

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

A Measurement Model of Gestures in an Embodied Learning Environment: Accounting for Temporal Dependencies

Alejandro Andrade

Indiana University, USA
laandrad@indiana.edu

Joshua A. Danish

Indiana University, USA

Adam V. Maltese

Indiana University, USA

ABSTRACT: Interactive learning environments with body-centric technologies lie at the intersection of the design of embodied learning activities and multimodal learning analytics. Sensing technologies can generate large amounts of fine-grained data automatically captured from student movements. Researchers can use these fine-grained data to create a high-resolution picture of the activity that takes place during these student–computer interactions and explore whether the sequence of movements has an effect on learning. We present a use-case modelling of temporal data in an interactive learning environment with hand gestures, and discuss some validity threats if temporal dependencies are not accounted for. In particular, we assess how, if ignored, the temporal dependencies in the measurement of hand gestures might affect the goodness of fit of the statistical model and would affect the measurement of the similarity between elicited and enacted movement. Our findings show that accounting for temporality is crucial for finding a meaningful fit to the data. In using temporal analytics, we are able to create a high-resolution picture of how sensorimotor coordination correlates with learning gains in our learning system.

Keywords: Embodied cognition, embodied learning, hidden Markov models, optimal matching, temporal analytics

NOTES FOR PRACTICE

- Sensing technologies make available the design of learning environments engaging the body and movement. As large amounts of data become available, appropriate analytic techniques are required to make correct inferences about the learning that takes place in these environments.
- We illustrate the application of temporal analytics in the analysis of gestures. Temporal analytics are important for teasing out the signal from the noise within sequences of hand movements.
- Incorporating these kinds of fine-grained multimodal data can prove transformative in the design of effective learning environments because of their potential for personalization and prediction.

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

[T]here will certainly be some occasions when bodily engagement can be an especially effective means for achieving some learning goals. [...] There will also be times in which the new technologies that work with the body will ultimately help to tell us something new and important about how and when we learn.

—V. R. Lee (2015)

1 INTRODUCTION

Interactive learning environments with body-centric technologies are gaining traction in educational research because they lie at the intersection of the design of embodied learning activities and multimodal learning analytics (e.g., Abrahamson, 2014; Black, Segal, Vitale, & Fadjo, 2012; Lee, 2015; Lindgren & Johnson-Glenberg, 2013; Worsley et al., 2016). Sensing technologies like the Kinect, the Leap Motion sensor, or automated visual tracking, allow students to interact with virtual objects on a computer screen via gestures or physical movement. For instance, in a technologically enhanced activity, students may act out the behaviours of planets and meteors by moving around their classroom to learn about Newtonian physics (Lindgren, 2015), or use their bodies to model how particles move in different states of matter to learn about the particulate nature of matter and the relationship between energy, motion, and state (Danish, Enyedy, Saleh, Lee, & Andrade, 2015). In other examples, students move their arms to control virtual objects on the computer screen to learn about proportionality (Abrahamson & Sánchez-García, 2016), geometry (Smith, King, & Hoyte, 2014), and mathematical proofs (Nathan et al., 2014). All these new STEM learning environments put physical movement at centre stage. In addition, they allow researchers to seamlessly capture large amounts of data about student movements, which opens up new opportunities for studying the role of embodied activity in learning. An ongoing goal of our research is to explore how we can use these data to refine both our designs, and our analyses of learning.

Sensing technologies can generate large amounts of fine-grained data, automatically captured from student movements, which researchers can use to analyze student activities. As more and more evidence accumulates to show how our cognition is grounded in the body, researchers are also increasingly interested in exploring how these fine-grained movement data might support inferences about how the body supports learning (or the lack thereof). For instance, Smith et al. (2014) used Kinect movement logs to study how two students with low and high learning gains created different embodied representations of geometrical angles with their arms. From the Kinect logs, the authors found the student with high learning gains was able to create a wider range of arm positions to represent the same types of angles, compared to the student with low learning gains. This suggests a clear link between the ability to represent those angles, and learning about them. However, Smith et al. (2014) did not develop a generalizable statistical model to show how students arrive at different movement solutions. Indeed, little is known about how to best use computer logs of these kinds of movements to generate predictive models of learning from student movements. Predicting learning from movement can be important for testing hypotheses about embodied learning. Furthermore, a predictive model has important consequences for the design of technologically enhanced learning environments. For instance, the statistical model can be

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

used as an assessment tool, or to tailor experiences for individual students depending on their performance. Yet, a predictive model first requires a measurement model of the physical movements — i.e., a theoretical link between the log data and what these data represent in the form of a latent variable.

Measuring physical movements — hand movements especially — is not a simple task. In particular, it requires the special consideration of how to measure and account for sequences of movements through time or *temporal dependencies*. To account for temporal dependencies in the data, special statistical models are required for relaxing the independence assumption. The independence assumption refers to the supposition that any two observations in a dataset are independent. That is, measuring observation at X_1 does not provide any information about observation at X_2 . But this is not true when modelling the movement of hands, as the position of a hand at time t is dependent in part on where that hand was at time $t-1$. As we show in this paper, these temporal dependencies take place at various levels of analysis. Specifically, a) when measuring hand movement direction, b) when modelling the sequence of movements throughout the activity, and c) when measuring the distance between elicited and enacted movement.

Our approach is to model hand movement data and time dependencies using two statistical tools, hidden Markov models (HMM) and optimal sequence matching. We use HMM to create a model of the relative movement of both hands (the combination of up, down, or static movement of each hand). In doing so, the HMM reduces the dimensionality of the data. This is important because our interest is in the relative movement of both hands as they simultaneously move, regardless of their absolute position. We are interested in this relational movement because it maps to the kinds of quantitative reasoning that the students are engaged in without being asked to address absolute locations in space. In addition to measuring relative position, we use an optimal matching (OM) algorithm to measure the distance (also known as similarity) between the computer-elicited and student-enacted movements. In using the OM algorithm, the temporal information within the movement sequences is accounted for because this algorithm includes information about shifts and state transitions, as will be explained below. The distance between elicited and enacted movements provides a measure of how well the student follows the automated movement elicited by the computer. Measuring this distance can be important when examining the relationship between how students coordinate their motion with the computer-based cues and their learning gains. For instance, do students who are better at coordinating their movements with the computer's also show better learning gains?

The aim of this paper is twofold: a) to present a use-case of two specific statistical methods for the modelling of temporal data in an embodied learning environment, and b) to discuss some possible validity threats if these temporal dependencies are not accounted for. In particular, we systematically analyze how ignoring temporal information affects the measurement of hand movement data. Specifically, we assess how, if ignored, the temporal dependencies might affect the goodness of fit of the statistical model. In addition, we examine how temporality affects the distance (or similarity) measure between elicited and

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

enacted movement. Thus, our research questions are as follows: *How can we use temporal analytics to model and visualize student hand movement data? What are some consequences of ignoring the temporal dependencies when modelling hand movement data? Are these statistical models of embodiment related to learning gains, and as a result are they helping us to explore the specific relationship between elicited gestures and learning outcomes?*

2 BACKGROUND

2.1 Embodied Learning

The embodied turn in cognitive science (Anderson, 2003; Barsalou, 2010; Wilson, 2002) has foregrounded the key role the body plays in human cognitive processes. Embodiment theories argue higher-order cognitive processes, including memory and symbolic thinking, are grounded in body-based perception and action within a physical environment (Abrahamson & Lindgren, 2014; Barsalou, 2008; Hutto, Kirchoff, & Abrahamson, 2015). Although there are many different approaches, embodied cognition theories range between two viewpoints. On the one hand, the strong view of embodied cognition argues that all cognition is situated and action-based. Concepts are tightly related to perceptual and motor schemas. For instance, according to the reflexive abstraction hypothesis, mental objects with abstract properties are internalized by coordinating various lower-level empirical abstractions which are built up by performing actions on physical or imagined objects (Abrahamson, Shayan, Bakker, & Van Der Schaaf, 2015; Piaget, 1952). On the other hand, the soft view of embodied cognition argues that some higher-order abstract schemas can interact with perceptual and motor schemas to ground their meaning. Some concepts may be built upon other concepts, which in turn are based on perceptual and motor information. For instance, according to the metaphorical mapping theory, knowledge domains are related to one another by cross-domain mappings, which occur when a target domain receives the inferential structure of a source domain (Anderson, 2003; Lakoff & Johnson, 1999). An extreme sociocultural version of this argument also notes that there is always a social role of embodiment, which leads to continuous change in the environment and thus feeds back into how individuals experience that space (Enyedy et al., 2017; Hall, Ma, & Nemirovsky, 2014; Ma, 2017).

