knowing just how to put this new knowledge to work (Mislevy et al., 2003b, p. 149).

Introduction

In most classrooms, assessment is the break between moments of learning. In stealth assessment (Shute, 2011), assessment is the lens that reveals learning as it unfolds. Rather than interrupting instruction with separate tests, stealth assessment seamlessly embeds evidence collection and inference directly into digital learning experiences such as games, simulations, and other interactive learning environments. As learners engage, their actions generate process data (i.e., observables) that can be interpreted to make principled claims about their knowledge, skills, and other attributes (i.e., unobservables). The result is an assessment that feels authentic (i.e., ecologically valid) and unobtrusive (learners focus on learning and problem solving rather than assessment). Moreover, real-time estimates of stealth assessment can be used for diagnostic purposes and personalizing the learning experience to maximize learning (Rahimi & Shute, in press).

Stealth assessment was first articulated over a decade ago (Shute, 2011; see also Shute, 2023). This approach is grounded in the evidence-centered design framework (ECD; Almond et al., 2015; Mislevy et al., 2003a), which guides assessment designers to create valid, reliable, and fair assessments that are high quality and psychometrically sound. ECD includes four models for the design and a four-process model that is used to implement stealth assessments. The design models include: the *Competency Model,* which specifies the targeted constructs (e.g., problem solving, creativity, collaboration, physics understanding); the *Evidence Model,* which details the behaviors or in some cases utterances (e.g., assessment of collaboration) that count as indicators and how they are modeled (i.e., identified, scored, and accumulated); the *Task Model*, which designs interactions that elicit those indicators; and the *Assembly Model,* which determines how multiple tasks and observations combine into coherent inferences about proficiency. When implemented well, ECD ensures that embedded assessments are not ad hoc analytics, but theory-driven claims supported by valid evidence.

The four-process architecture (another tool from ECD; Almond, Steinberg, & Mislevy, 2002) provides an ideal model for the implementation of a stealth assessment. The *presentation* or *evidence capture process* is the learning environment (e.g., game or simulation engine) and is responsible for logging events, which can provide evidence. The *evidence identification process* is responsible for summarizing what happens as the learner engages with a single task or activity. The *evidence accumulation process* summarizes evidence information across tasks, and finally, the *activity selection process* selects or recommends the next task. This architecture allows stealth assessments to be dynamic and formative when embedded within the activities.

The four design models described above act as the *blueprint*. In contrast, the four-process architecture functions as the *factory*—it operationalizes the design by continuously capturing learner actions, identifying evidence, accumulating results, and selecting the next activity. The blueprint and factory ensure that stealth assessments are both theoretically grounded and dynamically implemented.

It is helpful to situate stealth assessment within the broader distinction between assessment *of* learning (summative) and assessment *for* learning (formative). Assessments of learning certify what students know at the end of instruction (Harlen & James, 1997), whereas assessments for learning are used during learning to guide next steps (Black & Wiliam, 1998). Stealth assessment is intentionally designed to be formative (Shute & Rahimi, 2017), providing unobtrusive, real-time feedback and adaptively adjusting task difficulty to support learners' progress as well as informing instructional decision-making. Thus, a complete stealth assessment cycle entails capturing, scoring, and interpreting competency-related evidence, followed by meaningful actions such as providing learning supports (e.g., feedback), adaptive interventions, or personalized learning opportunities.

Over the past decade, multiple studies and papers have reported on how to design, implement, and validate stealth assessments. These efforts include stealth assessments of understanding Newtonian physics (Shute et al., 2021), reading and literacy (Fang et al., 2021; McNamara et al., 2023), as well as hard-to-assess competencies such as creativity (Shute & Rahimi, 2021), problem solving (Shute et al., 2016), persistence (DiCerbo, 2014; Rahimi et al., 2021), and conscientiousness (Moore & Shute, 2016). In addition to these empirical studies, other papers explain stealth assessment's steps in detail (e.g., Rahimi & Shute, 2024; Shute, 2011; Shute et al., 2019), as well as the software architecture and technical underpinnings of stealth assessments (Rahimi, Almond, & Shute, 2023), the psychometric properties of stealth assessment (Almond et al., 2017; DiCerbo, 2019; Rahimi, Almond et al., 2023), and the optimal development of competency models for stealth assessments (Rahimi, Almond et al., 2024).

