Disembodied AI and the limits to machine understanding of students' embodied interactions

Mitchell J. Nathan*

MAGIC Lab, Wisconsin Center for Education Research, Educational Psychology Department, School of Education at the University of Wisconsin–Madison, Madison, WI, United States

The *embodiment turn* in the Learning Sciences has fueled growth of multimodal learning analytics to understand embodied interactions and make consequential educational decisions about students more rapidly, more accurately, and more personalized than ever before. Managing demands of complexity and speed is leading to growing reliance by education systems on disembodied artificial intelligence (dAI) programs, which, ironically, are inherently incapable of interpreting students' embodied interactions. This is fueling a potential *crisis of complexity*. *Augmented intelligence* systems offer promising avenues for managing this crisis by integrating the strengths of omnipresent dAI to detect complex patterns of student behavior from multimodal datastreams, with the strengths of humans to meaningfully interpret embodied interactions in service of consequential decision making to achieve a balance between complexity, interpretability, and accountability for allocating education resources to children.

KEYWORDS

artificial intelligence, augmented intelligence, cognitive science, embodied learning, foundation models, learning sciences, multimodality

1. Introduction

The primary objective of this *Perspectives* article is to expose a looming crisis of complexity: educational systems are becoming more dependent on artificial intelligence (AI) programs to make consequential decisions about learning and learners from rich streams of multimodal data that emerge from many sources, including students' embodied interactions. However, disembodied AI (dAI) programs–I argue–are fundamentally incapable of understanding people's embodied interactions in the ways that humans understand them. Furthermore, the emergent dAI models are of such complexity that end users (and often the original programmers) cannot understand the models or recreate the chain of reasoning that led to these decisions. Therefore, dAIs should not be directing consequential educational decisions affecting the lives of children. The secondary objective is to offer potential paths forward from this crisis. One promising approach is the development of "augmented intelligence" systems (AISs) that amplify human performance using dAI resources while relying ultimately on human decision making.

2. Theoretical framework: The embodied turn and growth of multimodal learning analytics

2.1. The embodied turn in the learning sciences and education

Empirical evidence and arguments from philosophy, psychology, neuroscience, education, and critical theorists in education effectively dismantle the view of learning as information processing of ungrounded symbol systems by dAI that are amodal (i.e., non-sensorial), arbitrary (i.e., non-historical and non-cultural), and abstract (i.e., ungrounded) (Harnad, 1990; Varela et al., 1991; Glenberg, 1997; Shapiro, 2019). To the contrary, humans make meaning of events, ideas, and cultural and scientific inscriptions by grounding them to their sensorimotor experiences that are interpreted within sociocultural and historical contexts (Wilson, 2002; Barsalou, 2008; Newen et al., 2018).

In psychology, Glenberg and Robertson (2000) found that human readers judge the sensibility of sentences based on the sensorimotor affordances invoked by the actions described in the sentences, rather than their lexical interconnections in high-dimensional spaces, as modeled by dAI systems widely applied in education areas such as automated essay grading (LSA; Burgess and Lund, 1997; Landauer and Dumais, 1997).

Neural imaging data show that reading words with motor associations—such as kick, lick, and pick—selectively activates the motor areas of the brain for one's feet, tongue, and fingers, respectively (Pulvermüller, 2005). Botox patients whose injections temporarily paralyze the facial corrugator supercilli muscle used in frowning showed selective impairment in processing sentences that invoke anger but not those that invoked joy or were emotionally neutral (Havas et al., 2010).

Critical theorists in education reject the disembodied view that neglects the central role of culture in language, thinking, symbols, and emotion for educational attainment. McKinney de Royston et al. (2020) expressly identify the essential nature of embodied cultural experiences by framing learning as rooted in bodies and brains that are embedded in social and cultural practices and shaped by lifelong culturally organized activities.