While the specific mechanisms of embodied cognition are not yet clear, there is ample evidence that attention to gesture and movement in the design of learning environments can support learning. As a result, learning scientists are finding new ways to incorporate body-based movement within learning environments. For instance, Lindgren and Johnson-Glenberg (2013) propose six principles for the design of technologically enhanced learning environments — i.e., 1) ascribe benefits of body-based learning to everyone, 2) assert action–concept congruencies, 3) augmentation should augur well, 4) introduce opportunities for collaborative interaction, 5) pair lab studies with real-world implementations, and 6) re-envision assessment. These principles are aligned with a soft view of embodied cognition, where the body serves as a metaphor with which students can explore principles and relationships in math and science domains. Complementing this work, Abrahamson and Lindgren (2014), also propose three design

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

principles to guide the application of embodied cognition in the creation of learning environments — i.e., 1) activities should mobilize perceptual senses and kinesthetic coordination, 2) activities should have meaning and provide action–feedback loops, and 3) students will need guidance to become attuned to the hidden aspect of the environment. Abrahamson and Lindgren’s (2014) principles are aligned with a stronger view of embodied cognition, in which “math and science concepts are not abstract, conceptual mental entities, removed from the physical world. Rather they are deeply somatic, kinesthetic, and imagistic. Interactive tasks typical of embodied design thus steer learners to discover, refine, and practice physical action schemes that solve local problems but can then be signified as enacting the targeted content” (Abrahamson & Lindgren, 2014, p. 11). Enyedy and Danish (2015) take a different approach, aligned with a sociocultural view of learning and cognition, where the body is regarded as another semiotic resource. Enyedy and Danish (2015) argue that “the promise of embodied cognition for education lies not in the presence of these links but in the ways in which embodied cognition opens up new horizons for instructional design. Designing instruction to account for the body allows us to legitimize and blend together new modalities and new sets of intellectual resources for learning” (Enyedy & Danish, 2015, pp. 97–8). The approach we take in the design of an embodied simulation of population dynamics (ESPD) bridges these prior approaches by attending simultaneously to how the body provides resources to reason with, and how this is situated in a meaningful social context. In the next section, we briefly describe the approach we followed for the design of our ESPD.

2.2 A Study with the Embodied Simulation of Population Dynamics

As noted above, we see embodiments such as gesture providing both an individual and a social resource. Thus, our working hypothesis was that we can use elicited gestures to support the way students learn about quantitative patterns of complex systems, and as these gestures become an object to think with about the graphical patterns, learners will continue to find interactional value in using these gestures during later explanations and collaborations. Quantitative patterns of complex systems are nonlinear changes in the quantities of a system. These quantitative patterns are nonlinear because systems usually display cycles and delays, as well as variable rates of change. An example of a nonlinear system dynamic is the feedback loop between the size of a fox’s skulk (group) and the size of a rabbit’s colony due to their predator–prey interrelationship (Wilensky & Reisman, 2006).

Elicited gestures are hand movements cued by the learning system or the experimenter. An example of an elicited gesture is to ask a student to use their hands to depict the population levels in the system being studied. In the current system, this might mean asking a student to move her left hand down and her right hand up, simultaneously, while thinking about the inverse relationship between the size of the population of foxes versus rabbits. She would use her left hand to represent the rabbit population, and the downward movement would represent the shrinking of this population size as the foxes feed on the rabbits. She would use her right hand to represent the fox population, and the upward movement to represent its

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

growing size as more foxes live to reproduce because they feed on the rabbits. We believe that this kind of elicited gesture has the potential to transform how elementary students learn such interrelationships.

In our learning environment (Andrade, Danish, & Maltese, 2017), called the embodied simulation of population dynamics (ESPD), a student explores the quantitative patterns of complex systems in the context of predator–prey dynamics (e.g., foxes and rabbits). The ESPD cues the learner to represent the unstable equilibrium between foxes and rabbits via hand gestures (see Figure 1). Our intention is that by moving their hands in this way, the student will make deeper connections to how the two populations are related, both by connecting physically to the movement patterns, and by reflecting explicitly on these relationships, which we believe become more salient through this embodiment. The student’s goal is to match the bar graphs, which she controls using two balls of different colours, with the horizontal markers on the computer screen (see Figure 1 on the left). For instance, the student moves her right hand down to match the fox population marker because the horizontal marker is lower than the depicted fox population; her left hand moves up to match the rabbit population marker because the rabbit population is lower than the horizontal marker. Through the use of these “elicited gestures,” the ESPD makes the nonlinear quantitative patterns salient to the student, who will learn about them via embodied mechanisms.

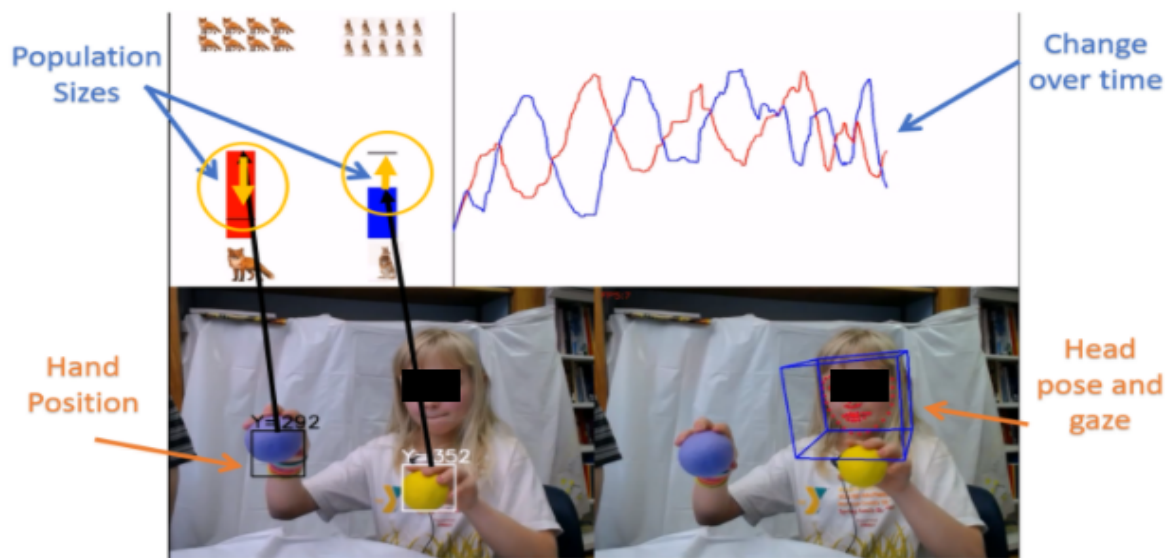


Figure 1. The ESPD learning environment. Top: What the student sees — a bar graph of population sizes on the left, and a line graph of the changes in population sizes over time on the right. Bottom: What the computer sees — colour blobs representing the hand positioning on the left, and head pose and gaze on the right. The student’s goal is to follow the movement of the horizontal markers by moving the population bars with the coloured balls. The horizontal markers automatically move to represent a nonlinear dynamic relationship between predator and prey population sizes.

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

We also hypothesize that we can explore ways in which physical action is connected with conceptual learning by creating a statistical model of the sequence of a student’s movements while interacting with the learning system. For instance, during our initial intervention, we started noticing some students used gestures to support their explanations of the relationship between predator–prey populations in a related scenario of fish and dolphins after interacting with the ESPD. In the following example, a student gestured with a simultaneous movement of her hands, left hand upward and right hand downward, and said, “When the dolphins go up, fish would go down because there’s so many dolphins...” (see Figure 2a). Then, she moved her left hand down and said, “But then the dolphins go down because there’s not enough fish...” (see Figure 2b) “and then the fish would go up because there’s less dolphins to eat them...” while moving her right hand up (see Figure 2c). As these spontaneous gestures look like the elicited gestures, we entertained the possibility that learning gains had something to do with the ability some students displayed in appropriating the elicited gestures. Because the elicited gestures were presented during the interaction with the computer, and the computer automatically logs all hand movements, we decided to build a model that would allow us to measure the difference between the sequence of elicited movements and the sequence of enacted movements.



(a) Simultaneously moves left hand up and right hand down

(b) Moves left hand down

(c) Moves right hand up

Figure 2. Student gestures while explaining at a later moment, similar to the elicited ones.

3 METHODS

3.1 Participants and Research Design

In an exploratory research study, fifteen third and fourth graders (F = 8, M = 7, Avg. Age = 9.13, SD Age = 0.8), from a mixed-age class at a private school in the mid-western US participated in a task-based cognitive interview. Students were individually interviewed and answered a pre-tutorial questionnaire, then interacted with the ESPD, and then answered a post-tutorial questionnaire. Interviews were videotaped and took 30 minutes on average. The pre- and post-tutorial questions were adapted from Hokayem, Ma, and Jin (2015). To score the answers, we used an adaptation of Hokayem et al.’s coding

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

scheme, which includes seven complexity levels of reasoning about predation dynamics. The interaction with the ESPD consisted of nine tasks, which were divided in three phases (briefing, training, and demonstration). The first three tasks familiarize the student with the tracking system and display. For the training phase (tasks 4–6), the student follows the automatic movement of the bars and is told to notice the patterns in the line graph that her hand movements create. In the demonstration phase (tasks 7–9), the student only sees the line graph display and is challenged to demonstrate the elicited movement by creating the appropriate changes in one population with respect to the other. To answer the research question of whether students show learning gains about quantitative patterns of predation ecosystems, a Wilcoxon Signed-Rank test for repeated-measures was used to compare the changes from pre-test to post-test scores. Furthermore, we wanted to see if there was a connection between learning and physical movement. We conceived that it might be possible to find a significant correlation between student hand movements and learning gains. Specifically, we hypothesized better learning gains should be associated with an improvement in student-enacted movements. That is, we hypothesized that students who increased their similarity with the computer-elicited movement would also have higher learning gains. This hypothesis was tested using the log data comparing the change in similarity from task 6 to task 9 and a Spearman's rank correlation test.

