Where Do We Go from Here?

Exploring the Limits of Intelligent Tutoring Systems.

`

Erica D. Kessler, Ph.D.

U.S. Army Research Institute for the Behavioral and Social Sciences

Introduction

Several researchers have described one-on-one human tutoring as the instructional gold-standard for training and education, as it enables learners and tutors to understand how each is approaching the learning process and to tailor instruction to individual learning needs (D'Mello & Graesser, 2023; Fox, 2020; Kraft et al., 2022; Nickow et al., 2020). For the Army, a problem arises in that this instructional gold standard is labor-intensive and does not support economies of scale. Given limited time and budget for training and education, the Army has historically focused on instructional methods that support economies of scale such as traditional, classroom instruction (e.g., lectures), distance instruction (e.g., IMI), and large-group demonstrations and hands-on exercises. While effective, these scalable methods do not achieve the same level of learning outcomes associated with one-on-one tutoring because they are not tailorable to individual learning needs (D'Mello & Graesser, 2023; Fox 2020; Kraft et al., 2022; Nickow et al., 2020).

Emerging technologies such as artificial intelligence (AI), machine learning (ML), and natural language processing (NLP) suggest solutions to address the scalability problem presented by human-to-human tutoring. For instance, the U.S. Army Learning Concept for 2030-2040 (TRADOC Pamphlet 525-8-2) argues that the Army may leverage these emerging technologies to support leader development at every echelon (TRADOC, 2024). Given continued rapid technological development, applications of AI, ML, and NLP in Professional Military Education (PME) are increasingly feasible. That said, these are evolving technologies that continue to require enhanced capabilities to mirror the instructional context provided by human-to-human tutoring. Here, I argue that we should focus on three potential paths to mature current AI, ML, and NLP technologies to be applied in automated learning environments, with the intention being to mirror with greater fidelity the epistemological processes associated with human-to-human tutoring. We need to: (a) pursue sophisticated tailoring of instruction to learner needs by leveraging the theories of mind concept, (b) develop AI capacity for principled and contextualized real-world understanding, and (c) enhance learner trust in the veracity and practical value of the instructional content provided by automated systems.

Existing scalable instructional methods are challenged in that they lack the adaptability to address individual learning needs precisely. If what the Army seeks is to provide the right learning opportunity to the right person at the right time, it must overcome the scalability problem inherent to individualized instruction (i.e., one-on-one tutoring). Emerging technologies promise to overcome the scalability problem through applications of AI, ML, and NLP within automated tutoring technologies, known as Intelligent Tutoring Systems (ITSs). For instance, researchers have sought to utilize ITSs to facilitate and maintain the precarious balance of learner emotional satisfaction and actual learning of information (Graesser et al., 2014).

When the instructional modality or process does not allow for emotional or personal enjoyment within the learning experience, learners may perceive instructional content (in both traditional and automated settings) to be challenging, useless, and repetitive. ITS designers seek to “put the student in a zone of optimal concentration that targets relevant knowledge about the subject matter at a pace that delivers the right challenges to the particular student at the right time” (Graesser et al., 2014). As explained by Lin et al. (2023), current ITSs can personalize learning experiences to account for preferred learning styles and other individual preferences. Moreover, ITSs can be blended into human-to-human interactions in instructional contexts to provide human tutors with insights on interaction, engagement, and performance using data gleaned from learning experiences.

Consider, for example, the AutoTutor technology (Graesser, 2016). AutoTutor is an AI-informed ITS that provides learners with conversations and interactions with an adaptive “tutor,” which has reportedly been successful in its applications (Graesser, 2016). Current research indicate that computer-based tutoring can be comparable to human tutoring in terms of learning gains, particularly in domains such as physics, mathematics, and reading comprehension (D’Mello & Graesser, 2023). Systems such as AutoTutor aid learning by using conversations informed by NLP to adapt to learners’ behaviors, rhetorical inputs, and even individual dispositions by using sophisticated machine learning techniques (Graesser, 2016). To consider how such technologies may support the vision described in the Army’s Learning Concept, we must better understand how human learners and tutors interact with each other in an instructional context.