Despite maturing theory and practice, two misconceptions about this assessment technique persist. One myth is that stealth assessment is "just for games." While games can be a natural learning environment and/or testbed, the approach generalizes to many digital contexts—including learning management systems, simulations, and creative platforms. For example, Rahimi et al. (2024) present a stealth assessment of creativity in a programming and music-remixing environment, illustrating applicability well beyond game play.

A second myth is that because learners are unaware they are being assessed, then stealth assessment is sneaky and unethical. However, "stealth" refers to its non-disruptive integration in the learning process, not secrecy. Further, transparent communication to the learner (e.g., through a dashboard) about assessment information can foster trust and motivation without compromising validity. This assessment technique is called stealth because there is no question in front of the learners to respond to thus learners don't feel like they are being assessed. Also, since stealth assessment is meant to be formative and low-stakes, communicating the fact that "the system is assessing your knowledge and skills to help you achieve your goal" to the learner can enrich their experience.

After nearly two decades of progress, the evidence base is substantial. An ongoing systematic review identifies 170 studies since 2004 spanning learning sciences, computer science, psychometrics, STEM education, and health (see Rahimi et al., 2023). These studies collectively show that stealth assessment can be psychometrically rigorous, instructionally useful, and engaging for learners.

Now the field is at an inflection point. Advances in computational psychometrics, learning sciences, machine learning and Generative AI (GenAI) are opening new possibilities for evidence identification, scoring, and adaptive support; at the same time, questions about fairness, privacy, transparency, and consent are rightly moving to the forefront. *This Journal of Research on Technology in Education (JRTE)* Special Issue aims to depict where we are and project where we could go next. In our call, we invited empirical, conceptual, and theoretical contributions across three focal areas: (1) Current research and development: validation studies, psychometric investigations, and the design/evaluation of formative supports within stealth assessment frameworks; (2) New directions: technical advances (e.g., AI/ML-enabled architectures), frameworks that extend or complement traditional ECD, and innovations in feedback, adaptivity, and assessment control; and (3) Ethical considerations: fairness and bias, data ownership and privacy, learner awareness and consent, and the responsible use of such assessments in authentic learning contexts.

The Process of Selecting Papers for the Special Issue

During the proposal stage, the three guest editors—Seyedahmad Rahimi, Valerie Shute, and Russell Almond—independently reviewed each proposal (~500 words or less) using a shared rubric. Proposals were evaluated on four criteria: Scope & Relevance (fit to stealth assessment and to *JRTE*'s readership); Soundness of Method *(if applicable)*; Clarity (of goals, design, and contributions); Novelty (theoretical, methodological, and/or practical significance). Each editor assigned an overall score (0–100) per proposal. After independent ratings, the editors met to reconcile differences and discuss borderline cases; the final proposal score was the average of the three independent ratings. Across all submissions, the median score was 70, which we adopted as the cutoff for inviting full manuscripts.

Invitations to submit full manuscripts were sent on December 20, 2024. Full papers (n = 15) were due April 1, 2025, and then proceeded through the journal's external peer-review process, coordinated by the first guest editor in accordance with *JRTE* policies. The three guest editors also served as one of the reviewers for the invited papers.

Included Papers

The resulting Special Issue includes 11 papers (see Table 1).

Table 1
Included Studies per Category in this Special Issue

Category	Author(s)	Context
Current Research and Development	Almond, Rowe, & Almeda (2026)	K–12 STEM education (computational thinking).
	Liu & Fulwider (2026)	K–12 computer science and coding education.