3.2 Computer-Elicited and Student-Enacted Data

The embodied system, the ESPD, is an instructional design that uses digital interaction via sensing technologies to help students make connections between physical movement and quantitative understanding of complex systems. These connections take place in the form of embodied representations of quantitative patterns facilitated via elicited gestures. The ESPD system has three components: a) a tracking system that follows two coloured balls, b) a display with two horizontal markers that cue nonlinear movement and depict where a student's hands are relative to the marker, and c) a line graph tracking the movement of the bar graphs over time (see Figure 1 above). The computer-vision algorithm captures the vertical position (in pixels) of the colour blob centres in each frame. It is assumed the position of the hands controlling the bar graphs is in response to the movement elicited by the horizontal markers. Figures 3 and 4 show the elicited and enacted data, respectively. Figure 3 shows the computer-elicited data in matrix and graphic form, and Figure 4 shows an example of the empirical movement tracked by the computer for Student 1.

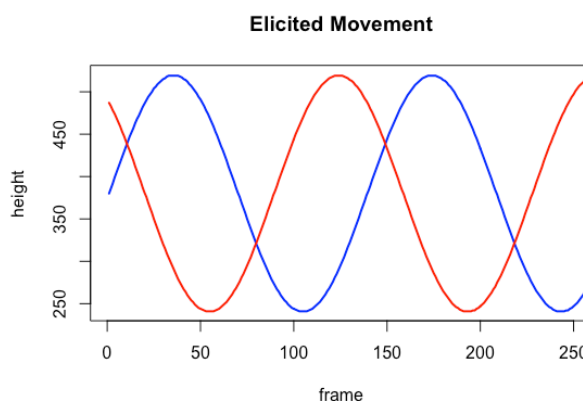
In what follows, we systematically analyze how ignoring temporal information affects the measurement of the hand movements, as recorded in the ESPD log data. We first analyze this issue at the level of the measurement of movement combinations. That is, we compare the goodness of fit between the empirical data and two latent models. We compare an HMM, which accounts for temporality, versus a latent class analysis (LCA) which does not account for temporality. The effect of accounting for the temporal dependency is assessed by comparing the fit indices for a similar number of latent states between these two models. Then, we analyze the similarity measures between elicited and enacted movement. That is,

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

we compare the distance measures from the OM algorithm versus the Hamming algorithm and a frequencies-only algorithm, based on the count of matching attributes. These two latter algorithms ignore the temporal information at different degrees. We analyze how results vary from the use of these various algorithms.

Frame	Rabbits (Blue)	Foxes (Red)
1	380	487
2	386	483
3	392	478
4	399	474
5	405	469
6	411	464
7	417	459
...

(a)

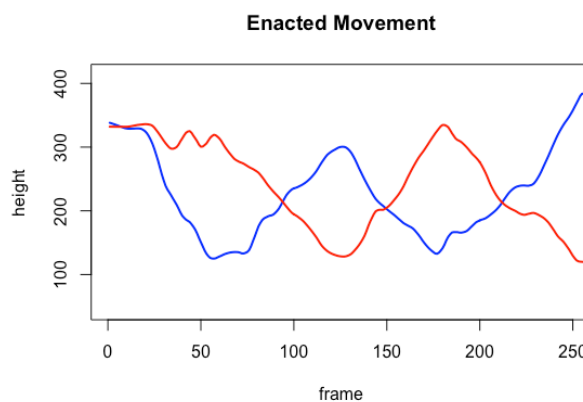


(b)

Figure 3. Computer-elicited movement data in both table and graphical form.

Frame	Left Hand (Blue)	Right Hand (Red)
1	337	332
2	337	333
3	337	333
4	337	334
5	335	336
6	334	336
7	331	335
...

(c)



(d)

Figure 4. Student-enacted movement data in both table and graphical form.

Having laid out the general purpose of the analysis, next we briefly explain how the first temporal analysis is used for computing the direction and qualitative motion of the hands from the ESPD logs (see Figures 3 and 4).

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

3.3 Preprocessing Data and Analysis of Hand Direction and Qualitative Motion

The most fine-grained logs are hand position values as the computer measures (in pixels) the position of the hands in each video frame, at a frequency of 7 frames per second. However, position values without temporality do not carry meaningful information about student behaviour. To represent behaviour, direction values are computed. Direction values capture motion information by considering the direction and magnitude of the movement. Direction values are calculated, first, by subtracting each hand’s vertical position at time t from position at time $t+1$. Second, by focusing only on the qualitative motion in direction categories (“up,” “down,” “static”), the dimensionality in the data is reduced. We focus on qualitative motion instead of absolute motion to simplify the representation of hand movement. This is because the overlap between absolute position data (in pixels) is not as important as the overlap in relative movement data (e.g., up-up, up-down, etc.). In computing the qualitative motion data, additional considerations include accounting for student idiosyncrasies, like individual differences in the amplitude of their movements, and noise. Thus, the magnitude values are first normalized (dividing each absolute motion value by the largest value), and a threshold filter is applied to avoid detecting small movements as meaningful movements. The selected threshold was the student’s semi-interquartile range. An example of the qualitative motion calculation is shown in Table 1 and Figure 5.

Table 1. Calculation Example of Hand Motion and Direction Values using a ± 1 threshold. The hand position is recorded in pixels, and two consecutive points are subtracted. Positive magnitude values imply the hand is going down because pixel values increase downwards in a video image.

Frame	Left	Right	magnitude_left	magnitude_right	direction_left	direction_right
1	337	332	—	—	—	—
2	337	333	0	1	static	static
3	337	333	0	0	static	static
4	337	334	0	1	static	static
5	335	336	-2	2	up	down
6	334	336	-1	0	static	static
7	331	335	-3	-1	up	static
8	331	335	0	0	static	static
...

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

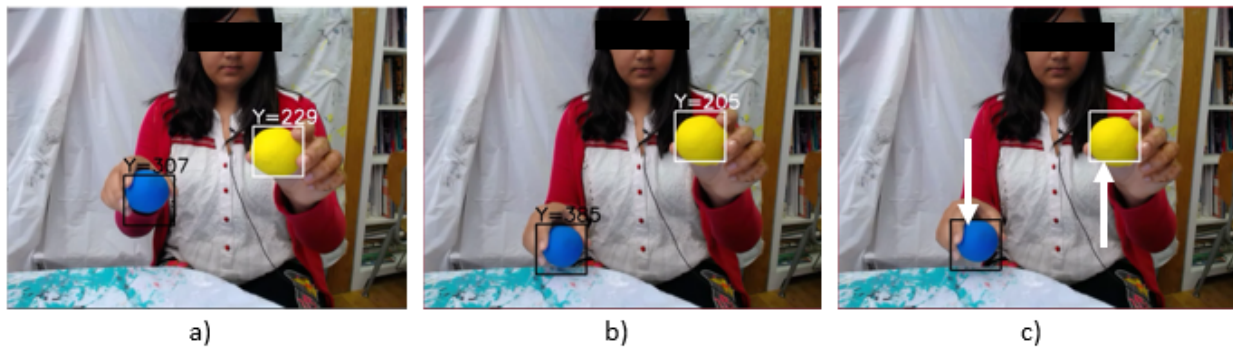


Figure 5. Calculating direction vectors from hand position data: c) is the sign of the difference between a) and b).

Direction vectors only indicate the direction of movement between time t and $t+1$. If the interest is in understanding the pattern of the bimanual movement, a statistical model can be used to represent the bimanual motion coordination at every two frames for the duration of the activity. For instance, the student might be trying to coordinate a simultaneous movement of one hand going up and the other hand going down. Or perhaps the student might be trying to coordinate the movement of one hand *after* she starts the movement of the other hand. When using a model-based statistical approach, latent states can be inferred from the patterns of fine-grained log data. This data reduction would go from two data streams of (categorical) direction vectors, to a sequence of motion states. However, ignoring the temporal dependency between observations in this step might produce too many latent states or too many state transitions, as will be shown later. To account for the autocorrelation between observation at time $t+1$ and observation at time t , we use an HMM to model the sequence of movements throughout the learning activity.

3.4 Using a Hidden Markov Model to describe the Hand Movement Sequence

An HMM is also referred to as a dependent finite mixture model (Gollery, 2008; MacDonald & Zucchini, 1997; Visser & Speekenbrink, 2010; Zucchini & MacDonald, 2009). HMMs have been used in various applications like speech recognition, EEG analysis, psychology, economics, and genetics. The purpose of this statistical model is to infer an unordered set of latent states that explain the correlation between a set of observed variables given an estimated transition rate between latent states. The HMM can fit univariate or multivariate data for continuous or discrete variables. The fundamental assumption is, at any point in time, the observations are distributed as mixtures given an r number of latent/hidden states, and time-dependencies between observations are due to time-dependencies between the hidden states following a first-order Markov process (Visser & Speekenbrink, 2010). A first-order Markov process assumes that, given a sequence of a discrete random variable, the occurrence at time $t+1$ is conditioned upon the most recent value of the random variable at time t (Zucchini & MacDonald, 2009). This property is a relaxation of the independence assumption, and can be displayed as a direct graph where any future

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

observation is dependent only on the present observation (see Figure 6a). The conditional probabilities associated with a Markov process are called transition probabilities, and convey the temporal association between the distinct hidden states. All the transitions from state i to state j create a matrix of transition probabilities. When this relaxation of independence is incorporated in a model for the analysis of finite mixture distributions, the model is thus called a hidden Markov model. Thus, the HMM has two parts, the first part represents the C latent parameter process of a Markov walk, and the second part represents the X state-dependent process such that the distribution of the observations depends only on the current state (see Figure 6b). In contrast, independent finite mixture models assume the X observations are independent when conditioned upon the latent states (see Figure 6c). One can think of an independent finite mixture model as a factor analysis with categorical observed variables instead of continuous variables. A factor is a latent state, which is responsible for a distinct combination of levels in the observed variables. An HMM would impose a Markov process on top of the factor analysis.

