Received: 20 November 2022 | Accepted: 16 March 2023

DOI: 10.1111/bjet.13320

REVIEW

British Journal of Educational Technology BERA

The role of learning theory in multimodal learning analytics

Michail Giannakos[1] | **Mutlu Cukurova**[2]

[1]Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway

[2]UCL Knowledge Lab, University College London, London, UK

Correspondence
Michail Giannakos, Department of Computer Science, Norwegian University of Science and Technology, Sem Sælandsvei 9, NO-7491, Trondheim, Norway.
Email: michailg@ntnu.no

Abstract

This study presents the outcomes of a semi-systematic literature review on the role of learning theory in multimodal learning analytics (MMLA) research. Based on previous systematic literature reviews in MMLA and an additional new search, 35 MMLA works were identified that use theory. The results show that MMLA studies do not always discuss their findings within an established theoretical framework. Most of the theory-driven MMLA studies are positioned in the cognitive and affective domains, and the three most frequently used theories are embodied cognition, cognitive load theory and control–value theory of achievement emotions. Often, the theories are only used to inform the study design, but there is a relationship between the most frequently used theories and the data modalities used to operationalize those theories. Although studies such as these are rare, the findings indicate that MMLA affordances can, indeed, lead to theoretical contributions to learning sciences. In this work, we discuss methods of accelerating theory-driven MMLA research and how this acceleration can extend or even create new theoretical knowledge.

KEYWORDS
learning theories, multimodal data, multimodal learning analytics

Practitioner notes

What is already known about this topic

- Multimodal learning analytics (MMLA) is an emerging field of research with inherent connections to advanced computational analyses of social phenomena.
- MMLA can help us monitor learning activity at the micro-level and model cognitive, affective and social factors associated with learning using data from *both* physical and digital spaces.
- MMLA provide new opportunities to support students' learning.

What this paper adds

- Some MMLA works use theory, but, overall, the role of theory is currently limited.
- The three theories dominating MMLA research are embodied cognition, control–value theory of achievement emotions and cognitive load theory.
- Most of the theory-driven MMLA papers use theory 'as is' and do not consider the analytical and synthetic role of theory or aim to contribute to it.

Implications for practice and/or policy

- If the ultimate goal of MMLA, and AI in Education in general, research is to understand and support human learning, these studies should be expected to align their findings (or not) with established relevant theories.
- MMLA research is mature enough to contribute to learning theory, and more research should aim to do so.
- MMLA researchers and practitioners, including technology designers, developers, educators and policy-makers, can use this review as an overview of the current state of theory-driven MMLA.

INTRODUCTION

Learning is a complex phenomenon moderated by numerous factors related not only to intraindividual aspects such as previous knowledge and metacognition but also to the social and contextual factors in which the learner is situated. Learning analytics (LA) uses 'digital traces' left behind when learners interact with technologies and the learning context, making it possible to model some of these factors to support learning. However, one of the most frequent critiques of LA and artificial intelligence (AI) in education (AIED) research has long been associated with challenges related to capturing the complexity of learning via logged data from digital platforms (eg, the issue of reductive representations[1]). In recent years, significant progress has been made to understand and support the multimodal nature of learning (Kress & van Leeuwen, 2001) via data from different modalities and interfaces instead of merely focusing on digital interactions on a computer screen. To some extent, the use of multimodal data (eg, via cameras and audio) has been central to learning-sciences research for several decades (eg, qualitative video analysis and discourse analysis; Derry et al., 2010). However, the use of multimodal data requires significant expertise and manual work (eg, chunking, wayfinding and annotation; ibid). In recent years, the proliferation of new methods, devices, models and algorithms has enabled the continuous, unobtrusive and automated collection and sensemaking of multimodal data to support learning (Blikstein & Worsley, 2016). These developments have given rise to a new research stream at the intersection of multimodal data, learning sciences and advanced statistical and computational analyses through analytics and AI: Multimodal LA (MMLA).

It is worth acknowledging early on in the manuscript that the MMLA community is not necessarily the exclusive or inaugural research community to explore multimodal data and computational techniques in the context of learning. Colleagues engaged in related fields and communities such as AIED (ie, D'Mello et al., 2006), Educational Data Mining (EDM) (Craig et al., 2008) and User Modelling, Adaptation and Personalization (UMAP) (D'mello & Graesser, 2010) have previously conducted research that uses computationally processed multimodal data. Notably, affective computing research within these communities has a significant history of using multiple data modalities for the analysis and enhancement of learning (ie, Baker et al., 2010). Although MMLA was initiated as a separate community (following LA paradigm and focus on empowering instructors/learners instead of automated adaption with no human in the loop; Siemens & Baker, 2012) its goals and developments building upon the ongoing and previous work from the EDM, AIED, UMAP and other related communities. Similarly as the LA community leveraged the work and methods (eg, interaction analysis, use of computer logs) of the greater learning technology community, MMLA seeks to advance the work previously published in the aforementioned neighbouring fields. In contrast to prior research, perhaps distinguishing features of MMLA are its emphasis on empowerment and amplification of teaching and learning (Giannakos et al., 2022) (instead of full automation and adaption) and on facilitating learning experiences where the computer screen is not always the central focus or object of interaction (Worsley et al., 2021). Those features require a shift from using MMLA technology to automate and outsource teaching and learning to increasing bandwidth between humans and technology and enhancing teaching and learning through human–MMLA collaboration (multimodal data, AI and other computational techniques) and from focusing data exclusively from digital interaction spaces to data sources and modalities from physical spaces.

Indeed, the synergistic relationship between the advances of AI and multimodal interaction data to extend LA data sources, methodologies and tools (Giannakos et al., 2022) as well as to increase the variety of learning activities for which analytics technologies can act as a virtual observer and analyst (Ochoa, 2017) has immediate value. Although little is known regarding the research paradigms and theoretical considerations that drive MMLA design, development and implementation as well as the potential contributions of MMLA to theoretical considerations on learning, several ideas have been put forward. MMLA research, with the qualities introduced from the confluence of multimodal data and advanced computational analyses, can improve our understanding of how humans learn by providing information on the cognitive, affective and social aspects of learning, which would not be possible with mainstream LA data and techniques (Cukurova et al., 2020). This is because MMLA can monitor the learning activity at the micro-level and help us model humans' cognitive, affective and social factors associated with learning processes. Although traditional educational research and analytics have long examined learners' cognitive, affective and social factors via various methods (eg, knowledge tests, standardized surveys, think-aloud methods, observations, interviews and logged interactions), they are limited when it comes to investigating the complex and dynamic constructs of learning (Abrahamson et al., 2022; Richardson & Chemero, 2014; Tancredi et al., 2021). Alternatively, MMLA can provide information in a temporal and unobtrusive manner (Giannakos et al., 2022). This momentary and ecologically valid information provides an opportunity for a paradigm shift in the field of learning sciences, as this new source of information might not necessarily fit existing paradigms that rely on a static notion of learning. Therefore, the produced realities from MMLA research have the potential to 'strike back' (Kuhn, 1962) and reinforce the development of new theories, extend current ones or even force the field to generate new ones that account for the dynamic nature of learning (eg, dynamic systems theory or complex adaptive systems perspective; Ouyang et al., 2022). In this way, MMLA has the capacity to challenge established 'truths' and allow different theoretical lenses to further our knowledge of how humans learn.

Similarly, interdisciplinary research from the learning sciences has provided insights concerning the way humans learn, and, as a result, we now have an improved understanding of how best to teach and train people. As argued by Luckin and Cukurova (2019) and Cukurova (2018), this knowledge should inform the collection, analysis and interpretation of data from multimodal learning interactions through deductive investigations. In a recent inspiring editorial, Abrahamson et al. (2022) discuss the potential of the confluence of MMLA and learning theories. In particular, they highlight that the intersection of MMLA and embodied design represents a promising discipline and that coordination of this line of inquiry into a theoretically coherent research program has the potential for important implications to create theory-driven empirically validated technological-learning environments and thereby advance the future development of the learning sciences. In the same vein, our proposition is that MMLA can reinforce our current understanding of how humans learn by investigating how the multimodal data produced during the learning process and the advanced computational analyses used for their sensemaking align with or complement contemporary theoretical knowledge.

Despite the meritorious propositions above, our knowledge about the relationship between learning theories and MMLA research is scarce at best. This review paper addresses this gap aiming to answer the following research question (RQ): *What is the role of learning theory in MMLA research and to what extent can MMLA research advance learning theory?* To address this RQ, we investigate the following sub-RQs:

- (RQ1): Which theoretical positions and theories of learning are used in MMLA research, and how are they used?
- (RQ2): What is the relationship between the theories used in MMLA and the data modalities?
- (RQ3): What is the relationship between the theories used in MMLA and the intended goals of researchers?

To the best of our knowledge, this is the first literature review investigating the use of learning theory in MMLA. Previous reviews have described cases where MMLA can enrich our understanding beyond what can be done with mainstream LA (Blikstein & Worsley, 2016) or provide a conceptual framework using handpicked studies (Di Mitri et al., 2018). More recent works have also systematically reviewed the state of the MMLA field but with different foci, including a focus on MMLA capabilities (Sharma & Giannakos, 2020), on the evidence of impact and ethical considerations of MMLA research and practice (Alwahaby et al., 2022a), on the methodological developments in the field (Chua et al., 2019), and on MMLA research with children (Crescenzi-Lanna, 2020). Despite the importance of these review studies, they have limited focus (or none at all) on the role of theory. This paper presents a semi-systematic literature review of the field to examine the relationship between the theories of learning and MMLA research. In particular, this review will allow us to identify which theoretical positions and theories of learning are the most common in MMLA research, how these theories are being used, their relation to the data modalities and advanced computational analyses used, and the potential of MMLA research to advance the learning theories.

