Gesture Patterns and Learning in an Embodied XR Science Simulation

Jina Kang[1*], Morgan Diederich[1], Robb Lindgren[2] and Michael Junokas[3]
[1]Instructional Technology and Learning Sciences, Utah State University, USA // [2]Curriculum & Instruction, University of Illinois at Urbana-Champaign, USA // [3]Media and Cinema Studies, University of Illinois at Urbana-Champaign, USA // jina.kang@usu.edu // morgan.wood23@aggiemail.usu.edu // robblind@illinois.edu // junokasm@gmail.com
[*]Corresponding author

ABSTRACT: Recent research has emphasized the importance of leveraging embodied interactions for learning critical STEM concepts. ELASTIC^{3}S—an embodied environment for learning about cross-cutting concepts (i.e., non-linear growth)—allows learners to interact with different science simulations through whole-body gestures. Technological advances in gesture recognition can track and respond to students' gestures, however, there has been little investigation into how the gestures performed in these environments relate to subsequent learning. The need for sequential pattern recognition methods is critical in embodied learning if we are to understand how gestural interaction with a simulation facilitates learning. Using data collected via Microsoft Kinect V2 from twelve college students, we applied multivariate Dynamic Time Warping for clustering to identify gestural patterns in ELASTIC^{3}S as evidence for embodied learning processes. Our findings showed that identified trends of simulation use were indicative of students' struggles to understand the underlying ideas or use of the system and were associated with learning performance. These indicators can potentially be used to leverage real time, in-simulation assistance and promote a more adaptive learning experience via embodied simulations.

Keywords: Embodied learning, XR Science education simulations, Gesture recognition, DTW clustering, Time series analysis

1. Introduction

There has been significant interest in leveraging learners' embodied interactions for teaching critical STEM concepts (Lindgren et al., 2016; Nathan & Walkington, 2017; Stieff et al., 2016). This interest builds upon theories of embodied cognition that assert a fundamental connection between the actions/perceptual processes of the body and how people think and learn (Glenberg, 2010; Shapiro, 2019; Wilson, 2002). Research has shown that learners can be prompted to perform gestures and enact their emerging understanding of STEM ideas in ways that promotes new learning (Gallaher & Lindgren, 2015; Lindgren, 2014). This seems to be true even for abstract ideas and unseen processes such as molecular interactions (e.g., Mathayas et al., 2019). A particularly challenging concept that cuts across STEM domains is non-linear growth: understanding where it is present and how it differs from linear growth (Tretter et al., 2006).

We designed an XR embodied learning environment to target students' ideas about non-linear growth called ELASTIC^{3}S (Embodied Learning Augmented through Simulation Theaters for Interacting with Cross-Cutting Concepts in Science). The ELASTIC^{3}S platform allows learners to control different science simulations with user-defined whole-body gestures (e.g., hand waving, kicking). This particular implementation of XR uses gesture recognition technology paired with multiple large digital displays to create an interactive and immersive environment where students engage with science simulations through gesture. Although these types of environments have shown promise for summative learning outcomes (Johnson-Glenberg et al., 2014; Lindgren et al., 2016), there has been little investigation into how the progression of gestures that learners perform in these environments relate to their subsequent learning, and how the gestures themselves can be used for the purposes of assessment and monitoring. The purpose of this work is to apply pattern recognition algorithms to the gestures that students perform in an embodied XR science education simulation as a means of understanding what is learned and when personalized feedback should be presented.