Current Generation of ITS Technology

It is important to consider both the surface-level and deep-level attributes of human-to-human instruction, as ITS systems tend to mirror the surface-level characteristics of information exchange. There is also a deeper level to human interaction in which understanding and meaning are being ongoingly negotiated. Here, we focus on the psychological characteristics of human-to-human instruction providing a foundation for the observable exchanges of information taking place in instructional contexts. The next generation of ITSs should aim to support the complex and iterative psychological processes that take place between human learners and tutors, ensuring that these processes supporting knowledge creation are incorporated into their design and instructional approach. To address this challenge, we need to understand and model the epistemological framework within which human-to-human interaction facilitates the co-construction of contextualized and principled knowledge, providing learners with viable means to understand their world.

Given that the current features of ITSs tend to mirror surface level aspects of one-on-one instruction by focusing on information exchange, they lack the epistemic depth to truly replicate the interactions that a learner has with a human tutor. In recent years, researchers have focused on expanding AI and specifically ITSs to leverage current learning theories and strategies to account for more abstract learner characteristics, such as identifying inputs related to cognitive and emotional states during learning to inform subsequent learner-tutor interactions (D'Mello & Graesser, 2013). This is a move in the right direction. For example, a metanalysis by Karumbaiah et al. (2022) showed that outside of the more traditional affective states learners experience during learning (e.g., understanding/confusion, interest/boredom, engagement/disengagement), ITSs like AutoTutor can detect learners' affective states (e.g., anxiety, surprise, enjoyment). Working to improve the ability and accuracy of detecting learners' affective states could support AI in leveraging this information to tailor the instructional context, enabling these systems to engage more effectively with learners based on a nascent theory of mind.

When we view one-on-one tutoring as a back-and-forth exchange of information, as it appears on the surface, we neglect the deeper co-construction of meaning that is taking place between the learner and the tutor. We better position ITSs for success when we design these systems to mirror the processes through which human-to-human interaction already creates contextualized and principled knowledge. By understanding the interactions between learners and tutors, we are also able to establish a path for technological development to enhance the capabilities of ITSs.

At their current level of maturity, ITSs rely on large-language models to simulate human-to-human interaction. NLP can be implemented in ITSs to mirror a basic level of interaction between human tutors and learners, it lacks the contextualized understanding and theory of mind necessary to diagnose learners' conceptual (i.e., non-factual) errors and to tailor instruction to anticipate and meet learners' real-time learning needs (TRADOC, 2024). Likewise, current ITS capabilities do not facilitate learners in forming a theory of mind about the non-human instructional system with which they are interacting, which may disrupt a learner in taking on the "perspective" and understanding the "mental state" of the non-human instructional system.

Adult learners are highly skilled in forming theories of others' minds. In fact, the process of forming a theory of mind serves as a tacit ground for communicating and is essential to achieving a deep understanding of principled knowledge developed through social interaction (Bandura, 2001). In a traditional human-to-human tutoring environment, a tutor develops and applies a theory of mind about a learner to diagnose and overcome a learner's individual learning challenges. Likewise, a learner forms a theory of mind about the tutor to understand the meaning the tutor intends to convey through their instructional interaction. By applying theory of mind, tutors facilitate tailored instruction within a constructivist context by providing a means to make sense of their instructional exchange with a learner as well as to adapt instructional content and techniques to meet the unique needs they perceive a learner to have (Quesque & Rossetti, 2020).

A similar process is not yet possible within the instructional exchange between an ITS and learners. When learners and ITSs are unable to form theories of mind, it challenges the emerging consensus of principled understanding that drives human-to-human co-construction of knowledge. A key step in developing ITS technology will concern working to overcome its inability to use sophisticated learner errors to tailor instruction to meet learners' needs in real-time. Additionally, ITSs require a capability to recognize and empathize with the challenges that human learners face during the instructional interaction, as a human tutor would.