METHODS

The use of full or partial systematic literature reviews in learning-technologies research has increased in recent years (Eriksson et al., 2022). Instead of focusing on quantitative data, semi-systematic reviews identify themes, theoretical perspectives and other qualitative information related to the topic (Wong et al., 2013). These review types can be particularly helpful for a historical topic overview, for developing a theoretical understanding, for creating a research agenda for a given field (Snyder, 2019) and for identifying components of a theoretical concept (Ward et al., 2009). Given our interest in improving our understanding of the relationship between learning theory and MMLA research (rather than summarizing the topics), we opted for a semi-systematic approach. This approach allowed us to have non-systematic components such as a versatile search strategy (additional search of relevant papers to be added to the corpus from two systematic literature reviews), a mixed-method analysis of the selected research papers (both qualitative and quantitative) and a focus on constructing a topic overview and theoretical understanding with critical discussions when labelling and synthesizing the findings (instead of focusing on the analytical synthesis of the available evidence of impact; Snyder, 2019).

Collection of papers

To ensure a high-quality literature review on the role of learning theory in MMLA research, we first identified the corpus. The selection of papers was primarily based on recent systematic MMLA literature reviews by Sharma and Giannakos (2020) and Alwahaby et al. (2022a) for 2010–20 and an additional search by the authors. In particular, we manually reviewed the papers highlighted by Sharma and Giannakos (2020)[2] and Alwahaby et al. (2022a)[3] and identified 28 papers from both that explicitly address theoretical positions (13 and 15 papers, respectively). Moreover, we manually checked the proceedings of LAK 2021–22 and recent special issues by Abrahamson et al. (2022), Azevedo and Gašević (2019) and Cukurova et al. (2020). We identified three additional papers from LAK 2021–22, three from Abrahamson et al. (2022) and one from Azevedo and Gašević (2019). Therefore, the final corpus consisted of 35 research papers. This search strategy followed the guidelines of the selected approach (a search combination of systematic and non-systematic characteristics) and allowed us to leverage previous relevant systematic searches with a focused manual search on highly relevant outputs driven by the professional knowledge and experience of the researchers (MMLA special issues and the primary LA venue).

Data coding and analysis

To consolidate the essence and main focus of the studies and to extract the information required to address our RQs, we coded the identified studies according to specific areas. In particular, each study was analysed using the following coding categories (CCs):

CC 1 *What theories inform the research?*
CC 2 *The focus of the learning domain* (*cognitive, social and affective*).

We examined theories that provided framing for grounding or interpreting the phenomena of the studies. The analysis included only the theories that are explicitly named or referenced in the papers, including overarching frameworks. Coders did not interpret implicit theoretical considerations. The data were captured by thorough reading and analysis (each author coded half with discussions and cross-coding conducted when needed, eg, if one of the coders felt uncertain). After coding, clustering was made using a bottom-up approach, by also considering the association of the theories (Schunk, 2012) and the associations described in the papers, the following three main learning theory clusters emerged:

1. Theories focusing primarily on the cognitive aspects of learning. These theories stress the acquisition of knowledge and skills, the formation of mental models and the processing of information and beliefs (Schunk, 2012). Prominent theories focusing primarily on the cognitive aspects of learning include Piaget's and Bruner's theories as well as theory associated with information processing (Schunk, 2012).
2. Theories focusing primarily on the affective aspects of learning. These theories stress the affective mediation assumption, which suggests that affective and motivational factors mediate the learning process. Prominent theories focusing primarily on the affective aspects of learning include Moreno's (2006) cognitive–affective theory of learning with media (CATLM) and Csikszentmihalyi's (1997) flow theory.
3. Theories focusing primarily on the social aspects of learning. These theories stress the idea that learning is a socially mediated process. All learning is mediated by tools such as language, symbols and signs, which are acquired/internalized during learners' social interactions with others and then used as mediators to support learning (Schunk, 2012). Although most learning theories address the social aspects of learning, in this category, we focus on situations where the social mediation of learning is the central (primary) element (eg, Vygotsky's theory; Moll, 2001). The cognitive, affective and social processes can hardly be regarded as completely separate mediating factors, and the boundaries between them are not always clear (eg, the cognitive–affective–social theory of learning in digital environments; Schneider et al., 2022). We labelled the domain of focus for each theory, as described in the papers of interest.

CC 3 *What is the role of the learning theory?* (*application, description, analysis and synthesis*).

To examine how theories of learning are used in MMLA research (second part of RQ1), based on Rogers (2012) and Eriksson et al. (2022), we used the following four roles that the theories take and conducted coding across the selected papers:

1. *The descriptive role* is when a theory is used to frame the work without further engagement (theory is mentioned without elaborating on its use).
2. *The application role* is when a theory is used to inform the work (eg, selection of data or analysis) without engaging with the merits or qualities of the theory itself.
3. *The analysis role* involves deeper engagement with theory with a focus on discussing its merits and qualities and even sometimes developing its inner concepts (extending or contextualizing it).
4. *The synthesis role* refers to works that not only discuss the merits and qualities of a theory but also compare and synthesize multiple theories with the goal of integrating multiple ways of knowing. The description and application roles use the theory as is, whereas the analysis and synthesis take on a more generative perspective.

CC 4 *What modalities are used and how?*

CC 5 *Main findings/outcomes/approaches as claimed*.

To address the relationship between the theories and data modalities used in MMLA research (RQ2) as well as the connection between the theories used and the approaches taken (RQ3), we identified the data modalities used and the intended goal of each paper. Then, we combined this information with the coded information on the theories used and constructed mind map diagrams[4] depicting the links between them, the goals and the data modalities.

CCs 1, 2 and 3 address RQ1; CCs 1 and 4 address RQ2; and CCs 1 and 5 address RQ3. We took notes while coding the papers and used this information to discuss the results of our analysis. It is important to highlight that papers were coded based on the reported information and that, in some cases, the information was very rich and, in others, it was fragmented.

The coding process was iterative with regular consensus meetings held between the two coders (authors of this paper), where unclear or in-explicit information was discussed. Although we did not conduct any systematic process to assess the reliability of our coding (eg, calculate Cohen's kappa or a similar index), we coded a small number of papers independently (four papers each) and discussed our results to develop a common understanding. This process provides a degree of reliability in terms of coding consistency and what Krippendorff (2018, p. 278) describes as reliability—'the degree to which members of a designated community concur on the readings, interpretations, responses to or uses of given texts or data'. To ensure coding completeness, we coded a number of additional items (see Appendix S1 for detailed coding).

RESULTS

Theoretical positioning domains in multimodal learning analytics research

The analysis shows that the cognitive domain is the most frequently targeted theoretical domain. All the papers (except five) are associated with the cognitive domain (19 focus exclusively on the cognitive domain and 11 on the confluence of the cognitive domain with either the affective or the social domain). The remaining papers focus exclusively on social (Martinez-Maldonado et al., 2019; Prieto et al., 2018; Yan et al., 2022) and affective (Lew & Tang, 2017) domains.

Theoretical positionings in the cognitive, affective and social domains

For papers focusing exclusively on the cognitive domain, the most commonly used theoretical positionings are EC ($n=11$) and cognitive load theory (CLT; $n=6$). Moreover, some studies use the theory of cognitive development (Piaget), dual-coding theory, distributed cognition and habituation. In terms of papers focusing on cognitive and affective domains, two papers use the control–value theory of achievement emotions (CVTAE) and the remaining papers use flow theory, cognitive process of attention and circumplex model of affect, EC/epistemological frames, cognition–arousal theory of emotions and CLT. Figure 1 gives an overview of the papers focusing exclusively on the cognitive domain (up) and the papers focusing on the intersection of the cognitive and affective domains as well as their primary and secondary theoretical positionings.

When it comes to the social domain, we found three papers focusing exclusively on the social domain and four papers focusing on the intersection of the social and cognitive domains. Papers focusing on the social domain mainly use socio-constructivist and collaborative learning theories and use MMLA to (automatically/semi-automatically) identify social and collaborative learning events (Figure 5). The four papers that focus on the intersection of social and cognitive domains focus on identifying social and collaborative learning events, but they also cater for MMLA related to learners' cognition (Figure 3).