2. Relevant work

2.1. Embodied learning

Processes of human cognition are deeply rooted in how the body interacts with the environment (Gallagher, 2006). Our understanding of how the world works is organized around the human sensorimotor system and our various modes of perception and action (Barsalou, 2008). Embodiment has increasingly become a focus in

various research domains including cognitive psychology (Glenberg, 2010), linguistics (Lakoff & Johnson, 1980), and the performing arts (Noice & Noice, 2006). In education, researchers are applying ideas from embodied cognition to the design of learning environments. "Embodied learning" is essentially the forging of meaningful connections between body movements, artifacts, and learning content (Duijzer et al., 2019; Lindgren & Johnson-Glenberg, 2013; Skulmowski & Rey, 2018). This type of learning is based around the idea that a student has agency and an active role in their learning experience, and that learning activities can be designed to effectively leverage alignments between physical modes of interaction and target concepts. Relevant work has found that the use of embodied learning leads to improved learning outcomes and understanding in multiple domains (e.g., Han & Black, 2011; Glenberg, 2008; Goldin-Meadow, 2011; Segal et al., 2014). Specifically, research has been investigated in STEM education that identified a relationship between gestures and improved scientific reasoning (Crowder, 1996).

Studies have demonstrated the learning effectiveness of embodied learning environments compared to more traditional learning settings. Johnson-Glenberg et al. (2011) conducted an experimental design studying if embodied learning using XR was more effective than traditional classroom learning in which seventy-one 9th graders participated. Results indicated that the embodied environment led to greater knowledge gains. As a follow-up study, they examined if the XR embodied learning environment was more effective than a desktop simulation, and they found that embodied learning yielded significant learning gains for chemistry and disease transmission (Johnson-Glenberg et al., 2014) and in the abstract domain of the electric field (Johnson-Glenberg & Megowan-Romanowicz, 2017). Lindgren et al. (2016) also evaluated the effects of embodied interaction on conceptual understanding and learning engagement where their results corroborated that of Johnson-Glenberg's et al. (2014) desktop simulation and embodied learning findings. Specifically, Lindgren et al. (2016) identified that the embodied learning simulation that was designed to teach critical concepts in physics led to positive results in terms of students' learning gains, engagement, and attitudes towards science.

2.2. Multimodal learning in XR

Learning is often multimodal (Jewitt, 2006) and associated with a variety of modes of communication (Ochao, 2017). Being able to capture the change of mode is critical to the understanding of learning processes. Multimodality contributes to learning via both multimodal instruction and complex multimodal representations created by learners. Multimodal instruction facilitates learning with effectively integrated representations of content across different sensory modalities (e.g., Birchfield et al., 2008).

Various methods have been developed by integrating multimodal data sources and have shown promise to further understanding of learning in an embodied environment. Traditional data such as observation, audio/video recording, and student and instructor discourse can be used to investigate embodied learning. However, such analytical processing of these data has limitations including being time consuming, error-prone, and having limited scalability (Prieto et al., 2018). Advanced technologies now enable the collection of a larger spectrum of multimodal data sources that reflect students' embodied experiences and further inform multimodal learning and instruction. The overall effectiveness of body-based learning activities has been demonstrated, however, there has been less attention given to the progression of embodied actions (e.g., gestures) performed within such a learning environment and the ways that these progressions may be conducive to learning (e.g., Smith et al., 2016). As the availability of multimodal data collected from embodied learning environments increases, identifying analytical approaches that allow for investigation and interpretation of these data becomes imperative.

2.3. Time series clustering analyses

Given the large quantities and ever-increasing complexity of data available, the need for scalable, time-series based analyses are critical (Lin et al., 2012). Scaling and performance necessitate additional revolutions in time series-driven pattern recognition. Many techniques, specifically in clustering, have evolved to meet this challenge. For instance, the KmL (K-Means for Longitudinal data) algorithm was developed out of K-means for longitudinal data (Genolini & Falissard, 2011). A benefit to using this over traditional K-means is that KmL can handle missing data seen frequently in time. Another time series clustering method, Dynamic Time Warping (DTW) clustering, has shown benefits. DTW calculates the minimum differences between two time series that can differ in length and amplitude based on the creation of an optimal warping path to assess similarities through one-to-many mapping (Li et al., 2010). The output of DTW results in a (dis)similarity metric that can then be leveraged in a clustering algorithm. For instance, Mezari and Maglogiannis (2017) used DTW on motion data to

recognize gestures. Shen and Chi (2017) also applied DTW by using 36 variables dealing with autonomy (i.e., hitcounts), temporal information (i.e., average times), and actions extracted from student interactions with a tutoring system. They explored different types of clustering methods, suggesting that popular clustering techniques such as K-Means do not account for the differing sequential nature of many educational-based problems. They explored DTW compared to the Euclidean distance and found DTW a viable method for clustering data where participants have differing lengths of time and numbers of interactions. The powerful applications of DTW are often lauded for the ability to evaluate time series of differing lengths and scalability to large quantities of data for pattern recognition.