Drawing on these critiques, some education scholars conclude that the knowledge and educational practices of students and teachers are fundamentally determined by people's individual and collective embodied processes in order to make sense of their school-based learning experiences (e.g., Shapiro and Stolz, 2019; Nathan, 2021; Macrine and Fugate, 2022). This has led to innovative designs in embodied learning through educational technology (Papert, 2020; Abrahamson and Lindgren, 2022), embodiment in AI and education (Timms, 2016) and embodied conversational agents (Cassell, 2001) that promote student learning and intellectual development.

2.2. Growth of multimodal learning analytics

With the embodiment turn has emerged methods for collecting and analyzing *multimodal data* to model embodied interactions (Worsley and Blikstein, 2018; Abrahamson et al., 2021). These include data for analyzing gestures (Closser et al., 2021), eye gaze (Schneider and Pea, 2013; Shvarts and Abrahamson, 2019), facial expression (Monkaresi et al., 2016; Sinha, 2021), grip intensity (Laukkonen et al., 2021), and so on, coupled with traditional statistical methods, qualitative methods, and deep learning algorithms that model human behavior based on massive amounts of mouse click and text-based data (e.g., Facebook's DeepText, Google's RankBrain). This shift in research methods has been enabled by the proliferation of low-cost, high-bandwidth cameras and sensors that track biometrics, facial, and body movement that supplement field notes, speech, text chat, and click log data (Schneider and Radu, 2022).

Work with multimodal data has historically been labor-intensive and subject to the severely limited processing capacities of humans that constrain the amount of data under consideration, its dimensionality, and the cycle time between data collection, interpretation, and action. This restricted the ability to use multimodal data to identify latent patterns and inform practitioners in real time about embodied interactions relevant to on-task and off-task behavior. Some of the forces that propelled educational data mining and learning analytics (Aldowah et al., 2019; Baker and Siemens, 2022) have motivated the creation of more efficient data analytic tools and algorithms to process massive multimodal corpora (e.g., An et al., 2019; Järvelä et al., 2019). This is leading to the emergence of new methodological practices of *multimodal learning analytics and data mining* (hereafter MMLA; Blikstein and Worsley, 2016).

3. Analytic method and evidence: The disconnect between dAI and human meaning making

An analysis of the computational architectures of classical and contemporary AI systems that underly the tools for MMLA reveals that they are fundamentally incapable of understanding the meaning of people's embodied interactions, even as they give the appearance of mimicking intelligent embodied behavior.

Classical, symbol-based AI systems were designed and implemented by human programmers to emulate human intelligence. The arbitrary, amodal, and abstract nature of these symbol systems was a feature, not a bug, and key to the power of these computational algorithms to operate consistently and efficiently, across a wide range of domains. For example, semantic nets presumably could model any organization of memory (Collins and Loftus, 1975). Although classical AI systems excelled at the analytic tasks that are the signature of adult intellect, such as complicated calculations and hierarchical inference-making, they were wholly inadequate at performing culturally familiar tasks well within reach of children, such as balance, face recognition, and

basic social interactions (e.g., Resnick, 1987) and struggled to be adaptive in the face of task, environmental, and user variation.

Connectionist architectures arose that addressed many limitations of classical AI. Often, these drew on parallel and distributed forms of computation that adapted to training experiences through the adjustment of strengths of connections among simple nodes in large networks, mediated by hidden layers (McClelland et al., 1986; Rumelhart et al., 1988). These systems excelled at simple pattern learning and prediction, and at many of the sensorimotor skills that eluded early symbolic AI systems. Yet these connectionist systems found many symbol analytic tasks cumbersome. These systems depended heavily on carefully cultivated training sets and pre-coded sensory inputs for successful learning, underscoring their disembodied nature.

New approaches arose that exploited high-dimensional spaces for computing variability and similarity, greatly expanding the training sets they could accommodate and the complexity of the associations they could encode (e.g., Burgess and Lund, 1997; Landauer and Dumais, 1997). Thus, attention in AI development turned to the importance of training experiences and the sheer number of nodes and inter-nodal connections used by these systems.