Finally, trust-building is essential to establish the perceived viability of the interaction between a human learner and human tutor (TRADOC, 2024). However, the current generation of ITSs technology does not allow for the learner to form a theory of mind about the algorithm with which they are interacting. Current AI technologies that support ITSs are based on large-language models that are prone to confabulation and hallucinations. Both confabulations (i.e., AI misinterpreting or misrepresenting existing information or stimuli) and hallucinations (i.e., AI creating new output or information based on nonexistent stimuli) can degrade trust in an ITS, if the learner is aware of these errors (Smith et al., 2023). Depending on learners' current level of knowledge in a domain, they may be able to identify that the ITS is making these types of errors. Based on that understanding, learners may determine that they have no means to exercise supervisory control over the ITS, which could lead to mistrust in the system itself. If a human learner does not trust the information provided by the ITS, the value of such an instructional modality would be lost. Trust should both be established between the learner and the ITS, and maintained throughout the learning experience.

Understanding the Current Applications of ITSs in Training and Education

Understanding how AI informs automated tutoring systems and learning environments is crucial to updating ITSs. While the mathematical and theoretical components of AI are far beyond the scope of this paper, a simple definition can provide the context for our current concerns. ITSs respond to learner input (i.e., interactivity) in a systematic procedure by producing information conditional to the interactivity (i.e., adaptivity). In real-time, the ITS then provides strategies, advice, and learning aids based on learners' needs (i.e., feedback). The continuous loop of interactivity, adaptation, and feedback occurs until the ITS system recognizes that an acceptable level of learning has been achieved (D'Mello & Graesser, 2023). This loop is fed by various ML techniques that incorporate NLP. However, it is often difficult to gain insight into the processes that take place within this loop, which is a broader challenge in AI research.

A more theoretical definition of how AI functions is also relevant. Bearman and Ajjawi (2023) define AI "according to relational epistemology, where in the context of a particular interaction, a computational artefact provides a judgement about an optimal course of action and that this judgement cannot be traced" (Bearman & Ajjawi, 2023). By this definition, AI acts as a black box that researchers have spent a great of effort seeking to explain and refine. Even so, a better course of action may be to develop the pedagogy of how we use AI in a social context, situating efforts into supporting the ability of AI to account for ambiguity, partial context, incomplete information, or fragmented mental representations—all of which better mirror the dynamic interactions present in one-on-one, human-to-human tutoring. Focusing research efforts on working *with* the black box instead of attempting to parse out what happens *inside* the black box could result in far more promising applications of AI in learning environments (Bearman & Ajjawi, 2023).

Within Army training and education, there is a need to better identify any organizational barriers that obstruct learning using ITSs and conduct research to better understand the sophisticated relationship that occurs when learners engage with, and form judgements about, ITSs during learning experiences (TRADOC, 2024; p. 41 calls for research section). Future training and education in the Army is currently envisioned to use machine learning informed by aggregate performance data to support the evaluation and refinement of the instructional approaches and training materials, which will continuously inform the refinement of current curricula to achieve performance objectives. The Army also plans to employ these technologies to analyze large amounts of data from learners to plan learning trajectories and to tailor these trajectories to the goals and needs of target audiences (TRADOC, 2024; p. 26). For example, the Synthetic Cognition for Operational Team Training (SCOTT), the Cross-Platform Mission Visualization Tool, and the Benchmarked Experiential System for Training (BEST) are all examples of ITS technology the Army is implementing, or has implemented, in training and education to create and refine teamwork models, objectives, and measures (Freeman & Zachary, 2018). Additionally, Scenario Based, Free Response Agents are computer-based trainings that provide real time interactions with a virtual agent using NLP have been used to assess interpersonal leadership skills (Brou et al., 2018). To continue to build on this vision for the future of Army training, technological development should focus on refining the capabilities of these tools to handle the sophisticated needs of Army learners throughout their careers (Brunyé et al.).