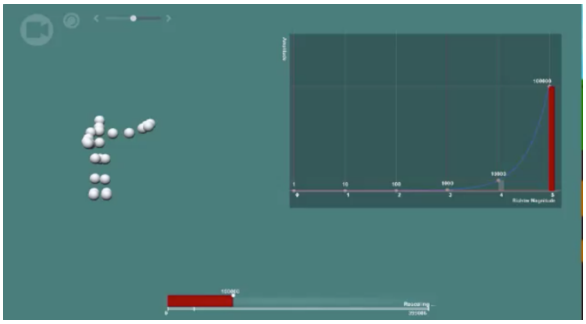
There is a paucity of research in investigating the analysis of fine-grained multimodal interaction data collected in a XR embodied learning environments, and how to interpret the interaction data of learners' embodied actions. The goal of this paper, therefore, is to investigate the massive and dynamic gesture data generated as students interact with an XR embodied science simulation and how such gesture data relate to students' embodied learning experiences. The main contribution of this paper is that we employ a novel analytical method (i.e., DTW clustering) to the fine-grained time-series gesture data. In particular, we adapt the data aggregation methods in Shen and Chi (2017) to investigate different granularities of gesture data and describe how each data granularity reveals different meanings. The present study is an exploratory one, in which we explore and describe different levels of analyses to discover the best data windowing techniques for identifying meaningful sequential patterns of students' embodied actions. In addition, we explore different features derived from gesture data, and we identify potential features for personalized feedback that facilitate students' productive whole-body movements and their learning in the future implementation of XR embodied learning environments.

Our primary hypothesis is that certain patterns identified from students' gesture data (e.g., time spent, volume, or speed on a specific gesture) will be related to their learning performance. For example, students who exhibit an increasing trend in their time spent on gestures that are misaligned with the target mathematical relationship (e.g., doing an addition gesture when a multiplication gesture is called for) to solve a problem will show lower learning gains. We aim to find new ways to analyze multimodal interaction data that will reveal any underlying connections between body movement and learning. This exploratory work lays the groundwork for automated detection of student embodied behaviors that ultimately supports personalized learning.

3. Methods

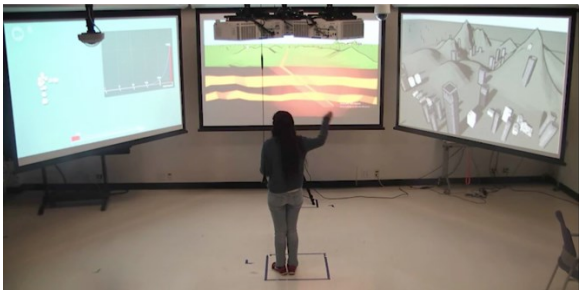
3.1. ELASTIC^{3}S

ELASTIC^{3}S (see Figure 1) was developed using the Unity engine for Microsoft Kinect V2 to empower high school and undergraduate students to build scientific knowledge around the crosscutting concept of scale, proportion, and quantity. This crosscutting concept was used to bridge the science topics of (1) earthquakes and (2) acidity/basicity. To detect students' movements and provide them with real-time skeletal-motion feedback (see Figure 1a), a gesture recognition system was developed using a hierarchical hidden Markov model (described in Junokas et al., 2018). The system is adaptive; that is, the system learns each participant's different types of gestures associated with four mathematical operations (i.e., +1, −1, ×10, and ÷10), which the system uses to recognize the participant's real-time skeletal data. This paper focuses on the gesture-based data that was collected during the earthquake simulation of ELASTIC^{3}S, which explored the concept of linear and non-linear growth through the application of the Richter scale. Students began by developing personal gestures using the metaphorical framing for each mathematical function we identified during early pilot interviews (Alameh et al., 2016). For example, participants were prompted to think of (1) addition as stacking a cube on top of a pile of other cubes, (2) subtraction as kicking one cube out of a pile, (3) multiplication as folding copies of a certain quantity on top of each other, and (4) division as splitting a stack into smaller groups. While students were cued to create gestures that adhered to this framing, the "one-shot" gesture recognition system (Junokas et al., 2018) meant that each student could develop their own personal gestural representation. The gestures that students created allowed them to explore the exponential concepts in different science topics (i.e., earthquakes and acidity/basicity).