This fueled the current movement to Foundation AI systems such as BERT, GPT-3, and DALL-E that are built to accommodate enormous training corpora with massive numbers of internodal connections (Bommasani et al., 2021). Foundation AI systems are designed to learn on their own and be adaptive to completely new, untrained conditions—often in ways that their creators cannot foresee. For example, GPT-3 is built on 175 billion parameters trained on 570 Gigabytes of text. GPT-3 can learn to write original essays, produce computer code, and generate reasonable responses to novel discourse (not just novel syntactic structures) it has never been trained on.

Still, these systems are working from disembodied patterns extracted from the regularities of how words and images occur in the training datastreams. GPT-3, as a representative example, "lacks intentions, goals, and the ability to understand cause and effect" [Percy Liang, Director of Stanford's Center for Research on Foundation Models (CRFM), in CRFM, 2021] that naturally come from human being's embodied interactions with one's environment and other people. Newer language models, such as ChatGPT, are based on GPT-3 architecture and develop their language generation and comprehension capabilities through these same basic analytic methods, coupled with a mechanism of Reinforcement Learning from Human Feedback (RLHF; Ouyang et al., 2022) from human labelers. Despite its fascination in the media, RLHF has significant limitations as noted by the developers (Ouyang et al., 2022). Its future performance is based on a number of subjective and untested sources of human bias; specifically: unaccounted for biases of the human labelers and the researchers who initially developed the instructions used by the labelers; the prompts provided by the developers and early users; and that the same human biases are present in the training and model evaluation process. Furthermore, foundation models like GPT-3, ChatGPT, and the like are completely opaque: the creators do not know how the models will work in new domains and cannot predict the future interactions of their creations What's more, in what is both a profound strength and a serious weakness, architectural and

training decisions made early on influence a system throughout its lifetime. Thus, when key considerations such as embodiment are neglected, one cannot simply go back and retrofit changes (Bommasani et al., 2021).

These issues of disembodiment, opaqueness, and developmental fixedness all converge to shape a distorted image of what the educational community should be drawn to. As Liang notes in a recent webinar (CRFM, 2021), ideally, "the ethical and social awareness needs to be integrated into the technological development." However, the norm for social and ethical considerations is to follow *after* the technology is built, trained, and deployed. Liang laments "At that point I think it's too late [Because of emergence and homogenization] some of the critical decisions have been made already, in a structural way" (CRFM, 2021).

Despite their enormous computing power, dAI programs for MMLA are fundamentally incapable of deriving human-centered meaning from embodied interactions. dAI programs fail along philosophical grounds to achieve intentionality (Searle, 1980). Instead, they generate ungrounded models of behavior linked to high-dimensional statistical regularities of behavior, rather than the meaningful embodied experiences they purport to model (Harnad, 1990). They fall short phenomenologically by relying on mathematical redescriptions that intervene between sensation and action (Gallagher, 2018). And the symbol structures they generate to describe human behavior have no cultural or historical bases (McKinney de Royston et al., 2020). As Barsalou (1999, p. 608) states, "computers should not be capable of implementing a human conceptual system, because they do not have the requisite sensory-motor systems for representing human concepts."

4. Urgency of the problem of dAI in educational decision making

A variety of automated detectors have been developed that use non-invasive methods to classify students' emotional states, engagement, and cognitive presence during their participation in on-line classes (e.g., Baker et al., 2010; Liu et al., 2019, 2023).

The increasing availability of multimodal data has coincided with growing expectations for computers to deliver data-driven, real-time directives for education, such as personalized learning (Walkington, 2013) and assessment, added pressures from a global pandemic that disrupted standard, in-person learning, and a lack of oversight or regulation on the access and use of such data by machines in educational settings (Crawford, 2021). The response has been a proliferation of dAI-based solutions to traditional educational problems such as formative and summative assessment and differentiated curricula using tools, such as *4 Little Trees*, that uses eye gaze, facial expression, and body movement to make educational decisions and evaluations about student attentiveness and level of engagement (Chan, 2021; Harper et al., 2022); and systems such as TalkMoves, that collect recordings of classroom discourse but ignore students' non-verbal interactions (Suresh et al., 2021).