Most of the MMLA studies that address the role of theory in their research focus on the cognitive domain, with a considerable number of works focusing on the intersection of cognitive and affective domains. The main theories used in papers focusing on the cognitive domain include EC and CLT, whereas the papers focusing on the affective domain mainly use CVTAE. A good number of works also focus on the social domain (either exclusively or at the intersection of the cognitive and social domains), and these leverage socio-constructivist

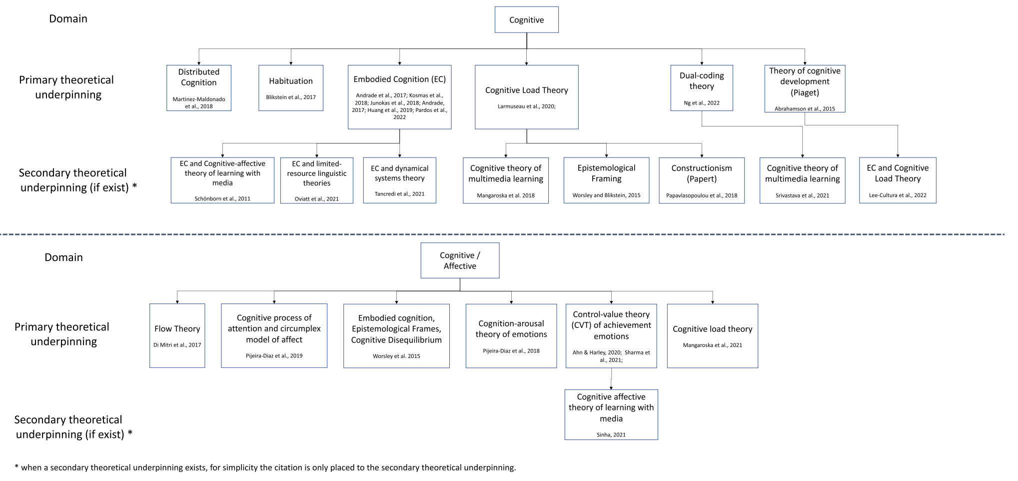

FIGURE 1 An overview (mind map diagram) of how multimodal learning analytics papers that focus on the cognitive domain and the intersection of cognitive and affective domains leverage theory.

theories, collaborative learning theories and socially shared regulated learning. Complete mind maps of how these theories are used in MMLA research can be found in Appendix S2.

Specific learning theories used in multimodal learning analytics research

The most widely used theory in MMLA research is EC from the cognitive sciences (Shapiro, 2010), which addresses the philosophical and theoretical questions related to the epistemic function of movement in cognition and learning and, according to many researchers, has a number of benefits for MMLA (Abrahamson et al., 2022). Contemporary MMLA research attempts to align MMLA with EC into theoretically coherent research, leveraging their confluence with the epistemic function of movement in learning, and this might be why we see many theory-related MMLA papers focusing on EC. These papers also focus on supporting STEM learning/concepts, populations that can be benefited from a high degree of embodiment (eg, math education; Andrade, 2017; Oviatt et al., 2015) and learners in a pre-operational stage or with special abilities (Kosmas et al., 2018). Another widely used theory in MMLA is CLT (Sweller, 1994), which was developed in the late 1980s from a problem-solving study. According to Sweller (1994), under CLT, instruction designs (and learning in general) should be used to reduce cognitive load in learners. CLT is being increasingly used, especially after the learning-technology community started using affordances that require high-level cognitive processing (eg, receiving automated hints and feedback while learning; Sun et al., 2019) and after the MMLA community started using techniques to infer different types of cognitive loads and mental effort (eg, eye-tracking, dual task and EEG).

Besides the focus on cognitive theories, MMLA research also focuses on affective theories. The most widely used theory from the affective domain is CVTAE (Pekrun, 2006). With CVTAE, MMLA research attempts to understand (and even sometimes account for) learners' emotional processes. The wide application of CVTAE in MMLA research can be justified by the inherent connection of MMLA with the affective computing community and the deep roots of intelligent tutoring systems with affective considerations as well as the numerous frameworks and tools that support CVTAE, for example, the work of Ekman et al. (2002) in the development of the Facial Action Coding System (FACS) and its application through the OpenFace library (Amos et al., 2016). Besides CVTAE, MMLA research leverages theories that consider learners' arousal state, such as studies by Di Mitri et al. (2017) that use flow theory and Csikszentmihalyi (1997) and theories that account for learner cognition and emotions (eg, cognitive–arousal theory of emotions and CATLM; see Figure 1).

When it comes to the social dimensions of learning, MMLA papers use a variety of theories, depending on the focus and context of each work. In particular, when it comes to physical classroom settings, several papers use collaborative learning theories (Martinez-Maldonado et al., 2019) and proxemics (Schneider & Blikstein, 2015; Yan et al., 2022) to account for learner movements in space and interactions during the learning process. Other works use learning regulation (Malmberg et al., 2019) and convergent conceptual change (Sharma et al., 2020) to account for learners' regulative behaviours and engagement when learning collaboratively.

How learning theories are used in multimodal learning analytics research

Our analysis shows that, in most papers ($n = 16$), theory is used as an application. The most common methods of applying theory include embedding it into, or using it to inform, the research design. For example, Prieto et al. (2018) applied Vygotsky's (1978) socio-constructivist theory by using MMLA to construct learners' and teachers' 'social plane' categories. Another example is Di Mitri et al. (2017), who applied flow theory (Csikszentmihalyi, 1997) by using MMLA to index learners' challenge and skill levels and, thus, predict their learning performances. Another example is Papavlasopoulou et al. (2018), who applied CLT (Sweller, 1994) by using MMLA to index learners' cognitive loads and associated these with their behaviours during a learning activity. Most of the studies found in this category apply CLT or EC to select MMLA measurements that can serve as proxies of learners' cognitive information or embodiment.

The second most frequent role involves using theory as a description (theory is mentioned without elaborating on its use), with 13 papers in this category. Examples of this category include papers by Lew and Tang (2017), who name CVTAE; Malmberg et al. (2019), who name the use of self-regulated learning theory; and Martinez-Maldonado et al. (2019), who name collaborative learning theories. In all cases, we see a description of the theory but no further elaboration on how it is used in relation to MMLA.

The third most frequent role involves using theory as an analytical tool (deeper engagement with the theory, discussing its merits and qualities and even sometimes developing its inner concepts), which was found in five papers (Abrahamson et al., 2015; Lee-Cultura et al., 2022; Oviatt et al., 2015, 2021; Tancredi et al., 2021). For example, Abrahamson et al. (2015) interpret their findings in ways that enable us to revisit, support, refine and, perhaps, elaborate on some of the seminal claims from Piaget's theory of genetic epistemology (eg, his insistence on the role of situated motor-action coordination in the process of reflective abstraction). Oviatt et al. (2021) elaborate on EC theory and limited resource theory, discuss their limitations in MMLA (eg, why the hands, in particular, provide a sensitive window for learners' mental state and why iconic gestures constitute an exception of reducing gestural activity as expertise increases) and outline how MMLA research qualities can complement these theories (eg, by leveraging insights that are not observable by the human eye). One last example is Tancredi et al. (2021), who combined EC with dynamic systems theory and coordination dynamics to extend its qualities and investigate the bimanual coordination patterns in different phases of mathematics embodied-design learning. Moreover, they discussed the benefits of going beyond the theoretical lenses of Piaget and Vygotsky and of embracing qualities from other theories, such as dynamic systems theory.

When it comes to using theory for synthesis, we did not find any papers that juxtapose multiple theories and integrate their concepts into a unified framework in relation to MMLA.

Although most of the studies in the analysis category juxtapose multiple theories, such as EC with dynamic systems theory (Tancredi et al., 2021), EC with limited resource theory (Oviatt et al., 2021) and EC with CLT (Lee-Cultura et al., 2022), the integration of these theories was limited in terms of synthesis. Therefore, most of the MMLA studies use theory 'as is' to inform their work (descriptive and application roles), whereas some studies take on a more generative perspective.

Alignment of learning theories with the intended goals and data modalities

EC is used in 11 papers (see Figure 2), and, in four of these, it is combined with other theories to account for other cognitive (Schönborn et al., 2011) and affective (Worsley et al., 2015) aspects of learning as well as the temporality of learning (Tancredi et al., 2021) and limited human resources (Oviatt et al., 2021). For studies that use EC, most of the intended goals are to test a concrete hypothesis that relates either to the use of certain parts of learners' bodies (eg, gestures; Andrade et al., 2017; Junokas et al., 2018), pose (Worsley et al., 2015), with a challenging learning domain and the benefits of embodiment—for example, math education (Oviatt et al., 2021; Tancredi et al., 2021), ecosystem education (Andrade, 2017) and biology (Schönborn et al., 2011)—or with a population with special abilities and the benefits of embodiment (eg, special educational needs; Kosmas et al., 2018). In terms of data modalities, 11 papers use motion data extracted either from typical videos/computer vision techniques (eg, Andrade, 2017; Oviatt et al., 2021), by devices that rely on motion detection such as Kinect and RealSense (which extract users' skeleton points; for example, Kosmas et al., 2018; Junokas et al., 2018) or by dedicated devices that log interactions in 3D or 2D space such as haptic 3D or 2D pens (eg, Oviatt et al., 2021; Schönborn et al., 2011) and touchscreen interactions (Pardos et al., 2022). Therefore, studies that leverage EC as the main theoretical lens are driven either by a challenging learning domain in which embodiment is vital (usually STEM domains), to support very young learners (eg, helping them to practice fine motor skills and hand-eye coordination) or to support learners with special abilities. To operationalize the benefits of EC to test hypotheses, these papers use motion-data modalities, with the most common data modality being motion cameras that generate a skeleton frame of joints and allow researchers to operationalize embodiment without any sophisticated computer vision technique and with limited constraints (eg, high accuracy and no personal data collection/processing). Figure 2 provides an overview of the relationships between the MMLA papers that leverage EC, their intended goals and the modalities used.