(a) Student's Skeleton and Two Bar Graphs

(b) Earth's Fault Line



(c) Three-Screen Simulation Space

Figure 1. Screenshots of the ELASTIC^{3}S XR Earthquake Simulation. *Note.* (a) Two bar graphs indicate the student's input magnitude; (b) The Earth's fault line shows the building of pressure associated with the magnitude; (c) A student is making a multiplication gesture.

Once students completed the training phase, students were provided five sequential tasks that varied in difficulty. First, they are prompted to set the amplitude of the seismic waves to create a different magnitude of earthquake. Students begin with a straightforward task, which we call M2, where they create a magnitude 2 earthquake (corresponds to 100 amplitude units, 10^2). The most efficient use of gestures to complete this task would be, starting at 0, to add 1, then multiply by 10 and multiply by another 10, resulting in an amplitude of 100 units.

Each task moved students through varying complexities of their gestures as well as conceptual understandings; for example, M3.5 requires students to apply their acquired knowledge of the exponential nature of the Richter scale in order to create a magnitude of 3.5 earthquake. This magnitude is not halfway between amplitudes of magnitude 3 and 4 and requires an in depth understanding of exponential growth. The task names, the corresponding amplitude, and level of difficulty are listed in Table 1.

Table 1. Simulation tasks and difficulty

Task order	Task name	Corresponding amplitude	Difficulty level*
1	M2	$10^2 = 100$	Mid - Hard
2	M3	$10^3 = 1{,}000$	Easy - Mid
3	M3.5	$10^{3.5} = 3{,}162.28$	Hard
4	M7	$10^7 = 10{,}000{,}000$	Easy
5	M8	$10^8 = 100{,}000{,}000$	Easy

Note. *Relative to other given tasks.

3.2. Participants

A total of twenty-four undergraduate students from the midwestern region of the United States participated in the earthquake simulation, which consisted of a pre-test, simulation session, and post-test. Following the IRB-approved protocol, each participant individually completed a task-based interview consisting of (1) a pre-test, (2) a simulation session lasting about 30 minutes, and (3) a post-test in a lab with a facilitator in the room. The pre-/post-tests consisted of questions that assessed their understanding of earthquakes and linear/non-linear growth in both the context of earthquakes and new contexts (see details in 3.3.3). During the simulation session, the participants were asked to engage with the earthquake simulation. Both the pre-/post-tests and simulation session were audio and video recorded. Unfortunately, the logs of early student participants were not recorded, with one additional student identified as an outlier via visualizations (see section 3.3.2) thus leaving twelve students' data

available for this study. Of these twelve students (mean age = 19.2), 58% (n = 7) stated they were female, 66% were white, two preferred not to answer, one student was Asian, and one identified as multiracial.

3.3. Data sources

3.3.1. Kinect data

The participants completed five tasks in a 30-minute earthquake simulation session. This paper focuses on the data collected via Microsoft Kinect V2 during the simulation session. Datapoints were collected at a rate of approximately 30 frames per second, in which the 3D coordinates (x, y, z) of 25 skeleton joints relative to the Kinect were tracked and recorded. Particularly, each joint's coordinates were recorded based on positions relative to the position of the Kinect sensor (see Figure 2).