	Gupta, Min, Carpenter, Azevedo, & Lester (2026)	K–12 computer science and coding education.
	Beigman Klebanov & Hoang (2026)	K–12 literacy education.
	Wang, O'Reilly, Beigman Klebanov, & Suhan (2026)	Elementary education (early literacy and reading fluency).
New Directions	Adair & Gobert (2026)	Middle and high school science education.
	Hadyaoui & Cheniti-Belcadhi (2026)	K–12 and higher education (collaborative and social-emotional learning).
	Cao, Etemadi, Dede, & Wheeler (2026)	Higher education and professional learning.
Ethical Considerations	Acosta, Min, Hong, Lee, Mott, Hmelo-Silver, & Lester (2026)	K–12 STEM and collaborative learning.
	Oliveri & Poe (2026)	Higher education (engineering education).
	Dever, Wiedbusch, & Azevedo (2026)	K–12 science game-based learning.

These papers reflect a range of domains, competencies, and methodological approaches. Collectively, these contributions demonstrate the current state of the art in stealth assessment and point toward promising directions for research, practice, and policy. We briefly discuss these papers.

Current Research and Development

This Special Issue includes several foundational contributions that strengthen the empirical and psychometric base of stealth assessment research. *Validating Game-based Learning Assessment of Students' Computational Thinking Practices using Bayesian Networks and Machine-learning based Detectors* (Rowe et al., 2026) validates a game-based assessment of students' computational thinking using the game Zoombinis. This study demonstrates psychometric validity for reasoning and problem-solving by triangulating in-game evidence with external measures and showing how Bayesian and machine-learning detectors can complement one another in score inference. Another study, *Modeling Hidden States in Learning: A Hidden Markov Model Approach to Stealth Assessment for Problem Solving in Computational Thinking* (Liu & Fulwider, 2026), models learning states using Hidden Markov Models to represent the dynamic, temporal progression of students' problem-solving behaviors—such as exploration, development, and debugging. This study links process data to transfer and learning outcomes. On a similar path, *Enhancing Stealth Assessment in Game-Based Learning through Goal Recognition* (Gupta et al., 2026), focuses on enhancing stealth assessment through goal recognition which shows how inferring learners' intentions from gameplay can improve evidence quality and enable more targeted, concept-level scaffolding in real time.

Two additional studies broaden the scope of stealth assessment into literacy contexts. *Towards Stealth Assessment of Reading Comprehension* introduces a stealth assessment approach to assessing reading comprehension by leveraging natural interaction signals during narrative gameplay, moving beyond traditional comprehension quizzes toward continuous, process-based indicators of understanding (Beigman Klebanov & Hoang, 2026). *Stealth Assessment of Oral Reading Fluency During Interactive Book Reading* explores stealth assessment of oral reading fluency within interactive book reading environments, using real-time analyses to provide immediate, formative feedback while maintaining alignment with established literacy benchmarks (Wang et al., 2026).

These papers exemplify rigorous ECD principles and demonstrate the diverse applicability of stealth assessment across domains and modalities. They collectively show that when grounded in solid psychometric reasoning and thoughtful task design, stealth assessments can yield valid, formative, and engaging insights into learners' developing competencies—without disrupting the natural flow of learning. These studies illustrate the current state of the art in stealth assessment. Next, we turn to papers that chart new directions for future development of stealth assessment.

New Directions

Three studies in this Special Issue push the boundaries of stealth assessment design through AI-enabled adaptivity and cultural responsiveness which illustrate how new technologies and frameworks can extend beyond traditional ECD models. *Rex to the Rescue: Evaluating the Effectiveness of Real-Time AI-Driven Stealth Assessment and Support for Mathematical Modeling in Virtual Science Investigations* (Adair & Gobert, 2026) integrates ECD-aligned evidence models with real-time, AI-based supports to promote students' scientific modeling skills. This work shows how adaptive guidance can be synchronized with embedded assessment, ensuring that formative supports emerge precisely when learners need them—without interrupting the authenticity of inquiry-based science learning.

Similarly, *AI-Driven Adaptive Stealth Assessment for Socially Regulated Learning in Collaborative Environments* (Hadyaoui & Cheniti-Belcadhi, 2026) introduces an adaptive framework capable of detecting and responding to collaborative and social-emotional learning (SEL) processes as they unfold. By monitoring interaction dynamics, the system provides group-aware, context-sensitive feedback, dynamically adjusting task complexity and scaffolding to sustain productive collaboration. This study exemplifies how stealth assessment can move beyond the individual learner to model and support team-based competencies in real time.