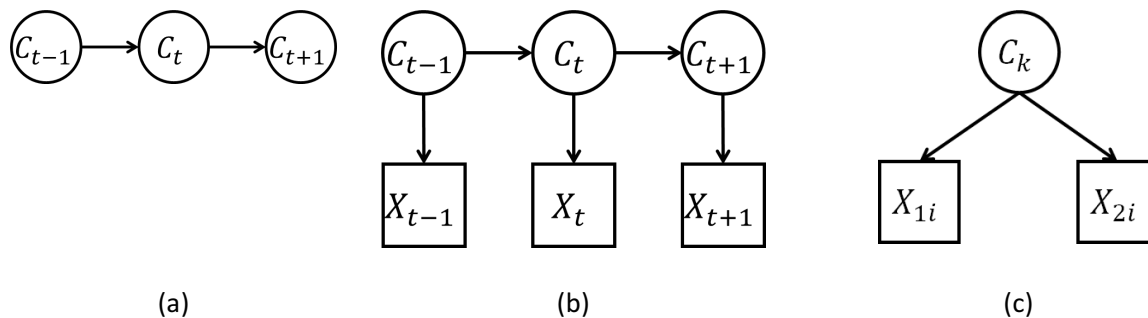


Figure 6. (a) A Markov chain, (b) a dependent finite mixture model also known as a hidden Markov model, and (c) an independent finite mixture model.

The HMM inputs a sequence of observations and predicts a latent state sequence of length $N-1$, where N is the number of time points in the data frame. The meaning of the latent states is evaluated by examining the composition of the mixture of observed variables. An example of an input data frame can be seen in the last two columns of Table 1 (*direction_left* and *direction_right* columns). An example of a hypothetical predicted six-state sequence, fit to the computer-elicited movement data, is shown in Figure 7. Figure 7 shows the plotted position of the hands over time as blue and red lines. The vertical lines show where there is a state change in the hands' relative movement. The periods between vertical lines correspond to latent states. The state sequence can also be plotted as a sequence of colours where each colour represents a distinct state (see Figure 8).

The meaning of these latent states can be inferred from the combination of movements. For instance, on the left side of Figure 7, between times 0 and around 15, the right hand (red) is moving downwards while the left hand (blue) is within the threshold of being static. This period corresponds to State 6. Between times 15 and 50, the right hand (red) keeps moving downwards while the left hand (blue) starts to move upwards. This period corresponds to State 3.

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

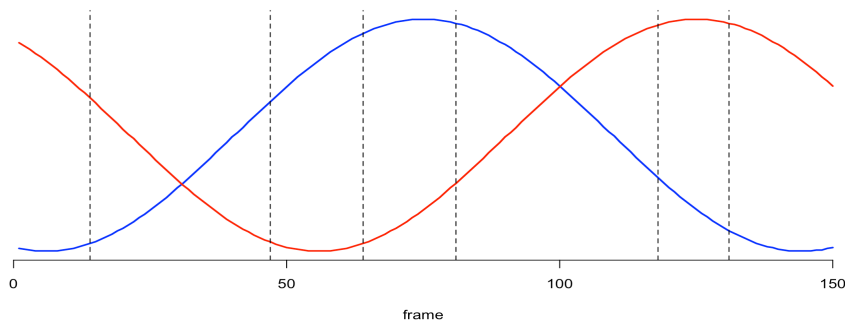


Figure 7. An example of a sequence of latent states. The trajectory of hand movements over time (right hand in red and left hand in blue). The vertical lines show the division of the hands’ trajectory by latent state.

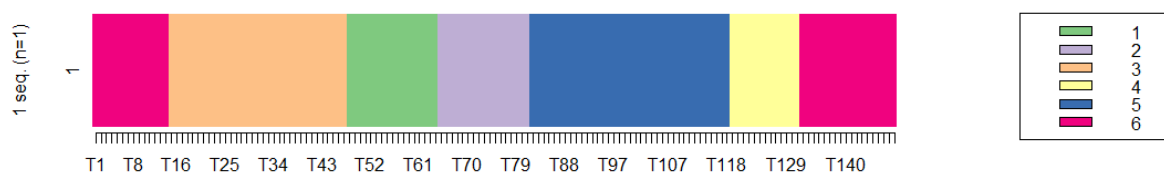


Figure 8. A colour-coded representation of the state sequence for the bimanual motor coordination example.

The meaning can also be inferred from the matrix of probability distributions of response categories (see Table 2). For instance, Table 2 shows the corresponding distribution of categories for the 6-state model. In State 1, the left hand has a probability of 100% to move down as the right hand remains static with a probability of 92%. Note the probabilities are not always 100%.

Table 2. Probability distribution of hand movements given latent states

	Left Hand			Right Hand		
	Down	Static	Up	Down	Static	Up
State 1: down – static	1.00	0.00	0.00	0.08	0.92	0.00
State 2: static – down	0.16	0.84	0.00	1.00	0.00	0.00
State 3: down – up	1.00	0.00	0.00	0.00	0.00	1.00
State 4: up – static	0.00	0.00	1.00	0.00	1.00	0.00
State 5: up – down	0.00	0.00	1.00	0.97	0.03	0.00
State 6: static – up	0.03	0.78	0.19	0.00	0.00	1.00

The HMM produces a transition probability matrix. An example of a matrix of transition probabilities between latent states is plotted in Figure 9. For instance, Figure 9 shows State 2 only occurs after State 5 and no other state. In a similar vein, State 4 only occurs after State 6, 5 after 4, 3 after 1, and 1 after 2.

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

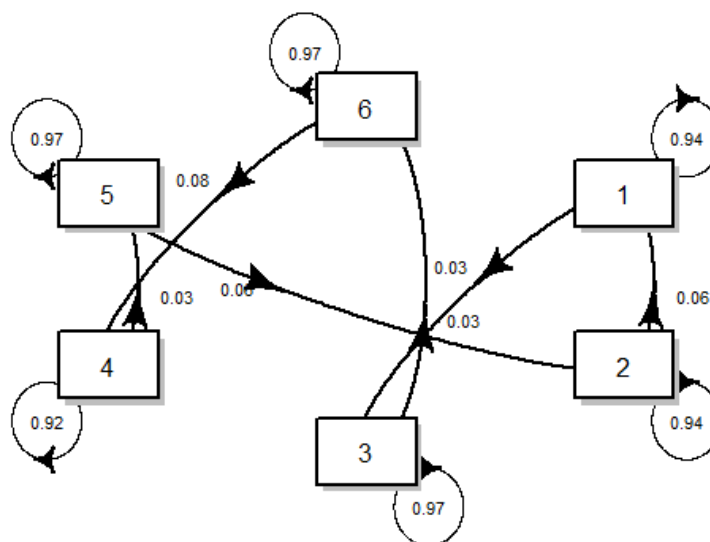


Figure 9. Transition probabilities between latent states of the computer-elicited data.

As the number of states is not defined *a priori*, several models can be fit to the data. Therefore, to select the best number of states for the empirical data, various fit indices can be employed. These indices evaluate how well the expected cell counts under a given model replicate the observed cell counts (Hagenaars & McCutcheon, 2002). The Akaike information criteria (AIC) and the Bayesian information criteria (BIC) account for the increment in the number of parameter estimates and therefore penalize the increment in the number of latent states. By using the AIC and the BIC, a balanced number of states fitting the data well can be found. The lower the value of these indices, the better the model fits to the data. Thus, to select an optimum number of latent states, a series of models with increasing numbers of latent states are fit to the data. Then, the model with the lowest AIC or BIC is selected. As we measure the entire sequence of movements throughout each task, however, it is expected that nine latent states will be the best fit to the data. This is because there are nine total possible movement combinations for the two hand direction values (see Table 3). Therefore, we would expect that the AIC and the BIC for the whole dataset would point to a 9-state model.

Table 3. Nine possible combinations of hand direction values

Hand	Direction								
Right	Up	Up	Up	Down	Down	Down	Static	Static	Static
Left	Up	Down	Static	Up	Down	Static	Up	Down	Static

In summary, an analysis of ESPD log data requires us to model the distinct combinations of relative hand movements as they unfold over time. Therefore, the HMM helps us translate the qualitative motion data into a sequence of latent states (i.e., a representation of the relative bimanual movement over time). In the next section, we entertain the possibility of using a different statistical model, one which does not

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

account for temporal dependencies. We do this to explore the validity of the HMM in modelling the hand movement data.

3.5 Ignoring Temporal Dependencies in the Data: HMM versus LCA

Let us pretend we were to ignore the Markov process on the latent states. Note we are not suggesting one should use a different model instead of an HMM. The following is, thus, a hypothetical example only for illustration and not for research purposes. If the independence assumption is not relaxed, the model reduces to an independent finite mixture model, as stated above. An independent finite mixture model for categorical variables is also known as LCA (Agresti, 2014; Collins & Lanza, 2013; Hagenaars & McCutcheon, 2002; Linzer & Lewis, 2011; Vermunt & Magidson, 2004; Vermunt, Tran, & Magidson, 2008). LCA has been used in various social science applications including econometrics, behavioural psychology, social psychology, biometrics, and consumer behaviour, among others. The purpose of this model is to infer an unordered latent categorical variable that explains the correlation between a set of observed categorical variables (Linzer & Lewis, 2011). The fundamental assumption is the instantiation of the observed categorical variables is conditioned upon the state of the latent categorical variable. Thus, the finite set of latent states explain the distinct mixtures of frequencies in the cross-classification table of observed variables. The model is called *independent* because it is assumed that the distinct latent states are independent of each other (Hagenaars & McCutcheon, 2002). This means that after the observations are conditioned upon the latent class, the observations are also independent, a property called local independence (Linzer & Lewis, 2011). In the results in section 4, we compare the outcome of the LCA to the HMM and show how the LCA does not provide a good account of hand data, precisely because it does not account for temporal dependencies — e.g., transition rates between states. Before moving on to the results section, however, we briefly explain our approach to creating a similarity measure between the elicited and enacted movements.