The Next Generation of ITSs in Army Training and Education

As noted, there are several challenges associated with understanding how to design ITSs to better mirror the epistemological transactions that occur between a tutor and a learner. While individual solutions for each of these challenges is likely warranted, it may be more logical to begin by making current ITS technology more sophisticated to handle the complex dynamic between a learner and a tutor. Specifically, research is needed to support the ability of an ITS to detect and adapt to learning errors or deviations that reflect attributes related to individual learner's mindsets about learning and the learning environment. The remainder of this theoretical paper will explore areas of current ITS technology that may expand our ability to capture and respond to learner and tutor interactions in learning environments, moving beyond a model focusing on information exchange to account for epistemological processes associated with learning in social contexts.

There is a need for more advanced adaptation and feedback techniques from ITSs, which should be motivated by identifying and establishing mutual theories of mind during a learning experience. The Army Learning Concept for 2030-2040 explains that an inability to connect research-based theories that support higher-level thinking to the structure and design of instructional technology and tools could lead to incompetent use of advanced AI technology (i.e., ITSs; p. 23). Theories of learning and development, specifically those with a focus on team building and collaborative problem solving, could be promising in terms of supporting current AI technology in Army leader education and development. Consider, for example, exploring ITS limitations in the context of social cognitivist theory. Social cognitivism is broadly concerned with instruction that provides learners with genuine experiences that require tangible solutions and provides guidance to learners to develop their understanding of instructional content (Vygotsky, 1978). Vygotsky (1978) proposed that, within an instructional setting, the interactions between learners and tutors should leverage their different levels of knowledge and experience, resulting in a joint effort to solve problems at the edge of learners' current capabilities, with tutors providing targeted support to learners to enable them to overcome specific challenges they are experiencing, e.g., scaffolding (Yilmaz, 2011).

In practice, ITSs like AutoTutor (Graesser et al., 2004; Nye et al., 2014) use a cognitivist approach (although AutoTutor is based on multiple theories of learning) to support learners through conversation-based instruction. Across domains, when interacting with AutoTutor, a learner will engage with a "tutor" to achieve a certain level of understanding or mastery of an instructional domain. For example, AutoTutor-ARC (Adult Reading Comprehension) was developed to help adult learners increase their literacy and comprehension skills. This program introduces learners to two computer agents, one acting as a tutor and the other acting as a peer (Chen et al., 2021; Graesser et al., 2016). During each lesson, learners engage in a tutoring session with both agents by having three-way conversations (with each party representing varying degrees of expertise and insight on the topic) rooted in foundations of reading comprehension strategies.

When the learner, the tutor, and peer understand and approach the instructional content differently, it prompts individualized feedback from the ITS to motivate and guide learners to increase their comprehension skills. Additionally, the AutoTutor software integrates NLP, allowing for insights into learners' verbal and written responses (Allen et al., 2023). While insights on things like comprehension level and sentential content matches are important, applications of NLP to detecting or determining context are scant and often do not provide the same benchmark of accuracy as more traditional text analysis applications of NLP. A major limitation of AutoTutor technology is its inability to detect context if the context has not been predetermined (e.g., previously programmed into the software or algorithm). Karaumbaiah et al. (2022) argue that ITS requires improved capabilities to detect and adapt to learners' affective states, which could allow it to identify additional contexts. However, being able to detect and adapt to *the source* of these emotions is perhaps an additional area of interest. For example, it may be valuable for ITSs to detect learners' anxiety or emergent understanding during an instructional session, but the ability of these systems to detect the context of that affective state might prove equally important. Without determining the context of learners' affective states, the systems could not effectively adapt and engage with the learner in a specific way that meets their current learning needs. Consider a scenario in which a learner displays signs of anxiety during a learning session, but the system cannot determine if the anxiety is about the topic of study, the interaction with the ITS, something else entirely, or a combination of these factors. If the ITS was able to form a more detailed conceptualization of the learner's real-time learning needs related to the anxiety they are experiencing, the technology could provide more appliable strategies for the context of that specific learning experience.