Finally, *Keeping the "Glass Box" Transparent: Comparing Expert and AI-generated Ratings and Feedback in Stealth Assessment for Judgement-focused Negotiation Simulations* (Cao et al., 2026), addresses the growing integration of artificial intelligence—particularly Large Language Models (LLMs)—in assessment design. It systematically compares human-crafted and AI-generated evidence models to evaluate consistency, contextual sensitivity, and validity in assessing complex, judgment-oriented skills such as negotiation. The analysis uncovers both the promise and the risks of delegating evidence modeling to generative systems, highlighting implications for explainability, learner trust, and accountability in AI-assisted assessments.

Collectively, these studies demonstrate the next generation of stealth assessment—adaptive, context-aware, and equitable—driven by AI yet grounded in principled design and human-centered values.

Ethical Considerations

Three contributions in this Special Issue focus explicitly on ethics, fairness, and transparency in the design and implementation of stealth assessment systems. The first paper, *A Fairness-Centric Approach to Stealth Assessment in Collaborative Game-Based Learning* (Acosta et al., 2026), introduces a bias-diagnosis and mitigation toolkit tailored for embedded assessment environments. It emphasizes fairness-by-design, proposing procedures and reporting practices that account for subgroup performance differences, learners' opportunity to learn, and contextual factors that may inadvertently disadvantage certain groups. This framework moves beyond post-hoc bias detection to advocate for fairness as a guiding principle throughout the entire ECD process.

The second paper, *Fostering Industry-Ready Talent: Designing Culturally and Linguistically Responsive Stealth Performance Assessments in Engineering Education* (Oliveri & Poe, 2026) expands the conceptual and ethical reach of stealth assessment by emphasizing inclusivity and real-world relevance. The paper outlines design principles for creating culturally and linguistically responsive embedded assessments in engineering education, ensuring that the competencies measured reflect authentic workplace performance and diverse learner backgrounds.

Finally, *Investigating the Impacts of Stealth Assessment on Physiological Arousal During Game-Based Learning* (Dever et al., 2026) adds a critical human-centered dimension to the discussion by examining how embedded assessment moments relate to learners' physiological arousal and emotional regulation. The findings raise design considerations around challenge calibration and feedback timing, ensuring that stealth assessments remain non-disruptive, ethically responsible, and sensitive to learners' cognitive and affective states.

These studies underscore the importance of ethical, transparent, and learner-aware assessment practices in the age of "intelligent," data-rich educational environments. They remind us that the future of stealth assessment depends not only on technical sophistication but also on equity, trust, and empathy in how assessments are designed, communicated, and experienced.

Future of Stealth Assessment

Looking ahead, the future of stealth assessment will be shaped by four major developments—emerging from the editors' years of research in this field and reflected across the papers featured in this Special Issue. First, researchers and assessment designers will increasingly blend top-down and bottom-up approaches, combining the theoretical rigor of ECD with the adaptive power of ML and AI-based discoveries. Second, GenAI will play a growing role in identifying evidence, automating real-time scoring, and generating adaptive supports that personalize learning experiences. Third, the field will shift from research prototypes to scalable systems, expanding implementation into authentic classrooms and large-scale digital environments while addressing the infrastructure and policy challenges of real-world deployment. Finally, a design-first mentality will anchor these advances, emphasizing meaningful task design, learner experience, and ethical responsibility to ensure that stealth assessment remains not only powerful but also human-centered, equitable, and trustworthy.

Blending Top-Down and Bottom-Up Approaches

As stealth assessment enters its third decade, its continued success depends on finding synergy between top-down and bottom-up design paradigms. The top-down approach, grounded in ECD, provides theoretical rigor, construct validity, and interpretive transparency—ensuring that every piece of evidence collected aligns with a well-defined competency model. In contrast, the bottom-up approach, enabled by advances in data science, learning analytics, and artificial intelligence, allows assessment systems to adaptively evolve through pattern discovery and empirical validation. This perspective treats learners' interactions not only as data to confirm existing models but as sources for discovering new competencies, evidence patterns, and behavioral insights.