The urgency is that school leaders and classroom teachers looking to manage their workloads with limited resources see dAI-based systems as ready-made solutions (e.g., Tyson, 2020).

However, school leaders and teachers may be ill-informed about the actual inner workings of dAI systems and the inherent limitations of these systems to understanding people's embodied interactions in the ways that humans understand them, as described in section 2. This needs to change before educational practices become too dependent on dAI systems without proper considerations of ways to address these limitations (as outlined in the next section).

The potential risks are that students' embodied ways of expressing their reasoning are disregarded, thus providing impoverished accounts of their engagement and learning; or, that these non-verbal behaviors are incorrectly classified due to the limitations and biases built into the dAI systems. In both scenarios, dAI systems would be given authority over consequential decisions about students' educational experiences that can have lifelong consequences without adequate oversight by educators.

5. Pathways forward

Given dAI limitations, alternatives are needed to manage the complexities of embodied interactions while still offering time-sensitive, human-centered interpretations and accountable decision-making. The emergence of augmented intelligence systems (AISs; Dubova et al., 2022) in areas such as healthcare with high-levels of personal interactions (Crigger et al., 2022) and need for trust ([HLEG-AI] High-Level Expert Group on Artificial Intelligence, 2019) offer promising avenues for education. One exemplar is *detector-driven interviewing* (DDI) methods. DDIs use dAIs to continually monitor human behavior using non-invasive methods for cognitive and affective patterns that signal learning and engagement events of importance to educators (e.g., frustration detectors), then alert human researchers and practitioners of these events to trigger personalized attention, natural human interactions, and customized pedagogical support (Baker et al., 2021; Ocumpaugh et al., 2021; Hutt et al., 2022). Successful DDIs in the learning system *Betty's Brain* (Leelawong and Biswas, 2008) demonstrates its ability to improve educational responsiveness that enhances student engagement and contributes to scientific models of the cognitive and affective processes that shape learning.

6. Discussion

The *embodiment turn* in the Learning Sciences dismantles accounts of intellectual behavior that equates cognition with disembodied computation. The rise of MMLA applied to student education is fueling a quiet movement to accede human educational decision making to dAI systems. This essay uses an embodiment framework to argue that autonomous dAI systems are fundamentally incapable of understanding embodied interactions the ways that humans understand embodied interactions due to their disconnect from sensorimotor and sociocultural interactions with their environments, and therefore should not be directing consequential educational decisions.

Thus, there is a looming crisis of complexity as dAI systems fundamentally incapable of understanding embodied interactions will be enlisted to manage the enormous complexities of the multimodal models used to describe those embodied interactions and make consequential educational decisions for students. Ethical and embodied AI systems seem a long way off. The time is ripe to invest in alternatives such as augmented intelligence systems that cultivate the omnipresence and computational power of dAIs with the embodied meaning making of human interpreters and decision makers (as illustrated by approaches such as detector-driven interviewing) as a means to achieve an appropriate balance between complexity, interpretability, and accountability for allocating education resources to our children.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Acknowledgments

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abrahamson, D., and Lindgren, R. (2022). "Embodiment and embodied design," in *The Cambridge Handbook of the Learning Sciences*, ed R. K. Sawyer, 3rd ed. (Cambridge: Cambridge University Press), 301–320. doi: 10.1017/9781108888295.019

Abrahamson, D., Worsley, M., Pardos, Z. A., and Ou, L. (2021). Learning analytics of embodied design: enhancing synergy. *Int. J. Child Comput. Interact.* 32:100409. doi: 10.1016/j.ijcci.2021.100409

Aldowah, H., Al-Samarraie, H., and Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: a review and synthesis. *Telemat. Inform.* 37, 13–49. doi: 10.1016/j.tele.2019.01.007

An, P., Bakker, S., Ordanovski, S., Taconis, R., Paffen, C. L., and Eggen, B. (2019). "Unobtrusively enhancing reflection-in-action of teachers through spatially distributed ambient information," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow), 1–14. doi: 10.1145/3290605.3300321

Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., and Graesser, A. C. (2010). Better to be frustrated than bored: the incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *Int. J. Hum. Comput. Stud.* 68, 223–241. doi: 10.1016/j.ijhcs.2009.12.003

Baker, R. S., Nasiar, N., Ocumpaugh, J.L., Hutt, S., Andres, J.M.A.L., Slater, S., et al. (2021). "Affect-targeted interviews for understanding Student frustration," in *Proceedings of the International Conference on Artificial Intelligence and Education*, eds I. Roll, D. McNamara, S. Sosnovsky, R. Luckin, and V. Dimitrova. 52–63. doi: 10.1007/978-3-030-78292-4_5

Baker, R. S., and Siemens, G. (2022). "Learning analytics and educational data mining," in *Cambridge Handbook of the Learning Sciences*, ed R. K. Sawyer, 3rd ed. (Cambridge, UK: Cambridge University Press), 259–278. doi: 10.1017/9781108888295.016

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660. doi: 10.1017/S0140525X99002149

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645. doi: 10.1146/annurev.psych.59.103006.093639

Blikstein, P., and Worsley, M. (2016). Multimodal learning analytics and education data mining: using computational technologies to measure complex learning tasks. *J. Learn. Anal.* 3, 220–238. doi: 10.18608/jla.2016.32.11

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., et al. (2021). On the opportunities and risks of foundation models. *arXiv.* [preprint]. doi: 10.48550/arXiv.2108.07258

Burgess, C., and Lund, K. (1997). Modelling parsing constraints with high-dimensional context space. *Lang. Cogn. Process.* 12, 1–34.

Cassell, J. (2001). Embodied conversational agents: representation and intelligence in user interfaces. *AI Mag.* 22, 67–83. doi: 10.1609/aimag.v22i4.1593

Chan, M. (2021). This AI reads children's emotions as they learn. *CNN Business.* February 17, 2021.

Closser, A. H., Erickson, J. A., Smith, H., Varatharaj, A., and Botelho, A. F. (2021). Blending learning analytics and embodied design to model students' comprehension of measurement using their actions, speech, and gestures. *Int. J. Child Comput. Interact.* 32:100391. doi: 10.1016/j.ijcci.2021.100391

Collins, A. M., and Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychol. Rev.* 82, 407. doi: 10.1037/0033-295X.82.6.407

Crawford, K. (2021). Time to regulate AI that interprets human emotions. *Nature* 592, 167. doi: 10.1038/d41586-021-00868-5

CRFM (2021). *Workshop on Foundation Models.* Stanford University Human-Centered Artificial Intelligence. Available online at: https://crfm.stanford.edu/workshop.html

Crigger, E., Reinbold, K., Hanson, C., Kao, A., Blake, K., and Irons, M. (2022). Trustworthy augmented intelligence in health care. *J. Med. Syst.* 46, 1–11. doi: 10.1007/s10916-021-01790-z

Dubova, M., Galesic, M., and Goldstone, R. L. (2022). Cognitive science of augmented intelligence. *Cogn. Sci.* 46:e13229.

Gallagher, S. (2018). *A Well-Trodden Path: From Phenomenology to Enactivism.* Oslo: Shaun Gallagher Filosofisk Suplement.

Glenberg, A. M. (1997). What memory is for: creating meaning in the service of action. *Behav. Brain Sci.* 20, 41–50. doi: 10.1017/S0140525X97470012

Glenberg, A. M., and Robertson, D. A. (2000). Symbol grounding and meaning: a comparison of high-dimensional and embodied theories of meaning. *J. Mem. Lang.* 43, 379–401. doi: 10.1006/jmla.2000.2714

Harnad, S. (1990). The symbol grounding problem. *Phys. D Nonlinear Phenom.* 42, 335–346. doi: 10.1016/0167-2789(90)90087-6