3.6 Optimal Sequence Matching

Because the student movement is a response to the computer-elicited movement, a measure of the similarity between the student-enacted movement and the computer-elicited movement can serve as a proxy of the student's ability to respond to the elicited movement. This similarity can be conceived of as the student's sensorimotor coordination. We use the term sensorimotor coordination because the action combines the perceptual aspects of noticing the position of the bars with respect to the horizontal markers on the visual display, and the motor aspects of the movement of both hands as they respond to the perceptual information. We propose making use of the OM algorithm to measure the similarity between student-enacted and computer-elicited movement because this algorithm can account for time dependencies in the data.

The OM algorithm (Abbott & Tsay, 2000; Gabadinho, Ritschard, Mueller, & Studer, 2011) is a dissimilarity measure, part of the family of measures known as *edit* distances, based on the minimal cost of

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

transforming one sequence into the other (Gabadinho et al., 2011). The larger the cost of transforming one sequence into the other, the more dissimilar two sequences are. Conversely, a cost of 0 implies two identical sequences. The OM algorithm inputs a matrix of sequences, a cost value for the insertion/deletion of an element (or *indel* costs), and a matrix of substitution costs (e.g., the cost for swapping State A for State B). Depending on the relative costs between insertion/deletion and substitution, the minimum value can favour either insertions/deletions or substitutions. For instance, *indel* costs shift the position of all elements to the right of the sequence, allowing for position-dependent costs. For instance, take the state sequence $S1 = [A, B, C]$, and find how many insertions or deletions are required to transform it into $S2 = [B, C, A]$. It can be seen that one could take out State A at the beginning of $S1 = [B, C]$, and then add State A as its last element, indeed making it identical to $S2 = S1 = [B, C, A]$. If each transformation is worth 1, this operation would have a cost of $d = 2$. Note this *indel* process can be regarded as a left or right *shift* of all elements to the right of the deleted/inserted element. Yet, if the costs of deleting and inserting elements are sufficiently high, then the minimal cost will be dominated by substitution costs. For instance, if instead of deleting and inserting State A from $S1$, one decides to substitute each of its elements (i.e., three substitutions, one per element), one can transform $S1$ to be identical to $S2$. Depending on how much each substitution costs, this transaction can cost less or more than inserting and deleting one element. If substitution costs are 1, the total cost of this procedure is $d = 3$ (as three substitutions are required). However, if substitution costs are 0.5, the final substitution cost $d = 1.5$, making it cheaper than the computed *indel* cost. Therefore, a careful balance between *indel* and substitution costs should be determined to prevent either one from dominating the calculation. However, setting substitution and *indel* costs is not an easy task, and can be a controversial feature of the OM algorithm. Researchers have tried a variety of approaches to set substitution and *indel* costs. For instance, researchers have proposed using *indel* costs less than $cl / 2 * \max(sm)$, where cl is the common sequence length and $\max(sm)$ the highest substitution cost (Gabadinho et al., 2011). For substitution costs, researchers have tried several approaches, such as a linear order of some sort, some known linear property of the states, or theoretically generated costs (Abbott & Tsay, 2000). To incorporate information about the time dependencies among latent states in a sequence, the substitution-cost matrix can be specified to be equal to the estimated transition rates between latent states. In this way, the OM algorithm can include the information contained in the transition matrix from the HMM. Because of these various possible approaches, methodologists recommend a systematic analysis of how different cost schemes alter the results (Abbott & Tsay, 2000).

In the following dummy example, we show how the *indel* and substitution costs affect the distance values. Suppose that one has three two-state sequences of length 3 as shown in Figure 10, and assign an insertion or deletion cost of 1 — the cost of inserting or deleting a state in the sequence. For instance, deleting state A at time 3 in sequence 3 would cost 1. Then, inserting state B at time 3 in sequence 3 would also cost 1. Thus, the *indel* cost of transforming sequence 3 into sequence 2 is 2. Furthermore, suppose any substitution cost is also 1 — the cost of substituting a state in the sequence by any other state. For instance, substituting state A for state B at time 3 in sequence 3 would have a cost of 1. Thus, the

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

substitution cost of transforming sequence 3 into sequence 2 is 1. Therefore, given the above *indel* and substitution costs, the minimal cost of transforming sequence 3 into sequence 2 is 1 — because substituting is cheaper than deleting and inserting. In the same vein, the substitution cost of transforming sequence 1 into sequence 2 is 3 — because all three states need substitution. However, the *indel* cost of transforming sequence 1 into sequence 2 is 2 — because state A at time 1 needs to be deleted, shifting the sequence to the left, and then state B at time 3 is inserted. Therefore, the minimal cost of transforming sequence 1 into sequence 2 is 2. Finally, the cost of transforming sequence 2 into sequence 3 is 1 — because it requires only one substitution.

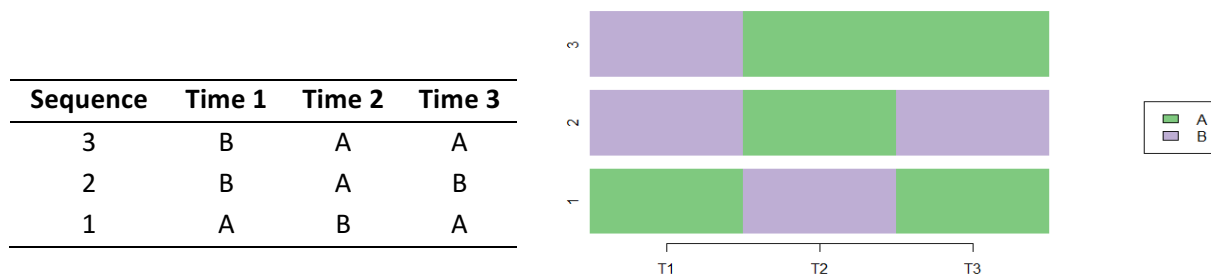


Figure 10. Three dummy sequences to illustrate the OM algorithm.

If *indel* costs are too high compared to substitution costs, most or all distance values will be determined by substitution. This is not an ideal situation when shifts are of importance. Shifts occur when a block of states at an initial position are very similar to the target sequence if moved to a different position. This circumstance occurs when two series (two temporal patterns) are similar but differ only by a few positions (when pattern A is equal to pattern B but shifted to the left or to right). In our study of sensorimotor coordination, this is of great importance because there may be a case where a student’s enacted movement is like the computer-elicited movement but is delayed by a few seconds. Thus, we should not expect the two sequences — elicited versus enacted — to be perfectly aligned, but instead we need to allow some latitude for the enacted motion pattern to catch up with the elicited motion pattern. We assume this latency is a function of a student’s reaction speed to the changes in the perceptual input from the screen.

Furthermore, as mentioned above, the OM algorithm allows a case in which substitution costs depend on the transition rates between states. Here, not all the substitution costs are 1, but depend on the transition probability from state *i* to state *j*. For instance, consider the three 3-state sequences in Table 4. If state A at time 3 in sequence 3 was to be substituted for state C, the cost would be 1.75. Compare that cost to the cost of substituting state B at time 3 in sequence 2 to state C — it would only cost 0.6. Thus, both substitution and *indel* costs can account for temporal information contained in the sequences and in the dependencies of transitioning from one state to the other. When applied to the sequence of gesture data, the OM algorithm will measure the steps required to transform a student-enacted motion sequence into the computer-enacted motion sequence.

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

Table 4. Three hypothetical three-state sequences and corresponding transition matrix

<i>Three-State Dummy Sequences</i>						<i>Transition Matrix</i>			
Sequence	Time 1	Time 2	Time 3	Time 4	Time 5		To A	To B	To C
3	B	A	A	C	B	From A	0.00	0.90	1.75
2	B	C	B	A	B	From B	0.90	0.00	0.60
1	A	B	C	B	A	From C	1.75	0.60	0.00

In the next section, we entertain the idea of using other similarity measures that account for temporal dependency at a lesser degree or not at all. We do this to study the validity of the OM algorithm to capture the similarities between the elicited and enacted sequence of movements in the ESPD log data. As is shown in the results, section 4, these other algorithms do not maintain the expected order of similarities suggested by a qualitative appraisal of various student sequences.

3.7 OM versus other Similarity Measures

One can also resort to using other similarity measures for quantifying the distance between computer-elicited and student-enacted movements, but these other measures might fail to account for temporality at various degrees. We compare the OM distances to a set of distances based on the count of matching attributes. These distances are proximity measures because they compare matching positions between two given sequences (Gabadinho et al., 2011). First, we compare the OM distance to the Hamming distance (Hamming, 1950), which measures the number of positions at which two sequences of equal length differ. Second, we compare the OM distance to a distance measure based on the frequency of attributes; that is, one that does not include any temporal information. This index is based on the Euclidean distance applied to the frequency of states. In summary, we compare the OM algorithm, which accounts for temporal information, to one method that does not allow for shifts (the Hamming distance) and another that does not use any temporal information at all (the Euclidean distance).