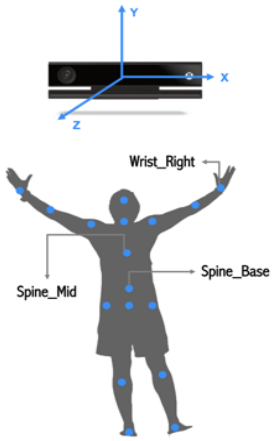


Figure 2. Visual representations of data collection including joints of interest

3.3.2. Data preprocessing and feature extraction

The first level of preprocessing involved data cleaning through annotation of session recorded video, documenting when participants started and stopped each gesture and the type of gesture being performed. This information was then synced with the Kinect data and resulted in a dataset that contained all recorded data, the gesture type, gesture order, timestamp, and the task being completed. We excluded any Kinect data without annotations. We also visualized the data to remove unintended movements recorded by the Kinect. It was identified that one participant had extreme outliers in M2 and M3.5 in both volume and speed, which appeared to be due to system glitches. That student was therefore removed from the dataset.

Using the cleaned data, we developed various simulation use measures (see Table 2). First, the *speed* of a gesture was measured by finding the magnitude of the velocity of a given joint position, formally expressed as (p_n: n_{th} joint position, t: time of a given joint position, s: speed):

$$\Delta p = p_1 - p_0$$
$$\Delta t = t_1 - t_0$$
$$\Delta s = \left|\frac{\Delta p}{\Delta t}\right|$$

This can be ultimately measured at individual, combinations, and/or the complete joint positions. The *volume* of a gesture was measured by finding the product of the Euclidean distance between the maximum and minimum points of given joint positions at each dimension (x, y, z), formally expressed as (v: volume):

$$d_x = \sqrt{(x_{max} - x_{min})^2}$$
$$d_y = \sqrt{(y_{max} - y_{min})^2}$$
$$d_z = \sqrt{(z_{max} - z_{min})^2}$$
$$V = d_X * d_Y * d_Z$$

This rectangular projection of volume was measured on all joints, providing a spatial perspective to a performed gesture. We then used the distance between two joint positions: spine_base and spine_mid (see Figure 2; these data points represent each participant's height) for the normalization, rescaled by each participant's body size.

While such frame-level data (referred to as "frame granularity") showed in-moment data, we wished to identify what level of aggregation was most important and made the most sense. This was completed by comparing multiple aggregations on frame granularity data (Shen & Chi, 2017). We therefore aggregated the frame granularity data by each gesture (referred to as "gesture granularity") to see if there would be a difference between our two types of granularity: gesture and frame. We found that compared to the frame granularity, the gesture granularity was more interpretable, as we were interested in the gesture as a whole, rather than a single frame. As for speed and volume of each gesture, the maximum, minimum, variance, and average were created for both variables on each distinct gesture to retain as much information as possible.

Further, we needed to know how long the gesture took (i.e., timestamp at the last frame – timestamp at the first frame) and the type of gesture that was made. We named these measures as: "total time spent on" addition, subtraction, multiplication, and division. In addition, during the aggregation process, we naturally lost the granularity of frames. We therefore included how many "frames" were made in the creation of each gesture, which was named as "number of frames." Table 2 shows the final list of variables (i.e., simulation use measures), the result of the aforementioned data preprocessing and feature extraction efforts.

Table 2. Simulation use measures for gesture level analysis

Variable name	Description	Data type	Example
Task	Earthquake simulation tasks	Character	M2
Volume	Four features (average, variance, max, min) generated from the aggregation of [1]normalized volume.	Decimals (m^3)	0.0033
Speed	Four features (average, variance, max, min) generated from the aggregation of [1]normalized speed.	Decimals (meter/second)	0.0437
Number of frames	Number of rows condensed to produce aggregated row at the gesture level	Whole Number	79
Time spent (on gesture)	Each gesture was considered a feature, resulting in four features (addition, subtraction, multiplication, division).	Decimals (second)	1.2853

Note. [1]Normalization occurred on the frame level which standardized each motion across all participants for comparison.