When integrated thoughtfully, these approaches can yield stealth assessment systems that are both principled and adaptive—balancing psychometric soundness with scalability and responsiveness to diverse learners and contexts. The future of stealth assessment thus lies in hybrid architectures that maintain theoretical coherence while embracing the flexibility and predictive power of data-driven discovery.

Using GenAI in Stealth Assessment

Recent advances in GenAI, particularly LLMs, foundational models, and multimodal transformers, are poised to redefine how stealth assessments identify evidence, score performance, and deliver personalized feedback. GenAI can automate portions of the evidence identification process by analyzing open-ended student responses, creative artifacts, or gameplay data to infer latent competencies. It can also support real-time scoring, enabling continuous updates to learner models without interrupting engagement.

Beyond analysis, GenAI can function as a support generator, producing adaptive hints, explanations, and reflective prompts tailored to individual learners' needs and emotional states. This dual capability—interpreting learner behavior while providing immediate, context-sensitive scaffolding—positions GenAI as both an assessment and learning partner. However, realizing this potential demands rigorous validation, transparency in algorithmic reasoning, and safeguards to prevent bias or over-reliance on automated judgments. The challenge ahead is to leverage GenAI's power while maintaining the human-centered integrity that underpins effective stealth assessment.

From Research Prototypes to Scalable Systems

Despite two decades of progress, most stealth assessment implementations remain research prototypes, constrained to controlled studies or specific learning environments. The next major step is scaling these systems to real-world educational settings—classrooms, after-school programs, and online platforms—where diversity, infrastructure, and usability present complex challenges. While many modern digital learning environments—such as automated programming platforms (e.g., CodinGame, Codewars)—implicitly apply stealth assessment principles by logging and evaluating learner interactions, these systems rarely do so within a principled design framework. They often emerge from software engineering or learning analytics traditions rather than from ECD or other systematic assessment frameworks. Consequently, while such environments demonstrate the potential scalability of stealth assessment concepts, few have achieved the theoretical coherence and psychometric fidelity characteristic of ECD-based implementations.

Achieving this transition (from prototype to scalable system) requires robust software architectures capable of handling large-scale data flows, real-time analytics, and integration with existing educational ecosystems (e.g., LMSs and adaptive learning platforms). Emerging fields such as Learning Engineering are already introducing stealth assessment as one of the key instrumentation methods in large-scale learning environments (Dede et al., 2018; Goodell & Kolodner, 2023). At the same time, the advent of Micro-credentials—which emphasize competency-based, cumulative assessments—demands use of stealth assessment more than ever before (Hunt et al., 2020; Zdunek et al., 2024).

Equally important is teacher orchestration and accessibility: educators must be able to interpret stealth assessment outputs and use them to inform instruction meaningfully. Scalable deployment also raises policy and ethical questions about data ownership, privacy, and sustainability.

As stealth assessment systems evolve toward broader use, success will hinge not only on computational sophistication but also on usability, transparency, and trust at scale.

Design-First Mentality

While technology enables stealth assessment, design determines its impact. Stealth assessment integrates both qualitative and quantitative approaches to understanding learning. It begins with a qualitative investigation through the literature and consultation with experts to define the competency of interest (i.e., competency model development; see Rahimi et al., 2024), then assessment designers deeply think about possible indicators for those competencies, and then deep analysis and design of tasks that can elicit meaningful evidence of the targeted competencies happen. These stages require thoughtful consideration of what constitutes valid, observable indicators of learning and how these can be captured within engaging, authentic tasks. Once implemented, the process transitions into a quantitative phase, where scored evidence is systematically gathered and analyzed to make inferences about learners' proficiency levels. These qualitative foundations highlight the importance of a design-first mentality—one that ensures the quantitative analyses are grounded in well-defined constructs and rich, contextually meaningful evidence. In other words, quantitative sophistication and technical novelty should not be prioritized over deep qualitative design.