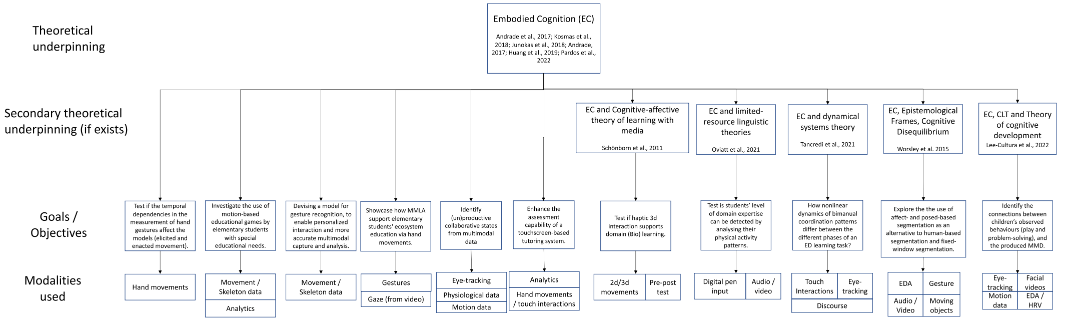

FIGURE 2 An overview (mind map diagram) of how multimodal learning analytics papers leverage EC, their intended goals and the modalities used.

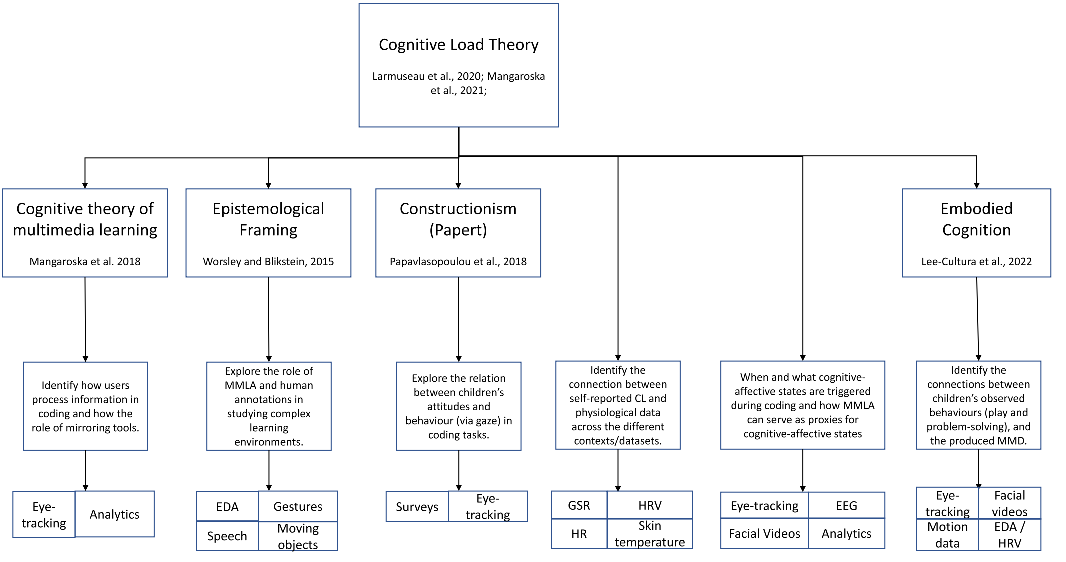

FIGURE 3 An overview (mind map diagram) of how multimodal learning analytics papers leverage CLT, their intended goals and the modalities used.

As described in Section 3.1.2, another frequently used theory in MMLA research is CLT, which is discussed in six papers (see Figure 3). In four of these works, CLT is combined with other theories to account for other cognitive (Mangaroska et al., 2018) aspects of learning as well as the embodiment (Lee-Cultura et al., 2022) and Papert's notion of constructionism (Papavlasopoulou et al., 2018). All papers focusing on CLT use MMLA with the goal of accessing learners' cognitive capacities (eg, how they process information; Mangaroska et al., 2018) and identifying differences and commonalities between the observed or self-reported cognition of learners and cognitive-related MMLA measurements that are unobservable by naked eye, such as CL based on galvanic skin responses (Larmuseau et al., 2020) and CL and arousal (Lee-Cultura et al., 2022). In terms of data modalities used, eye-tracking and physiological data (eg, EDA, GSR and HRV) are the most frequently used. This is unsurprising, as both eye-tracking and wristband-based physiological data have a high degree of reliability and objectivity, and they are ecologically valid (low intrusiveness) in indexing learners' cognitive processes (Giannakos et al., 2022). Figure 3 provides an overview of the relationships of the MMLA papers that leverage CLT, their intended goals and the modalities used.

The last theory with a relatively high frequency in MMLA research is CVTAE (Pekrun, 2006), with four papers using it. CVTAE is mostly used alone, but one paper combines it with a cognitive and affective theory of learning with media (see Figure 4). In all four papers, CVTAE is used to understand learners' emotions during the learning process. From Figure 4, it is evident that, in all studies, the goal is to examine the role of emotions in different contexts and different phases of learning. The data modalities used are almost exclusively learners' action units (AUs) from the FACS developed by Ekman et al. (2002); the exception is Lew and Tang (2017), who use learners' GSR and HR. Facial videos have an inherent connection to the affective domain and CVTAE due to a number of factors, such as the low cost of the apparatuses needed, the general acceptability of FACS as measurement units, the high ecological validity and the abundance of models and easy-to-use tools to measure AUs (eg, OpenFace library; Amos et al., 2016).

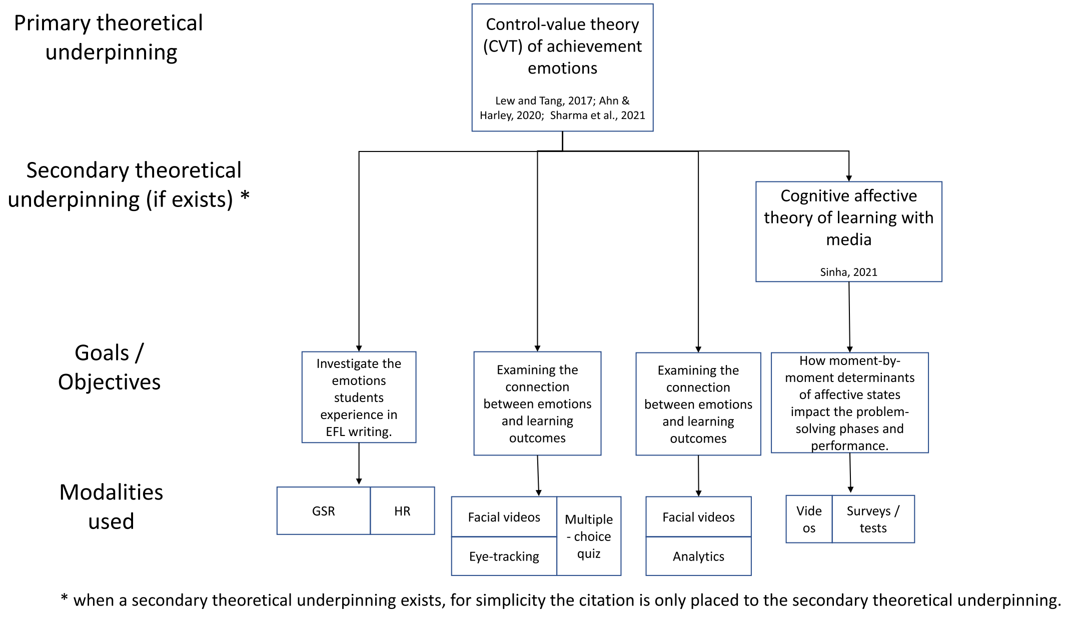

FIGURE 4 An overview (mind map diagram) of how multimodal learning analytics papers leverage CVTAE, their intended goals and the modalities used.

From the complete mind map diagrams provided in Appendix S2, we see that no other theory had a noticeable frequency (used more than twice). Therefore, investigating any potential alignment and relationship between these theories would be elusive. Nevertheless, it is interesting that there is no frequent use of a specific theory or a set of theories in the social domain. Therefore, we further examined the studies and respective theories for the social domain. Figures 5 and 6 provide an overview of how MMLA papers focus on the social domain, how they leverage theories, their intended goals and the data modalities used. When it comes to the three studies addressing exclusively the social domain (Figure 5), we see that they focus on Vygotsky's socio-constructivist theory of learning (Prieto et al., 2018), on collaborative learning theories (Martinez-Maldonado et al., 2019) and proxemics (Yan et al., 2022). The goal of these studies is to automatically tag or mine social interactions with respect to teachers' and/or learners' positions during physical learning activities. The data modalities used included positioning data collected from sensors in the room or wearables, which are sometimes complemented with additional data modalities such as analytics, eye-tracking and self-reports. Ultimately, these data are modelled to address positioning-based social interactions. An interesting observation here is that theory-relevant MMLA works focusing on the social domain only attempt to address physical settings and the positioning of teachers and learners. We could not find any papers addressing deeper social interactions that might have emerged from learners' and teachers' discourses or collaborative learning processes, despite the fact that automated speech analyses (eg, NLP) and artefact analysis (eg, co-development of an artefact such as a report or a code) might have revealed useful insights.