3.3.3. Learning performance data

The four researchers scored each participant's understanding of earthquake concepts and exponential growth during pre- and post-tests (Kang et al., 2018). Table 3 shows the descriptive statistics of learning performance data. During pre-/post-tests, the facilitator asked several questions to see the participant's understanding of earthquakes (i.e., conceptual knowledge) and exponential growth in the earthquakes (i.e., exponential knowledge) and new contexts (i.e., transfer knowledge) (See sample questions in Appendix A). In this study, the normalized gain score of each category (conceptual, exponential, and transfer knowledge) was calculated by ((posttest score) – (pretest score)) / ((total available score) – (pretest score)) (Hake, 1998). These learning performance measures were used later to examine if any identified clusters showed significant intergroup differences of learning performance.

Table 3. Descriptive statistics for learning performance

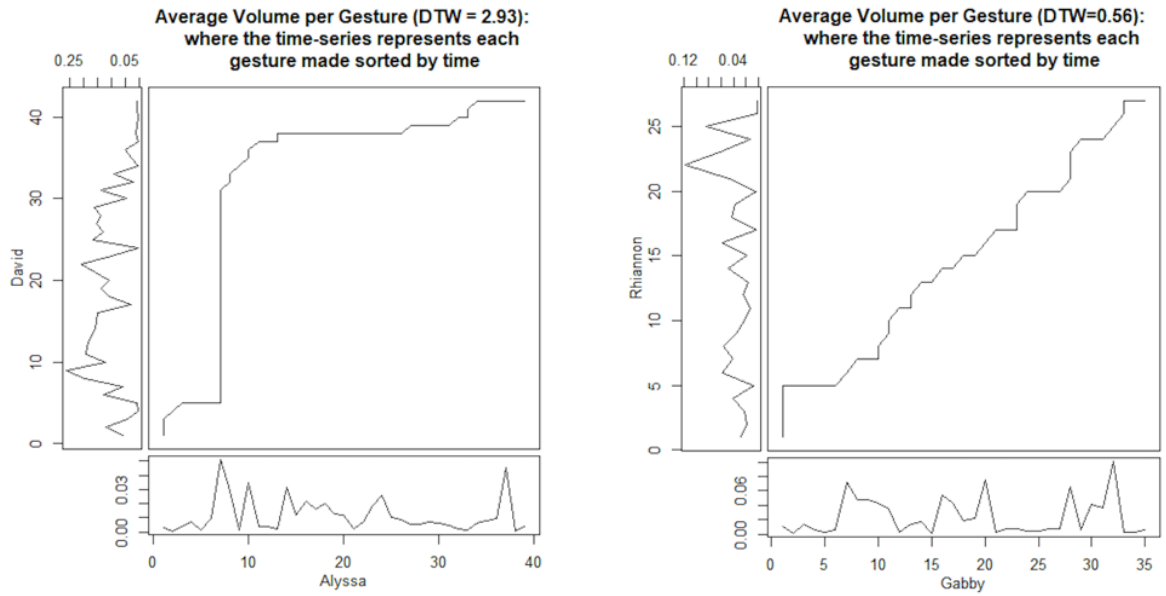
	Conceptual			Exponential			Transfer		
	Pre	Post	[1]Gain	Pre	Post	[1]Gain	Pre	Post	[1]Gain
Mean	3.08	8.25	0.75	4.63	7.67	0.28	3.00	2.88	-0.08
SD	0.90	1.08	0.15	3.48	2.63	0.28	1.65	1.86	0.50

Note. [1]Normalized gain score (Hake, 1998).

3.4. Analyses

The gesture granularity data utilized in this study are multivariate time series (see the variables in Table 2). Several considerations were observed to maximize effectiveness of this multivariate information in categorizing individuals based on their embodied learning. For example, time series data is a critical aspect of understanding student embodied learning process, and individuals did not complete a task in the same number of frames or gestures. DTW takes into consideration time series of different lengths, resulting in a computed similarity value of each pair of time series data, which is similar to a Euclidean distance (Shen & Chi, 2017). This results in a distance matrix comparing distance similarity values with each other.