Harper, D. J., Ellis, D., and Tucker, I. (2022). "Covert aspects of surveillance and the ethical issues they raise," in *Ethical Issues in Covert, Security and Surveillance Research Advances in Research Ethics and Integrity*, eds R. Iphofen, and D. O'Mathúna (Bingley: Emerald Publishing Limited), 177–197. doi: 10.1108/S2398-601820210000008013

Havas, D. A., Glenberg, A. M., Gutowski, K. A., Lucarelli, M. J., and Davidson, R. J. (2010). Cosmetic use of botulinum toxin-A affects processing of emotional language. *Psychol. Sci.* 21, 895–900. doi: 10.1177/0956797610374742

[HLEG-AI] High-Level Expert Group on Artificial Intelligence (2019). *Ethics Guidelines for Trustworthy AI.* Brussels: European Commission.

Hutt, S., Baker, R. S., Ocumpaugh, J., Munshi, A., Andres, J. M. A. L., Karumbaiah, S., et al. (2022). "Quick red fox: an app supporting a new paradigm in qualitative research on AIED for STEM," in *Artificial Intelligence in STEM Education: The Paradigmatic Shifts in Research, Education, and Technology*, eds F. Ouyang, P., Jiao, B. M. McLaren, and A. H. Alavi (Boca Raton, FL: CRC Press), 319–332. doi: 10.1201/9781003181187-26

Järvelä, S., Järvenoja, H., and Malmberg, J. (2019). Capturing the dynamic and cyclical nature of regulation: methodological progress in understanding socially shared regulation in learning. *Int. J. Comput. Support. Collab. Learn.* 14, 425–441. doi: 10.1007/s11412-019-09313-2

Landauer, T. K., and Dumais, S. T. (1997). A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol. Rev.* 104, 211. doi: 10.1037/0033-295X.104.2.211

Laukkonen, R. E., Ingledew, D. J., Grimmer, H. J., Schooler, J. W., and Tangen, J. M. (2021). Getting a grip on insight: real-time and embodied Aha experiences predict correct solutions. *Cogn. Emot.* 35, 918–935. doi: 10.1080/02699931.2021.1908230

Leelawong, K., and Biswas, G. (2008). Designing learning by teaching agents: the Betty's Brain system. *Int. J. Artif. Intell. Educ.* 18, 181–208.

Liu, Z., Kong, X., Chen, H., Liu, S., and Yang, Z. (2023). MOOC-BERT: automatically identifying learner cognitive presence from MOOC discussion data. *IEEE Transact. Learn. Technol.* 31:1–14. doi: 10.1109/TLT.2023.3240715

Liu, Z., Yang, C., Rüdian S., Liu, S., Zhao, L., and Wang, T. (2019). Temporal emotion-aspect modeling for discovering what students are concerned about in online course forums. *Interact. Learn. Environ.* 27, 598–627. doi: 10.1080/10494820.2019.1610449

Macrine, S., and Fugate, J. (2022). *Movement Matters: How Embodied Cognition Informs Teaching and Learning.* Cambridge, MA: MIT Press. doi: 10.7551/mitpress/13593.001.0001

McClelland, J. L., Rumelhart, D. E., and PDP Research Group (1986). *Parallel Distributed Processing* (Vol. 2). Cambridge, MA: MIT press.

McKinney de Royston, M., Lee, C., Nasir, N. S., and Pea, R. (2020). Rethinking schools, rethinking learning. *Phi Delta Kappan* 102, 8–13. doi: 10.1177/0031721720970693

Monkaresi, H., Bosch, N., Calvo, R. A., and D'Mello, S. K. (2016). Automated detection of engagement using video-based estimation of facial expressions and heart rate. *IEEE Transact. Affect. Comput.* 8, 15–28. doi: 10.1109/TAFFC.2016.2515084

Nathan, M. J. (2021). *Foundations of Embodied Learning: A Paradigm for Education.* London: Routledge. doi: 10.4324/9780429329098

Newen, A., De Bruin, L., and Gallagher, S. (Eds) (2018). *The Oxford Handbook of 4E Cognition.* Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780198735410.001.0001

Ocumpaugh, J., Hutt, S., Andres, J. M. A. L., Baker, R. S., Biswas, G., Bosch, N., et al. (2021). "Using qualitative data from targeted interviews to inform rapid AIED development," in *Proceedings of the 29th International Conference on Computers in Education* (Bangkok).