By examining the MMLA works that focus on the intersection of cognitive and social domains (Figure 6), it is evident that the focus on positioning data is less prominent; rather, there is a focus on assessing learners' cognitive development and associating it with their social interactions (eg, collaboration). Moreover, we see an effort to use learners' 'speech' with automated methods and assess their collaboration (Oviatt et al., 2015; Sharma et al., 2020) and co-regulation (Malmberg et al., 2019). These works provide a step forwards

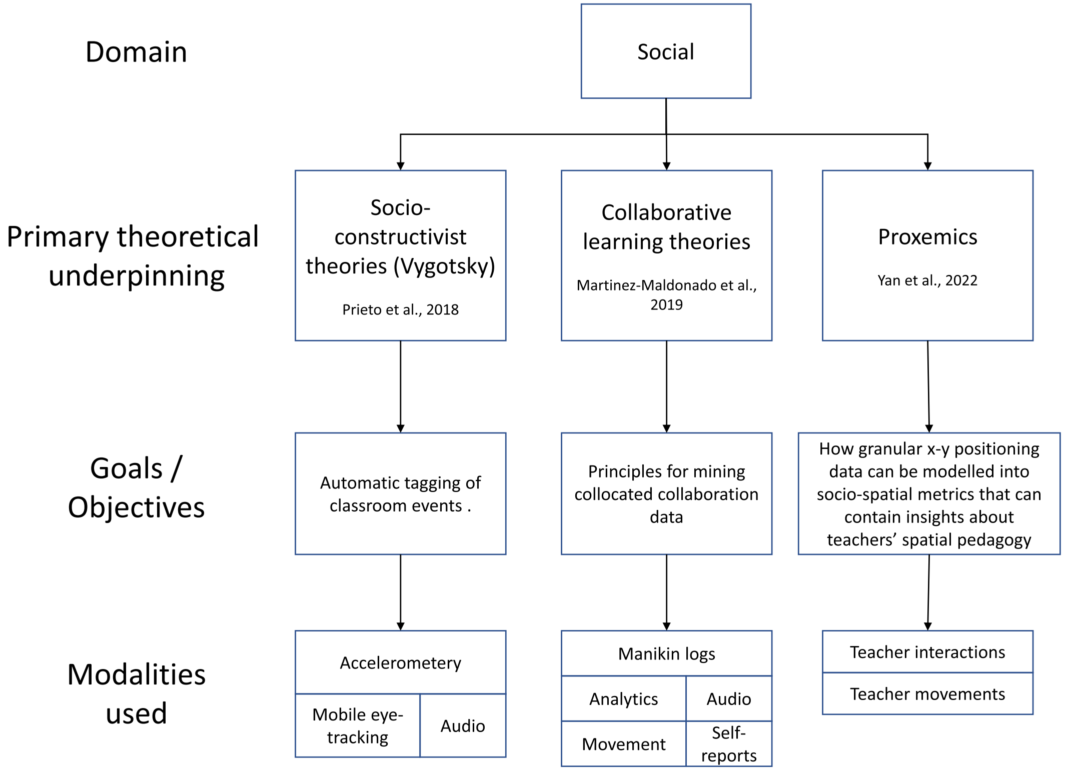

FIGURE 5 Mind map of how multimodal learning analytics papers focus on the social domain, how they leverage theories, what their intended goals and the modalities used.

by accounting for the speech modality and data features that recognize important elements of the social domain (eg, verbal communication and interruption), but they do not assess the quality of these elements (eg, via NLP), or they only rely on manual analysis of such quality evaluations. Given that most theories used (eg, self-regulated learning and convergent conceptual change) are operationalized and achieve their potential with deeper social interactions (eg, conflicting and misconceived) by conversational interaction, it is interesting that we could not find such works in the corpus.

DISCUSSION

From our analysis, it is clear that MMLA research does, indeed, leverage theories from different domains and traditions and that the majority can be categorized as cognitive theories. Cognition is a major part of understanding how humans learn. However, cognition can hardly be considered a merely intracranial biological activity, as it likely encompasses a person's total perception–action (Feiten et al., 2022), which are qualities that MMLA can potentially account for. This is one of the reasons why we see theories such as cognitive load theory being heavily used in MMLA research. In addition, any single dimension of cognition alone offers limited insights into how humans learn; rather, cognitive, affective and social processes are externalized during learning activities. These three processes are not proper 'parts' that can be neatly separated and re-assembled but, rather, form a system containing various interdependent elements and dynamic and adaptive interactions between them (Byrne & Callaghan, 2014). Therefore, insights about the interplay of these processes are important

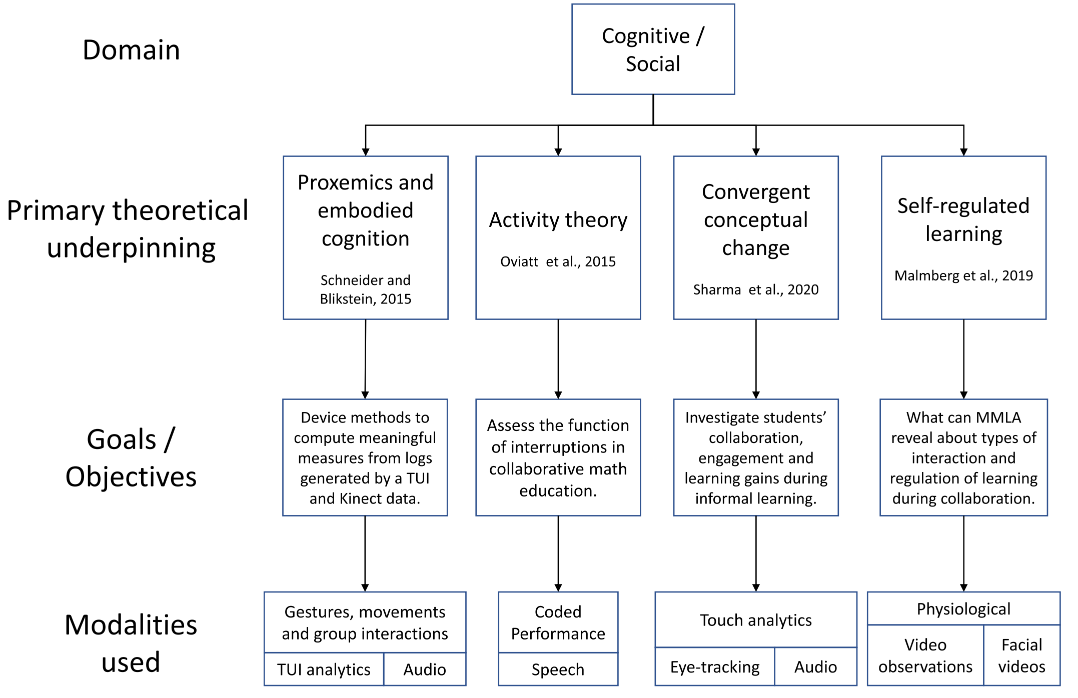

FIGURE 6 An overview (mind map diagram) of how multimodal learning analytics papers focus on the intersection of cognitive and social domains, how they leverage theories, what their intended goals are and the modalities used.

and have the capacity to further knowledge concerning how humans learn. One of the main goals of MMLA research is to access learners' cognitive, affective and social processes in a momentary and ecological manner (Giannakos et al., 2022). This review work indicates that, for MMLA research engaging in learning theories (either explicitly or implicitly), the main philosophical position is that our minds are not hidden within but are at least partially expressed in our behaviours and that these expressions are observable (although not always with the human eye). Accordingly, MMLA has the potential to be a 'new source of information' for understanding and supporting learning.

Three theories dominate MMLA research, namely EC, CLT and CVTAE. These theories are operationalized with data modalities of specific qualities. First, EC uses movement data (eg, skeleton data and hand movement) that are extracted from various devices (eg, motion sensors and cameras). This allows researchers not only to use learners' movements during the learning process (Kosmas et al., 2018; Pardos et al., 2022; Tancredi et al., 2021) but also to collect detailed movement data to improve our knowledge of the benefits of embodiment (eg, Andrade, 2017; Oviatt et al., 2021). This is in line with previous discussions on the intersection of MMLA and EC, which indicate the natural connection between, and potential of, those two domains (Abrahamson et al., 2022). Second, CLT uses data modalities that index our cognitive capacities (eg, pupillary diameter, GSR and HR/HRV) with the use of wearable and ubiquitous sensing (eg, eye-tracking, wristbands and EEG). Traditional methods of indexing learners' cognitive capacities (eg, how CLT is usually operationalized) are based on qualitative approaches, such as interaction analysis (investigating activities such as talking and problem solving) and think-aloud protocols (reveals the cognitive process but increases the CL and distracts the learner from the core task), as well as quantitative approaches such as standardized knowledge tests. Alternatively, MMLA provides new qualities that allow us to index CL in an unobtrusive and temporal manner. Such qualities can help us further CLT's initial goal (ie, proper design of learning that keeps cognitive load low; Sweller, 1994) by considering the dynamic and temporal nature of learning in CL. Third, CVTAE uses data modalities that index learners' emotions and affect. The most common method of operationalizing CVTAE is facial data (eg, facial videos). Learners' facial videos can be analysed via FACS (Ekman et al., 2002) and the OpenFace library (Amos et al., 2016), which computes the AUs for every frame in a video alongside a confidence value for each AU. Putting CVTAE into practice via MMLA is not only just one more way to operationalize this theory thanks to the qualities of MMLA (ie, temporality and computational approaches that deal with dynamic systems and phenomena), but it is also an opportunity to contribute back to the theory showing important connections between emotions and learning; for instance, Ahn and Harley (2020) found that emotions with positive valence might not always have a positive impact on learning and vice versa. In addition, MMLA provides opportunities to test some of the fundamental assumptions behind theories such as Ekman's AUs due to the scaling opportunities they provide (ie, AUs are cross-cultural values that can be used to detect emotions across the globe; Crawford, 2021).