Effective stealth assessments must align seamlessly with instructional goals, elicit authentic evidence through purposeful interaction, and integrate feedback mechanisms that are informative rather than intrusive or judgmental. Attention to aesthetics, narrative, and emotional design are equally important which ensure that assessment enhances, rather than disrupts, the learner's experience. Learning design frameworks such as Learning Experience Design (LXD; Ahn, 2019; Schmidt, & Huang, 2020) alongside ECD can enhance the design of stealth assessment in learning environments.

Finally, ethical and inclusive design must remain central to this vision. Stealth assessment systems should be transparent, fair, and adaptable to the diverse needs and contexts of learners. Ultimately, the next generation of stealth assessment will succeed not through algorithms alone, but through its commitment to human-centered, equitable, and inspiring learning environments—where assessment, instruction, and engagement are truly indistinguishable.

Conclusion

The contributions in this Special Issue collectively illustrate how far the field of stealth assessment has come—and where it is heading next. Over the past two decades, what began as an innovative idea for embedding assessment, seamlessly, within digital games has matured into a robust and multifaceted research domain that now spans literacy, computational thinking, collaboration, and social-emotional learning among other areas and competencies. The studies presented here reaffirm stealth assessment's potential to provide authentic, real-time, and equitable evaluations of learning while maintaining the joy and flow of engagement.

The future directions that we mentioned earlier point toward a future where assessment is no longer an isolated activity but an invisible, nondisruptive, integral part of learning itself—one that empowers educators, supports learners, and enriches the understanding of how learning happens. By combining scientific rigor with technological creativity and ethical responsibility, the next generation of stealth assessment promises not only to measure learning more effectively but to reimagine and transform it. Such learning can help learners to become interested in *learning* rather than just performing well on a test–they become lifetime learners.

Special Acknowledgement

The editors of this special issue would like to recognize the contributions of Robert J. Mislevy (1950–2025) to the field of stealth assessment. His work on both evidence-centered assessment

design and game-based assessment was foundational to the development of stealth assessment; our understanding of assessment has been greatly enriched through our conversations with him.

References

Adair, A. E., & Gobert, J. (2026). Rex to the rescue: Evaluating the effectiveness of real-time AI-driven stealth assessment and support for mathematical modeling in virtual science investigations. *Journal of Research on Technology in Education* (Special Issue on Stealth Assessment, in press).

Ahn, J. (2019). Drawing inspiration for learning experience design (LX) from diverse perspectives. *The Emerging Learning Design Journal, 6*(1), 1.

Almond, R., Row, E., & Almeda, V. M. (2026). Validating game-based learning assessment of students' computational thinking practices using Bayesian networks and machine-learning-based detectors. *Journal of Research on Technology in Education* (Special Issue on Stealth Assessment, in press).

Almond, R., Steinberg, L., and Mislevy, R. (2002) Enhancing the design and delivery of assessment systems: A four-process architecture. *Journal of Technology, Learning and Assessment.* (Vol 1.) (URL: https://ejournals.bc.edu/index.php/jtla/issue/view/210).

Almond, R., Tingir, S., Lu, X., Sun, C., & Rahimi, S. (2017, August 15). An elicitation tool for conditional probability tables (CPT) for Physics Playground. Uncertainty in Artificial Intelligence.

Acosta, H., Min, W., Hong, D., Lee, S., Mott, B., Hmelo-Silver, C., & Lester, J. (2026). A fairness-centric approach to stealth assessment in collaborative game-based learning. *Journal of Research on Technology in Education* (Special Issue on Stealth Assessment, in press).

Beigman Klebanov, B., & Hoang, A. (2026). Towards stealth assessment of reading comprehension. *Journal of Research on Technology in Education* (Special Issue on Stealth Assessment, in press).

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, *5*(1), 7-74.