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., et al. (2022). Training language models to follow instructions with human feedback. *arXiv.* [preprint] doi: 10.48550/arXiv.2203.02155

Papert, S. A. (2020). *Mindstorms: Children, Computers, and Powerful Ideas.* New York, NY: Basic books.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nat. Rev. Neurosci.* 6, 576–582. doi: 10.1038/nrn1706

Resnick, L. B. (1987). The 1987 presidential address learning in school and out. *Educ. Res.* 16, 13–54. doi: 10.3102/0013189X016009013

Rumelhart, D. E., McClelland, J. L., and PDP Research Group (1988). *Parallel Distributed Processing*, Vol. 1. Cambridge, MA: MIT press.

Schneider, B., and Pea, R. (2013). "Using eye-tracking technology to support visual coordination in collaborative problem-solving groups," in *To See the World and a Grain of Sand: Learning across Levels of Space, Time, and Scale: CSCL 2013 Conference Proceedings Volume 1—Full Papers and Symposia*, eds N. Rummel, M. Kapur, M. Nathan, and S. Puntambekar (Madison, WI: International Society of the Learning Sciences), 406–413.

Schneider, B., and Radu, I. (2022). "Augmented reality in the learning sciences," in *The Cambridge Handbook of the Learning Sciences*, ed R. K. Sawyer, 3rd ed. (Cambridge: Cambridge University Press), 340–361. doi: 10.1017/9781108888295.021

Searle, J. R. (1980). Minds, brains, and programs. *Behav. Brain Sci.* 3, 417–424. doi: 10.1017/S0140525X00005756

Shapiro, L. (2019). *Embodied Cognition,* 2nd ed. New York, NY: Routledge. doi: 10.4324/9781315180380

Shapiro, L., and Stolz, S. A. (2019). Embodied cognition and its significance for education. *Theory Res. Educ.* 17, 19–39. doi: 10.1177/1477878518822149

Shvarts, A., and Abrahamson, D. (2019). Dual-eye-tracking Vygotsky: a microgenetic account of a teaching/learning collaboration in an embodied interaction technological tutorial for mathematics. *Learn. Cult. Soc. Interact.* 22, 100316. doi: 10.1016/j.lcsi.2019.05.003

Sinha, T. (2021). Enriching problem-solving followed by instruction with explanatory accounts of emotions. *J. Learn. Sci.* 31, 151–198. doi: 10.1080/10508406.2021.1964506

Suresh, A., Jacobs, J., Lai, V., Tan, C., Ward, W., Martin, J. H., et al. (2021). Using transformers to provide teachers with personalized feedback on their classroom discourse: the TalkMoves application. *arXiv.* [preprint]. doi: 10.48550/arXiv.2105.07949

Timms, M. J. (2016). Letting artificial intelligence in education out of the box: educational cobots and smart classrooms. *Int. J. Artif. Intell. Educ.* 26, 701–712. doi: 10.1007/s40593-016-0095-y

Tyson, M. (2020). *Educational Leadership in the Age of Artificial Intelligence* [Dissertation]. Atlanta, GA: Georgia State University.

Varela, F. J., Thompson, E., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience.* Cambridge, MA: MIT Press. doi: 10.7551/mitpress/6730.001.0001

Walkington, C. A. (2013). Using adaptive learning technologies to personalize instruction to student interests: the impact of relevant contexts on performance and learning outcomes. *J. Educ. Psychol.* 105, 932. doi: 10.1037/a0031882

Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev.* 9, 625–636. doi: 10.3758/BF03196322

Worsley, M., and Blikstein, P. (2018). A multimodal analysis of making. *Int. J. Artif. Intell. Educ.* 28, 385–419. doi: 10.1007/s40593-017-0160-1