Researchers have been working to better design AI and machine learning software to detect and adapt to context, namely by focusing on things like attention, engagement, and motivation (Allen et al., 2016; Bosch & D'Mello, 2019; D'Mello & Graesser, 2012; Mills & D'Mello, 2015). For example, researchers recently developed an automated disengagement tracking system (DTS) that works within AutoTutor to detect behaviors that may be detrimental to learning (Chen et al., 2023). By establishing a baseline measure of engagement via aggregate learner response data on the first few questions of the lesson, the system can detect deviations from the baseline that signal disengagement from the learning experience (Chen et al., 2023). Additionally, by using a combination of individual differences measures, text indices, and keyboard stroke analyses, Allen et al. (2016) were able to accurately detect both high and low learner engagement, as well as boredom, across more than one-hundred writing samples. Utilizing this type of detection technology allows for improvement of predictive models within ITS and assists in setting an empirical foundation for obtaining measures of affect using AI technologies.

While disengagement is often indicative of a learner's attention and motivation, there is reason to believe that disengagement detection software could be also useful in assessing and adapting to contextual understanding and the learner's understanding of the ITS itself. For example, consider a scenario in which a Soldier is interacting with an ITS for a relevant training exercise. A system that functions like DTS, could establish a baseline measure of “contextual understanding” or “mindset” or “theory of mind” and detect when learners deviate from that baseline and in what ways. Consider a scenario in which a learner starts with a baseline mindset that gaining knowledge is something they can control and pursue, however, during an interaction with an ITS deviates such that their behaviors indicate they may be endorsing that mindset less during the learning session. The system could detect and respond in a way that encourages or supports the learner’s original mindset, redirecting them based on the context of the situation. Similarly, a baseline measure regarding an individual’s trust of the ITS could be assessed against their real-time learning behaviors, allowing the system to detect mistrust or lower levels of trust during a learning session. By detecting deviations in level of trust (either by determining the participant is losing or gaining trust in the system), the system can then adapt based on the learner’s needs in that moment, allowing space for more authentic learner-tutor interactions. While theoretical, applications of this type of detection software could provide promising avenues for tailoring current ITSs to better fit the individual learner’s theory of mind and level of trust in a particular learning environment.

An additional approach to gaining insights on context and state of mind during ITS interactions, is to design technology to detect and adapt attributes and behaviors related to emotions related to learning situations, for example confusion, frustration, boredom, etc. (Allen et al., 2016; D’Mello & Graesser, 2012). As previously mentioned, a common area of concern for ITS technology is an inability to detect and understand a learner's affect during learning situations (Allen et al., 2016). However, D'Mello & Graesser (2012) found that text-based emotion detection technology that uses machine learning techniques alongside AutoTutor technology has many potential applications for analyzing and predicting learner affect during learning. By employing AutoTutor technology that uses computational linguistics and computational discourse, systems like Coh-Metrix and the Linguistic Inquiry Word Count (LIWC) provide researchers with a more thorough analysis on the ways in which, and to what extent, these metrics can be used to detect and predict emotion. However, this technology originally only allowed for judgements about the learner's emotional state every 20 seconds, making it almost impossible to trace a learner's emotional state or change in emotional state to a particular real-time interaction with the tutoring system (D'Mello & Graesser, 2012a). To account for this limitation, AutoTutor technology was expanded to apply learning techniques after each turn between the learner and tutor allowing for affect judgements to be made more frequently and with more accuracy (D'Mello & Graesser, 2012b). Adapting current AI technology to account for learning-centered emotions, related to the material *and* the environment, could allow for a better sense of what and how attributes such as emotions and affect impact learning success.