Cao, L., Etemadi, A., Dede, C., & Wheeler, M. (2026). Keeping the "glass box" transparent: Comparing expert-designed and AI-generated evidence models in stealth assessment for judgment-focused negotiation simulations. *Journal of Research on Technology in Education* (Special Issue on Stealth Assessment, in press).

Dede, C., Saxberg, B., & Richards, J. (2018). *Learning engineering for online education* (Vol. 95). New York, NY: Routledge.

Dever, D., Wiedbusch, M., & Azevedo, R. (2026). Investigating the impacts of stealth assessment on physiological arousal during game-based learning. *Journal of Research on Technology in Education* (Special Issue on Stealth Assessment, in press).

DiCerbo, K. E. (2014). Game-based assessment of persistence. *Journal of Educational Technology & Society*, *17*(1), 17-28.

DiCerbo, K. E. (2017). Building the evidentiary argument in game-based assessment. *Journal of Applied Testing Technology*, 7-18.

Fang, Y., Li, T., Roscoe, R. D., & McNamara, D. S. (2021). Predicting literacy skills via stealth assessment in a simple vocabulary game. In *International Conference on Human-Computer Interaction* (pp. 32-44). Cham: Springer International Publishing.

Goodell, J., & Kolodner, J. (2023). *Learning engineering toolkit: Evidence-based practices from the learning sciences, instructional design, and beyond.* Instructional Design.

Gupta, R., Min, W., Carpenter, D., Azevedo, R., & Lester, J. (2026). Enhancing stealth assessment in game-based learning through goal recognition. *Journal of Research on Technology in Education* (Special Issue on Stealth Assessment, in press).

Hadyaoui, A., & Cheniti-Belcadhi, L. (2026). AI-driven adaptive stealth assessment for socially regulated learning in collaborative environments. *Journal of Research on Technology in Education* (Special Issue on Stealth Assessment, in press).

Harlen, W., & James, M. (1997). Assessment and learning: differences and relationships between formative and summative assessment. *Assessment in education: Principles, policy & practice*, *4*(3), 365-379.

Hunt, T., Carter, R., Zhang, L., & Yang, S. (2020). Micro-credentials: The potential of personalized professional development. *Development and Learning in Organizations: An International Journal, 34*(2), 33–35. https://doi.org/10.1108/DLO-03-2019-0074

Liu, Z., & Fulwider, C. (2026). Modeling hidden states in learning: A hidden Markov model approach to stealth assessment for problem solving in computational thinking. *Journal of Research on Technology in Education* (Special Issue on Stealth Assessment, in press).

McNamara, D. S., Arner, T., Butterfuss, R., Fang, Y., Watanabe, M., Newton, N., ... & Roscoe, R. D. (2023). iSTART: Adaptive comprehension strategy training and stealth literacy assessment. *International Journal of Human–Computer Interaction*, *39*(11), 2239-2252.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003a). Focus article: On the structure of educational assessments. *Measurement: Interdisciplinary Research & Perspective*, 1(1), 3–62. https://doi.org/10.1207/S15366359MEA0101_02

Mislevy, R. J., Steinberg, L. S., Almond, R. G., Haertel, G., & Penuel, W. (2003b). Leverage points for improving educational assessment. In B. Means & G. Haertel (Eds.), *Evaluating the Effects of Technology in Education* (pp. 149–180). Earlbaum.

Moore, G. R. & Shute, V. J. (2016). Improving learning through stealth assessment of conscientiousness. In A. Marcus-Quinn, & T. Hourigan (Eds.), *Handbook for digital learning in K-12 schools,* (pp. 355-369). New York, NY: Springer.

Oliveri, M. E. & Poe, M. (2026). Fostering industry-ready talent: Designing culturally and linguistically responsive stealth performance assessments in engineering education. *Journal of Research on Technology in Education* (Special Issue on Stealth Assessment, in press).