Learning theories and multimodal learning analytics

An established theory base is often seen as a sign of a mature research discipline. Researchers within disciplines frequently debate about what constitutes an acceptable/legitimate theory and what is the proper role of theory in research and practice. Learning theories provide a scientifically acceptable frame to interpret observations and explain learning phenomena (Suppes, 1974). They serve as an intellectual leash that allows us to organize and systematically link knowledge. Without overarching frameworks such as theories, research findings are scattered collections of data (Schunk, 2012). Theories explain various facets of learning (Bruner, 1985); some theories emphasize the associations between stimuli and responses (eg, behavioural), whereas others explain learning with lenses such as information processing and student perceptions (eg, cognitive). The former is appropriate for explaining simpler forms of learning (eg, multiplication facts/new-word acquisition in language learning), and the latter is appropriate for explaining more complex forms of learning (eg, problem solving/writing essays; Schunk, 2012).

Similar to other theories (Psillos, 2018), learning theories are not static or forever lasting; rather, they are designed to be exposed to continuous testing, extensions and modifications. Learning theories are modified when new research provides conflicting evidence or suggests additional factors to include. For instance, initial information-processing theories were not directly applicable to classroom learning due to their exclusive focus on knowledge processing (ie, they did not consider other important elements of classroom settings). These theories were then revised to incorporate the needed situational factors and account for classroom learning (Schunk, 2012). Along the same lines, learning has traditionally been described by cognitive-oriented theories (Schneider et al., 2022). However, in recent decades, theory has been extended to shift the focus to additional processes, such as affective (eg, Moreno, 2006) and social processes (eg, Moll, 2001). Therefore, learning theories should be considered not only as sources to frame MMLA but also as falsifiable conceptual frameworks that can be revised based on new insights from MMLA research.

Most of what is understood about learning is inferential (Schunk, 2012), and thus, we do not observe it directly but only through what can be inferred (eg, the outcomes of the learning process). Therefore, the only way for the learning scientists and practitioners to be aware of student learning is by observing and assessing the ‘products’ of the learning process. MMLA, by leveraging and combining capabilities stemming on multimodal/sensor data and advanced computational analyses and AI, extends our observation capabilities by allowing us to infer processes that are not observable or solely interpretable by the human eye. Thus, MMLA has the potential to reinforce our understanding about the cognitive, affective and social processes of learning and thereby enhance contemporary learning theory. Moreover, the inferential nature of investigating learning suggests that other phenomena, which remain unknown, can also influence learning. Based on the findings of this review, particularly use theory for synthesis, it is fair to argue that MMLA provides an opportunity to interpret the cognitive, affective and social processes of learning in ways that were not possible before (eg, in a temporal and unobtrusive manner[5] and in dynamical analyses that model how the various processes are interwind; Fleuchaus et al., 2020; Richardson & Chemero, 2014). These new sources of information provide an opportunity for a paradigm shift via developing new theories (or extending established theories) that can account for new information and discoveries (Kuhn, 1962). New theoretical lenses that leverage insights from MMLA investigations are likely to enrich our current understanding and further challenge some of the foundational and contemporary considerations concerning how people learn as well as how the digital traces produced during learning can be used to understand and support learning and the environment in which it occurs. This review revealed several examples of such attempts from MMLA literature, including Abrahamson et al. (2015) challenging the seminal claims of Piaget.

Therefore, MMLA has the potential to inform and extend learning theory, in other words, to contribute to the principal part of the power and elegance of the learning sciences (diSessa & Cobb, 2004). At the same time, future MMLA research should also consider the great potential of leveraging on and developing knowledge that is more generative than an instantiation (an experiment or an artefact) and yet not at the scope of generalized theory (eg, ontological innovations, strong concepts). Such intermediate-level knowledge may inform the practice of researchers and practitioners (eg, see Abrahamson et al. (2011) for an example ontological innovation and Höök and Löwgren (2012) for examples of strong concepts) and greatly advance MMLA research and practice (see Giannakos, 2023).

Implications for practice

Given the number of MMLA papers published in top venues (more than 100) and the number of papers identified that refer to theory (35), it seems that theory is not always well engaged with or that the findings are not connected or contribute to theoretical knowledge. This observation is in line with a recent thematic landscape analysis (Papamitsiou et al., 2020) of LAK and JLA papers, which have highlighted that the focus on theory is limited in those leading LA venues. Even the carefully selected papers of this review referring to theory do so in a limited way, with a few exceptions. The value of incorporating learning theory to (i) enhance our comprehension of the learning process and (ii) improve instructional practices currently appears to be downplayed in most published work in MMLA. Previous work in AIED highlighted that expertise in AI is essential to developing learning models, yet it is not sufficient on its own (Rosé et al., 2019) and the constraints of data-driven models without the incorporation of theoretical frameworks (Mavrikis, 2010) are apparent. In light of these findings, it is crucial for the MMLA community to recognize the importance of engaging with learning theory in a mutually beneficial manner, to facilitate a more comprehensive understanding of the learning process and use (or even develop) a scientifically acceptable frame to rationalize observations coming from MMLA research and explain learning phenomena. Currently, most MMLA studies, indeed, offer rich, useful and well-written discussion and implications (eg, to practice or to design) sections; nevertheless, very few (eg, Abrahamson et al., 2015;

Lee-Cultura et al., 2022; Oviatt et al., 2021; Tancredi et al., 2021) provide any implications for general-level knowledge and theory.

Limitations

The goal of this paper was not to provide a comprehensive review of the MMLA research. (Please refer to recent systematic reviews cited for this.) Rather, in accordance with the used methodology of semi-systematic review (Snyder, 2019), the goal was to obtain an overview of the field to identify and understand potentially relevant research traditions and their respective implications qualitatively. We carefully selected the methods based on this objective. However, as with any method, methodological decisions entail certain limitations. First, the selected papers include publications gathered from recent systematic literature reviews in MMLA complemented by a manual search by the authors. Despite the advantages of reusability and transparency of the process (eg, transparent process of selecting and coding of papers), this decision may result in study bias (the selected corpora might exclude relevant papers that use different terminology, for example, from AIED, ITS and UMAP). Thus, the corpora selection corresponds to relevant and quality-checked publications, but, at the same time, it might exclude conceptually relevant papers that might use different nomenclature (eg, affective learner modelling or multimodal affect detection instead of MMLA). However, this is in line with the selected methodology that focuses on topics where reviewing every single potentially relevant paper is not possible (Snyder, 2019). Moreover, the input included in our analysis (ie, corpora from two recent MMLA review papers) presents clear insights and provides a topic overview. Second, the coding of the selected papers might pose another possible bias, as the elements included in the papers were not always described properly (eg, not explicitly or accurately). Moreover, the terminology used to describe different theories and theoretical notions did not always follow the same approach (eg, embodied learning, EC, embodiment theory of cognitive sciences and embodied design), and this might lead to inconsistencies when overviewing the topic. However, two senior researchers coded the papers and discussed unclear aspects, which were resolved in regular consensus meetings. Thus, the amount of missing or unclear information is limited and unlikely to affect the results significantly. Finally, in our review, we specifically searched for explicit mentions of theory in the reviewed manuscripts. As a result, some published papers that did not explicitly discuss theory but did incorporate frameworks of actions or domain-specific strategies that were derived from theoretical considerations when designing MMLA support may have been inadvertently excluded.

CONCLUSION

In this work, we presented a semi-structured literature review on the role of theory in the MMLA field. We analysed papers from the perspective of learning theory to identify relationships between different theories, data modalities and intended goals. Finally, based on an overview of how theory is used in MMLA research, we discussed the results and potential implications.

MMLA is a growing community with a distinct role in LA (Giannakos et al., 2022), and there is an acute need to facilitate our understanding of the role and utilization of learning theory in MMLA. The current findings suggest that MMLA studies make limited use of learning theory, with three theories dominating MMLA research, namely EC, CLT and CVTAE. There is consistency between the use of theories, the data modalities and indented goals, but the theories are mostly used ‘as is’: they are not fully synthesized with findings, and existing theories are rarely extended. In the future, MMLA research should strive from a mentality of 'good to have' connection to theory, to considering theoretical discussions as an integral part of MMLA research and practice. This will likely ensure that MMLA can be operationalized for best practice or extended/challenged to enhance our understanding of the field; it will also help ensure the creation of new theoretical knowledge that will extend the boundaries of knowledge.

CONFLICT OF INTEREST STATEMENT

There is no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this paper.

ETHICS STATEMENT

The work does not include primary collection of data. Appropriate permissions and ethical approval were requested and granted, as needed.

ORCID

Mutlu Cukurova https://orcid.org/0000-0001-5843-4854

ENDNOTES

[1] LA and AIED tend to represent learning phenomena as sequences of logged actions (eg, video navigation, response time, and answer correctness), reducing complex realities to a manageable set of variables allowing us to model and improve learning. Nevertheless, the employed methods and algorithms are by nature reductive.

[2] Appendix S2 of Sharma and Giannakos (2020) provides a detailed list of the 42 papers.

[3] The Alwahaby et al. (2022b) open dataset provides a detailed list of the 100 papers.

[4] A mind map is a diagram that visualizes and organizes information into a hierarchy, thereby showcasing the relationships among pieces of the whole (Buzan, 2018).