A promising opportunity for adapting AI technology to address scalability and epistemological concerns in Army training and education could be related to the idea that learners do not necessarily experience one affective or emotional state at a time during learning (Bosch & D’Mello, 2014). For example, foundational work exploring affect during ITS learning situations assumed that a learner may feel confused, *then* they feel frustrated. However, it is more plausible that the learner experiences the co-occurrence of different affects during a single learning instance (i.e., feeling confused *and* frustrated at the same time). Determining not only what affective states co-occur, but also how the patterns of co-occurrence of different affective states relate to learning gains in ITSs, can inform technology on better strategies for detecting and adapting to varying learner needs. For example, consider a situation in which a learner is feeling both confused and frustrated over a learning topic. Should the ITS respond the confusion or the frustration first? Should the system resolve feelings of confusion completely before adapting to the learner's frustrations? How does that confusion or frustration impact the learner's trust in the ITS or shared theory of mind within the automated learning environment? Consider an environment in which the ITS is able to detect the various affective states related to epistemological factors (mistrust, shared understanding, collaborative tendencies, etc.) and adapt based on the learner, perhaps even calibrating to the learner's current affective state to reestablish trust or a shared understanding about the learning space. This could elicit the characteristics of a more traditional one-on-one tutoring session, in which the tutor is able to detect, adapt to, and support an individual learner's current needs in a way that ITSs have yet to demonstrate.

While current ITS technology is impressive in its ability to provide an interactive learning environment, tailoring the next generation of ITS technologies to better simulate the epistemological interaction, communication, and mutual trust that arises during a traditional tutor and learner tutoring setting could provide promising applications for overcoming scalability constraints of human tutoring. Additionally, enhancing the ability of ITSs to detect a learner's affect and context could provide an avenue for scaling the co-constructive value of a dedicated human tutor for a learner. Applying more advanced technologies to adaptation and feedback software within ITSs could result in more robust applications to Army leader training and education.

References

Allen, L. K., Graesser, A. C., & McNamara, D. S. (2023). Automated analyses of natural language in psychological research. In H. Cooper, M. N. Coutanche, L. M. McMullen, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), APA handbook of research methods in psychology: Foundations, planning, measures, and psychometrics (2nd ed., pp. 361–380). American Psychological Association. doi: 10.1037/0000318-017.

Allen, L. K., Mills, C., Jacovina, M. E., Crossley, S., D'Mello, S., and McNamara, D. S. (2016). Investigating boredom and engagement during writing using multiple sources of information: the essay, the writer, and keystrokes. In Proceedings of the sixth international conference on learning analytics & knowledge, (New York, NY, USA: Association for Computing Machinery), LAK '16, 114–123. doi: 10.1145/2883851.2883939.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, 52(1), 1-26. doi: 10.1146/annurev.psych.52.1.1.

Bearman, M., & Ajjawi, R. (2023). Learning to work with the black box: Pedagogy for a world with artificial intelligence. *British Journal of Educational Technology*, *54*(5), 1160-1173.

Bosch, N., & D'mello, S. K. (2019). Automatic detection of mind wandering from video in the lab and in the classroom. *IEEE Transactions on Affective Computing, 12*(4), 974-988. doi:10.1109/TAFFC.2019.2908837.

Brou, R., Ledford, B., Stallings, G., Normand, S., & Stearns, I. (2018). Building automated assessments of interpersonal skills. In Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC). Orlando, FL.

Brunyé, T. T., Brou, R., Doty, T. J., Gregory, F. D., Hussey, E. K., Lieberman, H. R., ... & Yu, A. B. (2020). A review of US Army research contributing to cognitive enhancement in military contexts. Journal of Cognitive Enhancement, 4, 453-468.

Chen, S., Fang, Y., Shi, G., Sabatini, J., Greenberg, D., Frijters, J., & Graesser, A. C. (2021). Automated disengagement tracking within an intelligent tutoring system. *Frontiers in Artificial Intelligence*, 3, 595-627. doi.org/10.3389/frai.2020.595627

D'Mello, S. K., & Graesser, A. (2023). Intelligent tutoring systems: How computers achieve learning gains that rival human tutors. In *Handbook of educational psychology* (pp. 603-629). Routledge.

D'Mello, S. K., & Graesser, A. (2013). AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *2*(4), 1-39. doi.acm.org/10.1145/0000000.0000000.