4 RESULTS

4.1 HMM

Using the *depmix6* R package (Visser & Speekenbrink, 2010), seven r -state HMM models, where $r = 4 \dots 10$, were fitted to a dataset with the computer-elicited data and data from the fifteen students. A 9-state HMM model best fits the data, according to the AIC (see Table 5). Note the BIC index suggests an 8-state model because it is a more conservative index than the AIC. However, after careful examination of the patterns, the response distribution for an 8-state model is not as clean as a 9-state model, as we would expect from the nine possible hand motion combinations. The response distribution for the 9-state model can be seen in Table 6.

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

Table 5. Fit Indices for *r*-state HMM models with both the student-enacted and computer-elicited data

<i>r</i>	AIC	BIC
4	1258.762	1394.86
5	1156.478	1349.648
6	1010.409	1269.433
7	929.0487	1262.707
8	771.8023	1188.875
9	700.3675	1209.635
10	742.2089	1352.452

Four predicted motion sequences for 150-frame excerpts (approximately 20 seconds) are plotted in Figure 11. Figures 11a and 11b compare the computer-elicited versus student 1 motion data. Two facts are apparent in this comparison. First, the states in the student-enacted sequence shift less smoothly than the computer-elicited. This is because the student 1’s movements are not as smooth as the computer’s, probably due to the student’s attempts at correcting her movement while trying to follow the elicited movement of the bar graphs. Second, the pattern of movements in the student-enacted sequence is shifted to the right. As anticipated, student 1’s movements lag behind the computer’s by about 20 frames (approximately 2 seconds), probably due to a lag in the student’s sensorimotor coordination. The graphs for students 2 and 3, however, are not as smooth as those for student 1 (see Figures 11c and 11d). These students seem to have struggled more than student 1 to shadow the computer movements (at least in this short window sequence).

Table 6. Probability distribution of hand movement given latent states

	Left Hand			Right Hand		
	<i>Down</i>	<i>Static</i>	<i>Up</i>	<i>Down</i>	<i>Static</i>	<i>Up</i>
State 1: static – static	0.00	1.00	0.00	0.00	1.00	0.00
State 2: down – down	0.97	0.03	0.00	1.00	0.00	0.00
State 3: static – up	0.02	0.98	0.00	0.00	0.00	1.00
State 4: up – static	0.00	0.00	1.00	0.00	1.00	0.00
State 5: up – down	0.00	0.00	1.00	0.99	0.01	0.00
State 6: down – static	1.00	0.00	0.00	0.02	0.98	0.00
State 7: down – up	1.00	0.00	0.00	0.00	0.00	1.00
State 8: up – up	0.00	0.07	0.93	0.00	0.00	1.00
State 9: static – down	0.00	1.00	0.00	1.00	0.00	0.00

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

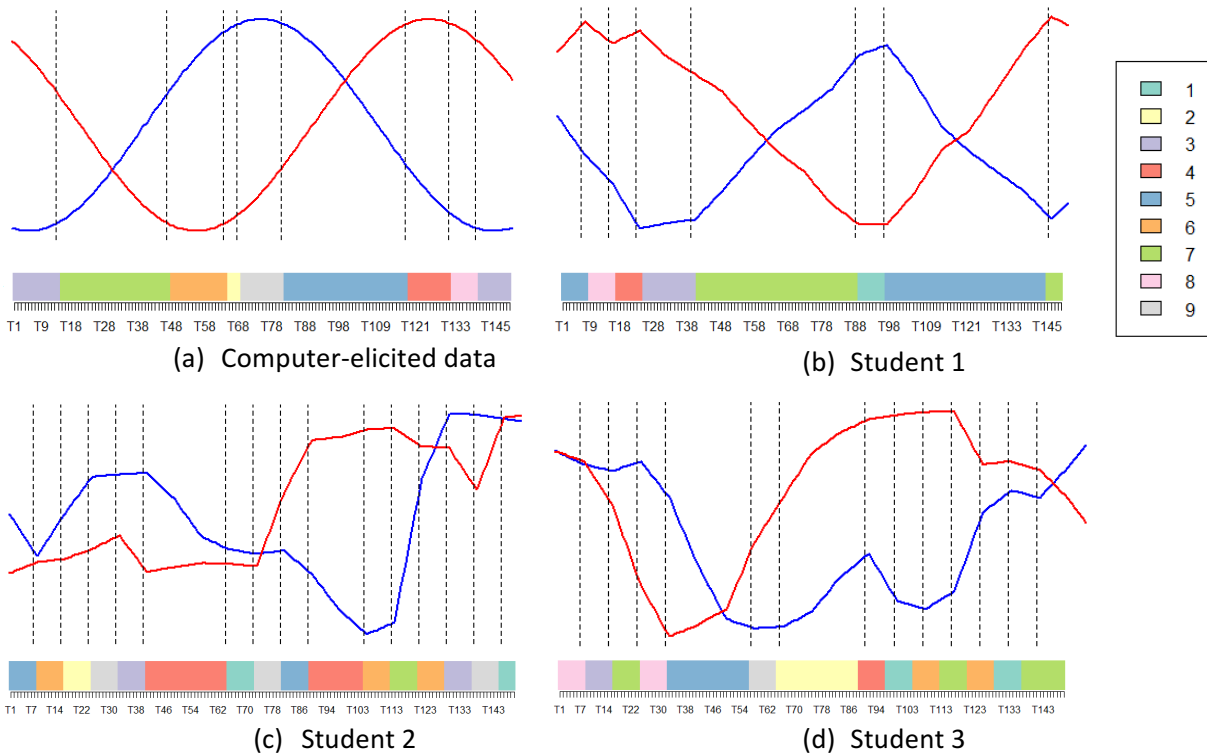


Figure 11. Graphical representation of the trajectories and latent states for the computer-elicited motion sequence and three student-enacted motion sequences.

In summary, the HMM captures student movements as a sequence of bimanual motion patterns, and there is a motion sequence per each student for tasks 6 and 9. We can then compare the similarity of these sequences to the computer-elicited sequence and study how close the student movements were to the ones from the computer and whether there are changes from task 6 to task 9. But before looking at this similarity measure, let us review the results if we were to ignore the transition probability between latent states. That is, examine the results from an LCA compared to the HMM.

4.2 HMM vs LCA

Using the *poLCA* R package (Linzer & Lewis, 2011), we ran seven *c*-state LCA models, where $c = 4 \dots 10$, with the same dataset as the HMM models above. Goodness of fit values are shown in Table 7. Two facts stand out from this table. First, the AIC and BIC values are much higher than the same fit indices for the HMM. Even the largest AIC value of the HMM model = 1394, is lower than the best AIC value from the LCA model = 2500. Second, the values keep increasing, which implies a worse fit, as more latent states are added to the model. Thus, as more latent states are added, the model does a worse job accounting for the observations.

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

Table 7. Fit Indices for *c*-state LCM models with the same data as the HMM

<i>c</i>	AIC	BIC
4	2500.85	2584.265
5	2510.85	2616.216
6	2520.85	2648.167
7	2530.85	2680.118
8	2540.85	2712.070
9	2550.85	2744.021

If we were to lay out the predicted sequence from a 6-state LCA, it would look like Figure 12b. It is apparent that states appear at strange moments, scattered throughout the sequence. This happens because latent states are not influenced by each previous state, and there is no continuity in the sequence. Put another way, these states are not temporarily stable. Furthermore, the latent states appear without a particular order throughout the sequence. Compare this sequence to the one constructed from the HMM (see Figure 12a), and it is easy to see that there is a remarkable difference in terms of the predicted states consistency.

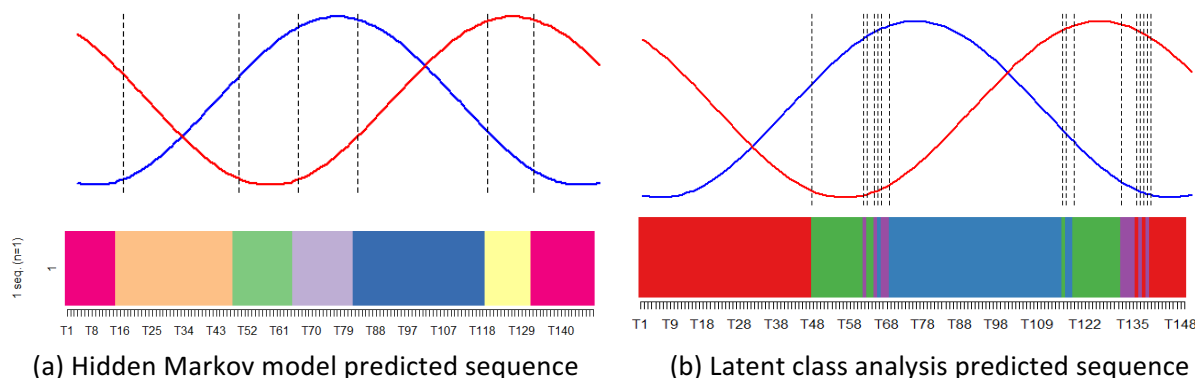


Figure 12. Graphical representation of the computer-elicited motion sequence comparing the use of HMM and LCA to predict the latent states.

Having explored the validity of the HMM approach, in comparison to the LCA model, we move on to examine the results of the similarity values between elicited and enacted movement using the OM algorithm.