[5] However, we should recognize that the learning sciences, to some extent, have engaged with the use of multimodal data in the past (eg, ITS and learner modelling).

REFERENCES

Abrahamson, D., Trninic, D., Gutiérrez, J. F., Huth, J., & Lee, R. G. (2011). Hooks and shifts: A dialectical study of mediated discovery. *Technology, Knowledge and Learning*, *16*, 55–85.

Abrahamson, D., Worsley, M., Pardos, Z. A., & Ou, L. (2022). Learning analytics of embodied design: Enhancing synergy. *International Journal of Child–Computer Interactions*, *32*, 100409.

Alwahaby, H., Cukurova, M., Papamitsiou, Z., & Giannakos, M. (2022a). The evidence of impact and ethical considerations of multimodal learning analytics: A systematic literature review. In D. Di Mitri, M. Giannakos, R. Hammad, X. Ochoa, K. Sharma, & D. Spikol (Eds.), *The Multimodal Learning Analytics Handbook*. Springer.

Alwahaby, H., Cukurova, M., Papamitsiou, Z., & Giannakos, M. (2022b). Dataset on studies related to multimodal learning analytics (to assess the evidence of impact and ethical considerations of multimodal learning analytics). Mendeley Data, v1. https://data.mendeley.com/v1/datasets/3d8gr3jjk2/draft?a=8b1ea571-f9dc-4a79-a7ec-849431d587a6

Amos, B., Ludwiczuk, B., & Satyanarayanan, M. (2016). Openface: A general-purpose face recognition library with mobile applications. *CMU School of Computer Science*, *6*(2), 20.

Azevedo, R., & Gašević, D. (2019). Analyzing multimodal multichannel data about self-regulated learning with advanced learning technologies: Issues and challenges. *Computers in Human Behavior*, *96*, 207–210.

Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & 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. *International Journal of Human-Computer Studies*, *68*(4), 223–241.

Blikstein, P., & Worsley, M. (2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, *3*(2), 220–238.

Bruner, J. (1985). Models of the learner. *Educational Researcher*, *14*(6), 5–8.

Buzan, T. (2018). *Mind map mastery: The complete guide to learning and using the most powerful thinking tool in the universe*. Watkins Media.

Byrne, D., & Callaghan, G. (2014). *Complexity theory and the social sciences: The state of the art*. Routledge.

Chua, Y. H. V., Dauwels, J., & Tan, S. C. (2019). Technologies for automated analysis of co-located, real-life, physical learning spaces: Where are we now? In *Proceedings of the 9th international conference on learning analytics & knowledge* (pp. 11–20). ACM.

Craig, S. D., D'Mello, S., Witherspoon, A., & Graesser, A. (2008). Emote aloud during learning with AutoTutor: Applying the facial action coding system to cognitive–affective states during learning. *Cognition and Emotion*, *22*(5), 777–788.

Crawford, K. (2021). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Crescenzi-Lanna, L. (2020). Multimodal learning analytics research with young children: A systematic review. *British Journal of Educational Technology*, *51*(5), 1485–1504.

Csikszentmihalyi, M. (1997). *Flow and the psychology of discovery and invention*. Harper Perennial.

Cukurova, M. (2018). A syllogism for designing collaborative learning technologies in the age of AI and multimodal data. In *European conference on technology enhanced learning* (pp. 291–296). Springer.

Cukurova, M., Giannakos, M., & Martinez-Maldonado, R. (2020). The promise and challenges of multimodal learning analytics. *British Journal of Educational Technology*, *51*(5), 1441–1449.

Derry, S. J., Pea, R. D., Barron, B., Engle, R. A., Erickson, F., Goldman, R., Hall, R., Koschmann, T., Lemke, J. L., Sherin, M., & Sherin, B. L. (2010). Conducting video research in the learning sciences: Guidance on selection, analysis, technology, and ethics. *The Journal of the Learning Sciences*, *19*(1), 3–53.

Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2018). From signals to knowledge: A conceptual model for multimodal learning analytics. *Journal of Computer Assisted Learning*, *34*(4), 338–349.

diSessa, A. A., & Cobb, P. (2004). Ontological innovation and the role of theory in design experiments. *The Journal of the Learning Sciences*, *13*(1), 77–103.

D'Mello, S. K., Craig, S. D., Sullins, J., & Graesser, A. C. (2006). Predicting affective states expressed through an emote-aloud procedure from AutoTutor's mixed-initiative dialogue. *International Journal of Artificial Intelligence in Education*, *16*(1), 3–28.

D'mello, S. K., & Graesser, A. (2010). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-Adapted Interaction*, *20*, 147–187.

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial action coding system: Facial action coding system: The manual: On CD-ROM*. Research Nexus.

Eriksson, E., Baykal, G. E., & Torgersson, O. (2022). The role of learning theory in child-computer interaction—A semi-systematic literature review. In *Interaction design and children (IDC '22)* (19 Pages. ACM. https://doi.org/10.1145/3501712.3529728

Feiten, T. E., Holland, K., & Chemero, A. (2022). Scalar symmetries and strange tools: Learning analytics and embodied design as techno-cultural confluence. *International Journal of Child-Computer Interaction*, *31*, 100452. https://doi.org/10.1016/j.ijcci.2021.100452

Fleuchaus, E., Kloos, H., Kiefer, A. W., & Silva, P. L. (2020). Complexity in science learning: Measuring the underlying dynamics of persistent mistakes. *Journal of Experimental Education*, *88*(3), 448–469.

Giannakos, M. (2023). Theory and intermediate-level knowledge in multimodal learning analytics. In K. Bartimote, S. Howard, & D. Gašević (Eds.), *Theory informing and arising from learning analytics*. Springer Nature.

Giannakos, M., Spikol, D., Di Mitri, D., Sharma, K., Ochoa, X., & Hammad, R. (2022). Introduction to multimodal learning analytics. In D. Di Mitri, M. Giannakos, R. Hammad, X. Ochoa, K. Sharma, & D. Spikol (Eds.), *The multimodal learning analytics handbook* (pp. 3–28). Springer.

Höök, K., & Löwgren, J. (2012). Strong concepts: Intermediate-level knowledge in interaction design research. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *19*(3), 1–18.

Kress, G. R., & van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication* (Vol. 312). Arnold.

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. Sage Publications.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. University of Chicago Press.

Luckin, R., & Cukurova, M. (2019). Designing educational technologies in the age of AI: A learning sciences-driven approach. *British Journal of Educational Technology*, *50*(6), 2824–2838.

Mavrikis, M. (2010). Modelling student interactions in intelligent learning environments: Constructing bayesian networks from data. *International Journal on Artificial Intelligence Tools*, *19*(6), 733–753.

Moll, L. C. (2001). Through the mediation of others: Vygotskian research on teaching. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 111–129). American Educational Research Association.

Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Journal of Computer Assisted Learning*, *22*(3), 149–158.

Ochoa, X. (2017). Multimodal learning analytics. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *The handbook of learning analytics* (pp. 129–141). Society for Learning Analytics Research (SoLAR).

Ouyang, F., Xu, W., & Cukurova, M. (2022). An artificial intelligence driven learning analytics method to examine the collaborative problem solving process from a complex adaptive systems perspective. *International Journal of Computer-supported Collaborative Learning*, arXiv:2210.16059.
https://doi.org/10.1007/s11412-023-09387-z

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, *18*(4), 315–341.

Psillos, S. (2018). The Stanford encyclopedia of philosophy. In E. N. Zalta (Ed.), *Realism and theory change in science*. Stanford University https://plato.stanford.edu/archives/sum2018/entries/realism-theory-change/

Richardson, M. J., & Chemero, A. (2014). Complex dynamical systems and embodiment. In L. Shapiro (Ed.), *The Routledge handbook of embodied cognition* (pp. 39–50). Routledge.

Rogers, Y. (2012). HCI theory: Classical, modern, and contemporary. *Synthesis Lectures on Human-Centered Informatics*, *5*(2), 1–129.

Rosé, C. P., McLaughlin, E. A., Liu, R., & Koedinger, K. R. (2019). Explanatory learner models: Why machine learning (alone) is not the answer. *British Journal of Educational Technology*, *50*(6), 2943–2958.

Schneider, S., Beege, M., Nebel, S., Schnaubert, L., & Rey, G. D. (2022). The cognitive-affective-social theory of learning in digital environments (CASTLE). *Educational Psychology Review*, *34*, 1–38. https://doi.org/10.1007/s10648-021-09626-5

Schunk, D. H. (2012). *Learning theories an educational perspective* (6th ed.). Pearson.

Shapiro, L. (2010). *Embodied cognition*. Routledge.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339.

Sun, J. C. Y., Yu, S. J., & Chao, C. H. (2019). Effects of intelligent feedback on online learners' engagement and cognitive load: The case of research ethics education. *Educational Psychology*, *39*(10), 1293–1310.

Suppes, P. (1974). The place of theory in educational research. *Educational Researcher*, *3*(6), 3–10.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312.

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

Ward, V., House, A., & Hamer, S. (2009). Developing a framework for transferring knowledge into action: A thematic analysis of the literature. *Journal of Health Services Research & Policy*, *14*(3), 156–164.

Wong, G., Greenhalgh, T., Westhorp, G., Buckingham, J., & Pawson, R. (2013). RAMESES publication standards: Meta-narrative reviews. *Journal of Advanced Nursing*, *69*(5), 987–1004.