D’Mello, S. K., and Graesser, A. (2012a). Language and discourse are powerful signals of student emotions during tutoring. *IEEE Transactions on Learning Technologies,* 5, 304–317. doi:10.1109/TLT.2012.10.

D’Mello, S., & Graesser, A. (2012b). Dynamics of affective states during complex learning. *Learning and Instruction,* 22(2), 145-157. doi.org/10.1016/j.learninstruc.2011.10.001

Fox, B. A. (2020). *The Human Tutorial Dialogue Project: Issues in the design of instructional systems*. CRC Press.

Freeman, J., & Zachary, W. (2018). Intelligent tutoring for team training: Lessons learned from US military research. In *Building intelligent tutoring systems for teams: What matters* (pp. 215-245). Emerald Publishing Limited. doi/10.1108/S1534-085620180000019013.

Graesser, A. C. (2016). Conversations with AutoTutor help students learn. *International Journal of Artificial Intelligence in Education*, *26*, 124-132.

Graesser, A. C., Cai, Z., Baer, W. O., Olney, A., Hu, X., Reed, M., et al. (2016). Reading comprehension lessons in autotutor for the center for the study of adult literacy. In *Adaptive educational technologies for literacy instruction*, (pp.288–293). Taylor & Francis Routledge.

Graesser, A. C., D'Mello, S. K., & Strain, A. C. (2014). Emotions in advanced learning technologies. In *International handbook of emotions in education* (pp. 473-493). Taylor & Francis Routledge.

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H. H.,Ventura, M., Olney, A., & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 180–192. https://doi.org/10.3758/BF03195563.

Karumbaiah, S., Baker, R., Tao, Y., & Liu, Z. (2022, March). How does students' affect in virtual learning relate to their outcomes? A systematic review challenging the positive-negative dichotomy. In *LAK22: 12th International Learning Analytics and Knowledge Conference* (pp. 24-33). doi.org/10.1145/3506860.3506863.

Kraft, M. A., List, J. A., Livingston, J. A., & Sadoff, S. (2022). Online tutoring by college volunteers: Experimental evidence from a pilot program. In *AEA Papers and Proceedings 112*, (pp. 614-618). doi.org/10.26300/b1ch-0g2.

Lin, C. C., Huang, A. Y., & Lu, O. H. (2023). Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review. *Smart Learning Environments*, *10*(1), 41. doi.org/10.1186/s40561-023-00260-y.

Mills, C., & D'Mello, S. (2015). Toward a Real-Time (Day) Dreamcatcher: Sensor-Free Detection of Mind Wandering during Online Reading. In *Proceedings of the 8th International Conference on Educational Data Mining* . (pp.69-76).

Nickow, A., Oreopoulos, P., & Quan, V. (2020). The impressive effects of tutoring on prek-12 learning: A systematic review and meta-analysis of the experimental evidence. NBER Working Paper No. 27476, JEL 2(J4).

Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and family: A review of 17 years of natural language tutoring. *International Journal of Artificial Intelligence in Education*, *24*(4), 427–469. https://doi.org/10.1007/s40593-014-0029-5.

Quesque, F., & Rossetti, Y. (2020). What do theory-of-mind tasks actually measure? Theory and practice. *Perspectives on Psychological Science*, 15(2), 384-396. DOI: 10.1177/030100662097698.

Smith, A. L., Greaves, F., & Panch, T. (2023). Hallucination or confabulation? neuroanatomy as metaphor in large language models. *PLOS Digital Health, 2*(11), doi.org/10.1371/journal.pdig.0000388.

U.S. Army Training and Doctrine Command. (2024). The *Army Learning Concept for 2030-2040.* TRADOC.

Vygotsky, L. S. (1978). *Mind in society: the development of higher psychological processes*. Harvard.

Yilmaz, K. (2011). The cognitive perspective on learning: Its theoretical underpinnings and implications for classroom practices. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 84*(5), 204-212. DOI: 10.1080/00098655.2011.56898.