4.3 OM Values

Using the *TramineR* R package (Gabadinho et al., 2011) we calculated the cost of transforming the student-enacted sequences into the computer-elicited sequence. For instance, student 1’s distance value was calculated to be 128, given an *indel* cost of 1 and a substitution cost matrix with the transition probabilities matrix (see Table 8). This means about 60 operations (between insertions, deletions, and substitutions) are required to transform this student-enacted sequence into the computer-elicited motion sequence.

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

We also examined how student 1’s distance value compares to the other two student sequences that do not appear as closely aligned to the computer-elicited movement. As anticipated, the calculated distances for student 2 and student 3 are 189 and 200, respectively. Compared to student 1’s sequence, these sequences are 48% and 56% less similar to the computer-elicited movement.

Table 8. Substitution cost matrix with the transition matrix of the 9-state HMM

	To 1	To 2	To 3	To 4	To 5	To 6	To 7	To 8	To 9
From 1	0.00	2.00	2.00	1.97	1.97	1.95	1.96	2.00	1.95
From 2	2.00	0.00	2.00	1.97	2.00	1.96	2.00	2.00	1.92
From 3	2.00	2.00	0.00	1.97	2.00	1.98	1.95	1.94	1.96
From 4	1.97	1.97	1.97	0.00	1.98	1.99	2.00	1.95	2.00
From 5	1.97	2.00	2.00	1.98	0.00	1.99	1.99	1.96	1.95
From 6	1.95	1.96	1.98	1.99	1.99	0.00	1.94	2.00	2.00
From 7	1.96	2.00	1.95	2.00	1.99	1.94	0.00	1.99	2.00
From 8	2.00	2.00	1.94	1.95	1.96	2.00	1.99	0.00	2.00
From 9	1.95	1.92	1.96	2.00	1.95	2.00	2.00	2.00	0.00

But, is all this trouble of setting up *indel* and substitution costs to input into the OM algorithm worth it? Why should we care? What if we were to ignore the possibility of shifts in pattern chunks or even the order of the states in these sequences? Can we obtain similar results from a more parsimonious approach? Having introduced the results from the OM algorithm, in the next section we take a look at the results for the similarity measures from other algorithms that account for temporal dependencies at different degrees. As we elaborate below, these other algorithms do not maintain the expected order in the similarities among students 1 through 3.

4.4 OM versus other Algorithms

Results for the three-student example are shown in Table 9. It is apparent the Hamming and Euclidean methods invert the sequences order for the three students. For instance, sequence 2 is less similar than sequence 3 for the Hamming method. Furthermore, the Hamming method shows sequence 1 is almost as dissimilar as sequences 2 and 3, when this is not really the case. For the Euclidean method, the differences are more contrasting. For instance, sequence 1 is no longer the most similar, and thus, this order does not agree with what is expected from a qualitative examination of the sequence plots (see Figure 11 above).

Table 9. Comparison of distance values from different methods. Distance values have been normalized for ease of comparison

Method	OM	HAM	EUC
Sequence 1	0.64	0.90	0.78
Sequence 2	0.95	1.00	1.00
Sequence 3	1.00	0.98	0.73

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

Having stated the validity of the OM algorithm to measure the similarity between elicited and enacted movements, in the next section we introduce the results from our exploratory study. In particular, we are interested in studying whether students were able to better coordinate their movements by task 9, in comparison to task 6, and whether improving their enacted movements is correlated to better learning gains.

4.5 Results from the Exploratory Study

Findings from our exploratory research support our two hypotheses. First, feedback loop reasoning significantly increased from pre- to post-tutorial scores. This increase in feedback reasoning scores (range = [0, 4]; median = 1) is significant, $Z(15) = 2.779$, $p\text{-value} = .008$, with a large effect size, $r = .718$. This effect size was calculated dividing the Z value by the square root of N , as suggested by Pallant (2007), and is equivalent to Spearman's rank correlation coefficient. Second, using the OM algorithm with *indel* costs set to 1 and substitution costs based on the transition matrix, OM scores were calculated for task 6 (range = [834.9, 1234]; median = 999.4) and task 9 (range = [807, 1230]; median = 1005.9). Although these results might suggest there is not a reduction in the average distance between elicited and enacted movement across students, we did find a significant yet moderate correlation between changes in the distance values from task 6 to task 9 and changes in the feedback loop reasoning from pre- to post-test, $r_s = -.599$, $p\text{-value} = .018$. Put another way, we found a significant difference in learning gains between students whose distance values decreased from task 6 to task 9 (median decrease in distance = -96.87 ; median learning gains = 2, $N = 7$) versus those students whose distance values increased (median increase in distance = 75.405; median learning gains = 0, $N = 8$), $Z(15) = 2.254$, $p\text{-value} = .032$. These findings suggest there is some evidence for our hypothesized relationship between hand movement and conceptual understanding. That is, an increase in coordination (i.e., bettering one's movements to make them more like the elicited movements by the end of the task-based interview) is associated with larger learning gains.

5 DISCUSSION

We measured student hand movements to study their sensorimotor skill in responding to elicited gestures to, in turn, understand embodied learning. Our results indicate that there is a relationship between measurable changes in how students employ elicited gestures and learning outcomes, which suggests there is value in not only eliciting content-specific gestures, but in also modelling student embodiments in a manner that accounts for temporality while still reducing dimensionality. Identifying and then analyzing the key features of such dense datasets is an important process that relies on both statistical procedures and an understanding of the learning processes to be modelled. Specifically, we show that when analyzing elicited hand movements, using statistical analyses that account for temporality, such as the HMM and the OM, is crucial for identifying those patterns that allow us to predict learning gains. Our systematic analyses make use of various statistical methods that account for temporal dependencies. These methods fit the data better than methods that do not account for temporality. We illustrate this by

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

comparing and contrasting the measurement of sensorimotor coordination at two levels of analysis. First, we compare the goodness of fit between two latent models. In using a latent model, our goal is to display patterns of bimanual coordination. The hidden Markov model proves to be a good choice for visualizing a sequence of latent states and representing sensorimotor coordination. The LCA, in contrast, proves to be a poor choice. This model has a poorer fit to the data, and the transitions between latent states are not smooth. Furthermore, the HMM produces a latent state transition matrix, whereas the LCA does not, which proves useful in further analysis of the data.

Second, we compare the similarity between the computer-elicited movement and the student-enacted movement by systematically decreasing the control of temporal dependencies in the algorithms that compute the similarity measures. This quantitative measure is important in that appraising the similarity provides a good idea of the level of sensorimotor coordination of each student. In comparing the OM distance values to other distance values computed from algorithms that do not control for different aspects of temporality (like the Hamming and Euclidean algorithms), we find the relative distances to the reference sequence (the computer-elicited sequence) are not maintained. Furthermore, not only are the relative distances not maintained, but also the distances do not seem to agree with a qualitative judgment of how each student-enacted sequence should be evaluated vis-a-vis the computer-elicited sequence. This is particularly serious for the Euclidean method, which ignores all temporality information.

Finally, we use the more robust models to predict learning outcomes. As we hypothesized, there is a clear relationship between student facility with matching the cued gestures and learning gains. This particular relationship would not be salient without these fine-grained data and the measurement model, and thus demonstrates the potential utility of these kinds of temporal analytics for predicting learning outcomes. Conceptually, we believe these findings help to demonstrate that there is a connection between student ability to fluidly model the predator–prey relationship with their hands, and their ability to learn and communicate this relationship. Future work is needed to unpack the exact nature of this relationship, but this is an important next step in demonstrating the value of both embodied designs for learning, and of high-resolution learning analytic approaches to help understand the role these designs play in learning. Specifically, these high-resolution pictures can guide future analyses exploring whether and how strong or soft embodied theories of learning operate at this time scale. For instance, because measurements of the change in sensorimotor coordination antecede measurements of learning gains, causal mechanisms like the reflexive abstraction hypothesis (Abrahamson et al., 2015; Piaget, 1952) could be tested following this approach. Furthermore, combining the results from these fine-grained analyses with classroom interaction analyses can help us understand the role of elicited gestures at larger time scales (i.e., from a sociocultural perspective of embodiment). For instance, one could study whether the ability to enact these movements coincides with the ability to transfer this understanding of quantitative patterns to a different domain (e.g., a social system like supply and demand), or whether these gestures can become a conversational resource during collaborative activities. All in all, this case study builds upon the development of multimodal learning analytics for the research and support of learning within complex

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

learning environments (Worsley et al., 2016) and related work endorsing the use of sensing technologies for the instruction of STEM concepts (Lee, 2015; Lindgren & Johnson-Glenberg, 2013).

While these findings are compelling, the study is not without limitations that we will seek to address in follow-up studies. First, in this exploratory work, we utilized a small sample size. Future studies will require a larger sample size to improve the stability of the statistical tests. Additionally, though the intervention has a pre–post design, the lack of a control group cannot rule out other sources of variation and thus the correlation detected here cannot imply causation. Despite these limitations, we feel this study sets out an initial framework for temporal analysis of patterns related to movement and gestures.