Rahimi, S., Almond, R., & Shute, V. J. (2023). Stealth assessment's technical architecture. In M. P. McCreery, & S. K., Krach (Eds.), *Games as stealth assessments* , (pp. 61-80). Hershey, PA: IGI Global. https://doi.org/10.4018/979-8-3693-0568-3.ch003

Rahimi, S., Almond, R., Shute, V. J., Sun, C. (2023). Getting the first and second decimals right: Psychometrics of stealth assessment. In M. P. McCreery, & S. K., Krach (Eds.), *Games as* *stealth assessments*, (pp. 125-153). Hershey, PA: IGI Global. https://doi.org/10.4018/979-8-3693-0568-3.ch006

Rahimi, S., Almond, R., Ramírez-Salgado, A., Wusylko, C., Weisberg, L., Song, Y., ... & Wright, E. (2024). Competency model development: The backbone of successful stealth assessments. *Journal of Computer Assisted Learning*, *40*(6), 2772-2789.

Rahimi, S., Shute, V. J., & Zhang, Q. (2021). The effects of game and student characteristics on persistence in educational games: A hierarchical linear modeling approach. *International Journal of Technology in Education and Science, 5*(2), 141-165. https://doi.org/10.46328/ijtes.118

Rahimi, S., & Shute, V. J. (2023). Stealth assessment: A theoretically grounded and psychometrically sound method to assess, support, and investigate learning in technology-rich environments. *Educational Technology Research and Development, 72*, 2417-2441. https://doi.org/10.1007/s11423-023-10232-1

Rahimi, S., Shute, V., Khodabandelou, R., Kuba, R., Babaee, M., & Esmaeiligoujar, S. (2023). Stealth assessment: A systematic review of the literature. In *Proceedings of the 17th International Conference of the Learning Sciences-ICLS 2023, pp. 1977-1978*. International Society of the Learning Sciences.

Rahimi, S., & Shute, V. J. (in press). Personalized learning in educational games using stealth assessment. In M. L. Bernacki, & C. Walkington (Eds.), *Handbook of personalized learning (Part I)*. New York, NY: Routledge.

Rahimi, S., Smith, J. B., Truesdell, E. J. K., Vinay, A., Boyer, K. E., Magerko, B., Freeman, J., & Mcklin, T. (2024). An automated, unobtrusive, formative assessment of creativity in a computer science and music remixing learning environment. *Psychology of Aesthetics, Creativity, and the Arts.* Advance online publication. https://doi.org/10.1037/aca0000683

Schmidt, M., & Huang, R. (2022). Defining learning experience design: Voices from the field of learning design and technology. *TechTrends, 66*(2), 141–158. https://doi.org/10.1007/s11528-021-00677-1

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.), *Computer Games and Instruction* (pp. 503-524). Charlotte, NC: Information Age Publishers.

Shute, V. J., & Rahimi, S. (2017). Review of computer-based assessment for learning in elementary and secondary education. *Journal of Computer Assisted Learning*, *33*(1), 1-19.

Shute, V. J., & Rahimi, S. (2021). Stealth assessment of creativity in a physics video game. *Computers in Human Behavior*, *116*, 106647.

Shute, V., Rahimi, S., Smith, G., Ke, F., Almond, R., Dai, C. P., ... & Sun, C. (2021). Maximizing learning without sacrificing the fun: Stealth assessment, adaptivity and learning supports in educational games. *Journal of Computer Assisted Learning*, *37*(1), 127-141.

Shute, V. J. (2023). History of stealth assessment and a peek into its future. In M. P. McCreery, & S. K., Krach (Eds.), *Games as stealth assessments*, (pp. 1-23). Hershey, PA: IGI Global. https://doi.org/10.4018/979-8-3693-0568-3.ch001

Wang, Z., O'Reilly, T., Beigman Klebanov, B., & Suhan, M. (2026). Stealth assessment of oral reading fluency during interactive book reading. *Journal of Research on Technology in Education* (Special Issue on Stealth Assessment, in press).

Zdunek, K., Dobrowolska, B., Dziurka, M., Galazzi, A., Chiappinotto, S., Palese, A., & Wells, J. (2024). Challenges and opportunities of micro-credentials as a new form of certification in health science education—A discussion paper. *BMC Medical Education, 24*(1), 1169. https://doi.org/10.1186/s12909-024-04656-y