Worsley, M., Martinez-Maldonado, R., & D'Angelo, C. (2021). A new era in multimodal learning analytics: Twelve core commitments to ground and grow MMLA. *Journal of Learning Analytics*, *8*(3), 10–27.

PAPERS INCLUDED IN THE LITERATURE REVIEW

Abrahamson, D., Shayan, S., Bakker, A., & Van Der Schaaf, M. (2015). Eye-tracking Piaget: Capturing the emergence of attentional anchors in the coordination of proportional motor action. *Human Development*, *58*(4–5), 218–244.

Ahn, B. T., & Harley, J. M. (2020). Facial expressions when learning with a queer history app: Application of the control value theory of achievement emotions. *British Journal of Educational Technology*, *51*(5), 1563–1576.

Andrade, A. (2017). Understanding student learning trajectories using multimodal learning analytics within an embodied-interaction learning environment. In *Proceedings of the seventh international learning analytics & knowledge conference* (pp. 70–79). ACM.

Andrade, A., Danish, J. A., & Maltese, A. V. (2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, *4*(3), 18–46.

Blikstein, P., Gomes, J. S., Akiba, H. T., & Schneider, B. (2017). The effect of highly scaffolded versus general instruction on students' exploratory behavior and arousal. *Technology, Knowledge and Learning*, *22*(1), 105–128.

Di Mitri, D., Scheffel, M., Drachsler, H., Börner, D., Ternier, S., & Specht, M. (2017). Learning pulse: A machine learning approach for predicting performance in self-regulated learning using multimodal data. In *Proceedings of the seventh international learning analytics & knowledge conference* (pp. 188–197). ACM.

Huang, K., Bryant, T., & Schneider, B. (2019). *Identifying collaborative learning states using unsupervised machine learning on eye-tracking, physiological and motion sensor data*. International Educational Data Mining Society.

Junokas, M. J., Lindgren, R., Kang, J., & Morphew, J. W. (2018). Enhancing multimodal learning through personalized gesture recognition. *Journal of Computer Assisted Learning*, *34*(4), 350–357.

Kosmas, P., Ioannou, A., & Retalis, S. (2018). Moving bodies to moving minds: A study of the use of motion-based games in special education. *TechTrends*, *62*(6), 594–601.

Larmuseau, C., Cornelis, J., Lancieri, L., Desmet, P., & Depaepe, F. (2020). Multimodal learning analytics to investigate cognitive load during online problem solving. *British Journal of Educational Technology*, *51*(5), 1548–1562.

Lee-Cultura, S., Sharma, K., & Giannakos, M. (2022). Children's play and problem-solving in motion-based learning technologies using a multi-modal mixed methods approach. *International Journal of Child–Computer Interaction*, *31*, 100355.

Lew, L., & Tang, T. Y. (2017). "Beyond EFL writing anxiety": Tapping into the individual emotionality of proficient EFL writers through semi-structured analysis and wearable sensing technology. In *International conference on learning and collaboration technologies* (pp. 170–181). Springer.

Malmberg, J., Järvelä, S., Holappa, J., Haataja, E., Huang, X., & Siipo, A. (2019). Going beyond what is visible: What multichannel data can reveal about interaction in the context of collaborative learning? *Computers in Human Behavior*, *96*, 235–245.

Mangaroska, K., Sharma, K., Gašević, D., & Giannakos, M. (2022). Exploring students' cognitive and affective states during problem solving through multimodal data: Lessons learned from a programming activity. *Journal of Computer Assisted Learning*, *38*(1), 40–59.

Mangaroska, K., Sharma, K., Giannakos, M., Træteberg, H., & Dillenbourg, P. (2018). Gaze-driven design insights to amplify debugging skills: A learner-centred analysis approach. *Journal of Learning Analytics*, *5*(3), 98–119.

Martinez-Maldonado, R., Echeverria, V., Santos, O. C., Santos, A. D. P. D., & Yacef, K. (2018). Physical learning analytics: A multimodal perspective. In *Proceedings of the 8th international conference on learning analytics and knowledge* (pp. 375–379).

Martinez-Maldonado, R., Kay, J., Buckingham Shum, S., & Yacef, K. (2019). Collocated collaboration analytics: Principles and dilemmas for mining multimodal interaction data. *Human–Computer Interaction*, *34*(1), 1–50.

Oviatt, S., Hang, K., Zhou, J., & Chen, F. (2015). Spoken interruptions signal productive problem solving and domain expertise in mathematics. In *Proceedings of the 2015 ACM on international conference on multimodal interaction* (pp. 311–318). ACM.

Oviatt, S., Lin, J., & Sriramulu, A. (2021). I know what you know: What hand movements reveal about domain expertise. *ACM Transactions on Interactive Intelligent Systems*, *11*(1), 1–26.

Papavlasopoulou, S., Sharma, K., & Giannakos, M. N. (2018). How do you feel about learning to code? Investigating the effect of children's attitudes towards coding using eye-tracking. *International Journal of Child–Computer Interaction*, *17*, 50–60.

Papamitsiou, Z., Giannakos, M. N., & Ochoa, X. (2020). From childhood to maturity: Are we there yet? Mapping the intellectual progress in learning analytics during the past decade. In *Proceedings of the tenth international conference on learning analytics & knowledge* (pp. 559–568).

Pardos, Z. A., Rosenbaum, L. F., & Abrahamson, D. (2022). Characterizing learner behavior from touchscreen data. *International Journal of Child-Computer Interaction*, *31*, 100357.

Pijeira-Díaz, H. J., Drachsler, H., Järvelä, S., & Kirschner, P. A. (2019). Sympathetic arousal commonalities and arousal contagion during collaborative learning: How attuned are triad members? *Computers in Human Behavior*, *92*, 188–197.

Pijeira-Díaz, H. J., Drachsler, H., Kirschner, P. A., & Järvelä, S. (2018). Profiling sympathetic arousal in a physics course: How active are students? *Journal of Computer Assisted Learning*, *34*(4), 397–408.

Prieto, L. P., Sharma, K., Kidzinski, Ł., Rodríguez-Triana, M. J., & Dillenbourg, P. (2018). Multimodal teaching analytics: Automated extraction of orchestration graphs from wearable sensor data. *Journal of Computer Assisted Learning*, *34*(2), 193–203.

Schneider, B., & Blikstein, P. (2015). Unraveling students' interaction around a tangible interface using multimodal learning analytics. *Journal of Educational Data Mining*, *7*(3), 89–116.

Schönborn, K. J., Bivall, P., & Tibell, L. A. (2011). Exploring relationships between students' interaction and learning with a haptic virtual biomolecular model. *Computers & Education*, *57*(3), 2095–2105.

Sharma, K., & Giannakos, M. (2020). Multimodal data capabilities for learning: What can multimodal data tell us about learning? *British Journal of Educational Technology*, *51*(5), 1450–1484.

Sharma, K., Leftheriotis, I., & Giannakos, M. (2020). Utilizing interactive surfaces to enhance learning, collaboration and engagement: Insights from learners' gaze and speech. *Sensors*, *20*(7), 1964.

Sharma, K., Papavlasopoulou, S., & Giannakos, M. (2021). Faces don't lie: Analysis of children's facial expressions during collaborative coding. In *FabLearn Europe/MakeEd 2021—An international conference on computing, design and making in education* (pp. 1–10). ACM.

Siemens, G., & Baker, R. S. D. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 252–254). ACM.

Sinha, T. (2022). Enriching problem-solving followed by instruction with explanatory accounts of emotions. *Journal of the Learning Sciences*, *31*(2), 151–198.

Srivastava, N., Fan, Y., Rakovic, M., Singh, S., Jovanovic, J., Van Der Graaf, J., Lim, L., Surendrannair, S., Kilgour, J., Molenaar, I., Bannert, M., Moore, J. D., & Gasevic, D. (2022). Effects of internal and external conditions on strategies of self-regulated learning: A learning analytics study. In *LAK22: 12th international learning analytics and knowledge conference* (pp. 392–403). ACM.

Tancredi, S., Abdu, R., Abrahamson, D., & Balasubramaniam, R. (2021). Modelling nonlinear dynamics of fluency development in an embodied-design mathematics learning environment with recurrence quantification analysis. *International Journal of Child-Computer Interaction*, *29*, 100297.

Tzi-Dong Ng, J., Hu, X., & Que, Y. (2022). Towards multi-modal evaluation of eye-tracked virtual heritage environment. In *LAK22: 12th international learning analytics and knowledge conference* (pp. 451–457). ACM.

Worsley, M., & Blikstein, P. (2015). Leveraging multimodal learning analytics to differentiate student learning strategies. In *Proceedings of the fifth international conference on learning analytics and knowledge* (pp. 360–367). ACM.

Worsley, M., Scherer, S., Morency, L. P., & Blikstein, P. (2015). Exploring behavior representation for learning analytics. In *Proceedings of the 2015 ACM on international conference on multimodal interaction* (pp. 251–258). ACM.

Yan, L., Martinez-Maldonado, R., Zhao, L., Deppeler, J., Corrigan, D., & Gasevic, D. (2022). How do teachers use open learning spaces? Mapping from teachers' socio-spatial data to spatial pedagogy. In *LAK22: 12th international learning analytics and knowledge conference* (pp. 87–97). ACM.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Giannakos, M., & Cukurova, M. (2023). The role of learning theory in multimodal learning analytics. *British Journal of Educational Technology*, *00*, 1–22. https://doi.org/10.1111/bjet.13320