6 CONCLUSION

The present analysis is important as a use-case to illustrate the application of temporal analytics in analyzing movement data, and for the study of the consequences of using statistical analyses when such information is ignored. The availability of sensing technologies is bound to increase in coming years, and researchers are taking advantage of this availability for the design of new learning environments that build upon embodied interfaces. The analyses we present in this paper can serve as a reference guide for how to include temporal analytics in movement data. Researchers who use sensing technologies might be interested in using temporal analytics, which can prove useful for increasing our understanding of the interaction between sensorimotor schema and student reasoning. The two statistical methods employed here, the HMM and OM, and the general approach outlined, however, are only one possible avenue for carrying out these analyses. The particular temporal statistical methods a researcher selects must account for the kinds of variables and types of datasets available to them. In any case, researchers should think about these aspects of temporality before data collection, especially if synchronization across data streams would be a requirement. Finally, our approach suggests that a clear relationship exists between how students take up elicited gestures, and how they learn the patterns being modelled by those gestures. This is an important step in advancing our understanding of the relationship between embodiment and learning. Future work can continue to unpack this relationship and suggest additional nuances in how to leverage gestures to support learning in this and other contexts.

REFERENCES

- Abbott, A., & Tsay, A. (2000). Sequence analysis and optimal matching methods in sociology review and prospect. *Sociological Methods & Research*, 29(1), 3–33. <http://dx.doi.org/10.1177/0049124100029001001>
- Abrahamson, D. (2014). Building educational activities for understanding: An elaboration on the embodied-design framework and its epistemic grounds. *International Journal of Child–Computer Interaction*, 2(1), 1–16. <http://dx.doi.org/10.1016/j.ijcci.2014.07.002>

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

- Abrahamson, D., & Lindgren, R. (2014). Embodiment and embodied design. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 358–376). Cambridge, UK: Cambridge University Press.
- 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. <http://dx.doi.org/10.1159/000443153>
- Abrahamson, D., & Sánchez-García, R. (2016). Learning is moving in new ways: The ecological dynamics of mathematics education. *Journal of the Learning Sciences*, 25(2), 203–239. <http://dx.doi.org/10.1080/10508406.2016.1143370>
- Agresti, A. (2014). *Categorical data analysis*. Hoboken, NJ: John Wiley & Sons. 10.1007/978-3-642-04898-2_161
- Anderson, M. L. (2003). Embodied cognition: A field guide. *Artificial Intelligence*, 149(1), 91–130. [http://dx.doi.org/10.1016/S0004-3702\(03\)00054-7](http://dx.doi.org/10.1016/S0004-3702(03)00054-7)
- Andrade, A., Danish, J. A., & Maltese, A. (2017). Why are you gesturing? Elicited gestures and learning gains in an embodied learning environment. Paper presented at the Annual Meeting of the American Educational Research Association AERA 2017, San Antonio, TX.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617–645.
- Barsalou, L. W. (2010). Grounded cognition: Past, present, and future. *Topics in Cognitive Science*, 2(4), 716–724. <http://dx.doi.org/10.1111/j.1756-8765.2010.01115.x>
- Black, J. B., Segal, A., Vitale, J., & Fadjo, C. (2012). Embodied cognition and learning environment design. In D. Jonassen & S. Land (Eds.), *Theoretical foundations of learning environments* (2nd ed., pp. 198–223). New York: Routledge.
- Collins, L. M., & Lanza, S. T. (2013). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences* (Vol. 718). Hoboken, NJ: John Wiley & Sons.
- Danish, J. A., Enyedy, N., Saleh, A., Lee, C. D., & Andrade, A. (2015). Science through technology enhanced play: Designing to support reflection through play and embodiment. In O. Lindwall, P. Hakkinen, T. Koschman, P. Tchounikine, & S. Ludvigsen (Eds.), *Exploring the Material Conditions of Learning: Proceedings of the 11th International Conference on Computer Supported Collaborative Learning (CSCL 2015)*, 7–11 June 2015, Gothenburg, Sweden (Vol. 1, pp. 332–339). International Society of the Learning Sciences. <http://dx.doi.org/10.22318/cscl2015.313>
- Enyedy, N., & Danish, J. A. (2015). Learning physics through play and embodied reflection in a mixed reality learning environment. In V. Lee (Ed.), *Learning technologies and the body: Integration and implementation in formal and informal learning environments* (pp. 97–111). New York: Routledge.
- Enyedy, N., Danish, J., DeLiema, D., Saleh, A., Lee, C., Morris, N., & Illum, R. (2017). Social affordances of mixed reality learning environments: A case from the science through technology enhanced play project (STEP). *Proceedings of the 50th Hawaii International Conference on System Sciences (HICSS-50)*, 4–7 January 2017, Waikoloa, HI, USA. IEEE Computer Society. <http://dx.doi.org/10.24251/HICSS.2017.254>

(2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>

- Gabadinho, A., Ritschard, G., Mueller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, 40(4), 1–37. <http://dx.doi.org/10.18637/jss.v040.i04>
- Gollery, M. (2008). *Handbook of hidden Markov models in bioinformatics*. London: Chapman & Hall/CRC Press.
- Hagenaars, J. A., & McCutcheon, A. L. (2002). *Applied latent class analysis*. New York: Cambridge University Press.
- Hall, R., Ma, J. Y., & Nemirovsky, R. (2014). Re-scaling bodies in/as representational instruments in GPS drawing. In V. Lee & M. Linn (Eds.), *Technology and the body: Perspectives from the learning sciences* (pp. 112–131). New York: Routledge.
- Hamming, R. W. (1950). Error detecting and error correcting codes. *Bell System Technical Journal*, 29(2), 147–160. <http://dx.doi.org/10.1002/j.1538-7305.1950.tb00463.x>
- Hokayem, H., Ma, J., & Jin, H. (2015). A learning progression for feedback loop reasoning at lower elementary level. *Journal of Biological Education*, 49(3), 246–260. <http://dx.doi.org/10.1080/00219266.2014.943789>
- Hutto, D. D., Kirchhoff, M. D., & Abrahamson, D. (2015). The enactive roots of STEM: Rethinking educational design in mathematics. *Educational Psychology Review*, 27(3), 371–389. <http://dx.doi.org/10.1007/s10648-015-9326-2>
- Lakoff, G., & Johnson, M. (1999). *Philosophy in the Flesh*. New York: Basic Books.
- Lee, V. (2015). *Learning technologies and the body: Integration and implementation in formal and informal learning environments*. New York: Routledge.
- Lindgren, R. (2015). Getting into the cue: Embracing technology-facilitated body movements as a starting point for learning. In V. Lee (Ed.), *Learning technologies and the body: Integration and implementation in formal and informal learning environments* (pp. 39–54). New York: Routledge.
- Lindgren, R., & Johnson-Glenberg, M. (2013). Emboldened by embodiment: Six precepts for research on embodied learning and mixed reality. *Educational Researcher*, 42(8), 445–452. <http://dx.doi.org/10.3102/0013189X13511661>
- Linzer, D. A., & Lewis, J. B. (2011). polCA: An R package for polytomous variable latent class analysis. *Journal of Statistical Software*, 42(10), 1–29.
- Ma, J. Y. (2017). Multi-party, whole-body interactions in mathematical activity. *Cognition and Instruction*, 35(2), 141–164. <http://dx.doi.org/10.1080/07370008.2017.1282485>
- MacDonald, I. L., & Zucchini, W. (1997). *Hidden Markov and other models for discrete-valued time series* (Vol. 110). London: Chapman & Hall.
- Nathan, M. J., Walkington, C., Boncoddio, R., Pier, E., Williams, C. C., & Alibali, M. W. (2014). Actions speak louder with words: The roles of action and pedagogical language for grounding mathematical proof. *Learning and Instruction*, 33, 182–193. <http://dx.doi.org/10.1016/j.learninstruc.2014.07.001>
- Pallant, J. (2007). *SPSS survival manual: A step-by-step guide to data analysis using SPSS version 15*. New York: McGraw Hill.

- (2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <http://dx.doi.org/10.18608/jla.2017.43.3>
- Piaget, J. (1952). *The origins of intelligence in children*. New York: International University Press.
- Smith, C. P., King, B., & Hoyte, J. (2014). Learning angles through movement: Critical actions for developing understanding in an embodied activity. *The Journal of Mathematical Behavior*, 36, 95–108. <http://dx.doi.org/10.1016/j.jmathb.2014.09.001>
- Vermunt, J. K., & Magidson, J. (2004). Latent class models. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 175–198). Thousand Oaks, CA: Sage Publications.
- Vermunt, J. K., Tran, B., & Magidson, J. (2008). Latent class models in longitudinal research. In S. Menard (Ed.), *Handbook of longitudinal research: Design, measurement, and analysis* (pp. 373–385). New York: Academic Press/Elsevier.
- Visser, I., & Speekenbrink, M. (2010). depmixS4: An R-package for hidden Markov models. *Journal of Statistical Software*, 36(7), 1–21.
- Wilensky, U., & Reisman, K. (2006). Thinking like a wolf, a sheep, or a firefly: Learning biology through constructing and testing computational theories — an embodied modeling approach. *Cognition and Instruction*, 24(2), 171–209. http://dx.doi.org/10.1207/s1532690xci2402_1
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4), 625–636. <http://dx.doi.org/10.3758/BF03196322>
- Worsley, M., Abrahamson, D., Blikstein, P., Grover, S., Schneider, B., & Tissenbaum, M. (2016). Situating multimodal learning analytics. In C. K. Looi, J. Polman, U. Cress, & P. Reimann (Eds.), *Transforming Learning, Empowering Learners: Proceedings of the 12th International Conference of the Learning Sciences (ICLS '16)*, 20–24 June 2016, Singapore (Vol. 2, pp. 1346–1349). International Society of the Learning Sciences.
- Zucchini, W., & MacDonald, I. L. (2009). *Hidden Markov models for time series: An introduction using R*. Boca Raton, FL: CRC Press.