For a single variate, visual example on warping paths and the generation of a similarity metric we present two examples in Figure 3: (1) DTW dissimilarity = 2.93 between David and Alyssa (Figure 3a), (2) DTW dissimilarity = 0.56 between Gabby and Rhiannon (Figure 3b). In each graph, we see three panels in each warping path. Each student's behavior over the course of time for a given variable is shown in each vertical (left) and horizontal (lower) panel. The central panel is the cost matrix, where we see the warping path between the two students' average volume patterns over time. A score of zero, or a diagonal line, indicates that the patterns are identical and no warping was required to "match" one pattern to another. The more warping that is required, the larger the metric is, indicating more dissimilarity between two students' behaviors.



(a) Comparison of David and Alyssa, DTW = 2.93

(b) Comparison of Gabby and Rhiannon, DTW = 0.56

Figure 3. Average volume comparison between a pair of students

Further, we sought to understand how different levels of the data could yield meaningful clustering results that address our objectives. Three levels: (1) task level, (2) subsequence level, and (3) all tasks level were created to view the data in multiple ways (see Table 4). For example, M3.5 on the task level used gesture data from only task 3.5. For M2/3/3.5 in subsequence level, we included the student's first gesture made in M2 to their last in M3.5. Therefore, subsequence level was ultimately used to understand the overall trajectory of the behavior up to that point. Task level and subsequence level clustered twelve matrices, each of which included each of twelve participants' multivariate gesture granularity data for each level of analysis. Finally, for all tasks level, we included all participants' gesture granularity wherein each student's behavior per task was evaluated against all other students' task behavior. That is, all task level clustered a total number of sixty matrices (i.e., twelve students × five tasks = sixty matrices), each of which included the multivariate gesture granularity data each participant made during each session.

Once the DTW value or metric was derived for each pair of data, the values were used in the clustering process. We selected the Hierarchical clustering method using Wards Linkage to ultimately partition students into clusters. The optimal number of clusters was identified using Silhouette method for each analysis level: 3 clusters. This was all completed in the tsclust package in R. To examine statistically significant differences of each variable across three clusters in each level of analysis, we first checked the assumptions of one-way

ANOVA. Along with our small sample size, the normality and homogeneity of variances were violated. Therefore, we performed Kruskal-Wallis non-parametric analyses of gesture and learning performance variables (i.e., normalized gain on conceptual, exponential, and transfer knowledge) of each cluster generated from task level, subsequence level, and all tasks level. Then, we conducted post hoc tests with the Bonferroni adjustment (Kruskal & Wallis, 1952) to examine statistically significant differences of each variable between each pair of three clusters.

Table 4. Different levels considered for clustering and other analyses

Analysis level	Tasks included in each analysis (# gesture granularity data)	Data structure examples[1]
Task level	M2 (n = 64)	A.M2, B.M2 … T.M2
	M3 (n = 21)	A.M3, B.M3 … T.M3
	M3.5 (n = 173)	A.M3.5, B.M3.5… T.M3.5
	M7 (n = 86)	A.M7, B.M7… T.M7
	M8 (n = 21)	A.M8, B.M8 … T.M8
Subsequence level	M2 (n = 64)	A.M2, B.M2 … T.M2
	M2, M3 (n = 85)	(A.M2+A.M3), (B.M2+B.M3) … (T.M2+T.M3)
	M2, M3, M3.5 (n = 258)	(A.M2+A.M3+A.M3.5) … (T.M2+T.M3+T.M3.5)
	M2, M3, M3.5, M7 (n = 344)	(A.M2+A.M3+A.M3.5+A.M7) …
	M2, M3, M3.5, M7, M8 (n = 365)	(A.M2+A.M3+A.M3.5+A.M7+A.M8) …
All tasks level	M2, M3, M3.5, M7, M8 (n = 365)	A.M2, B.M2 … T.M7, T.M8

Note. [1]Letters that come prior to task information are shortened aliases for students. A.M2 is Alyssa's data from task M2 exclusive.

4. Findings

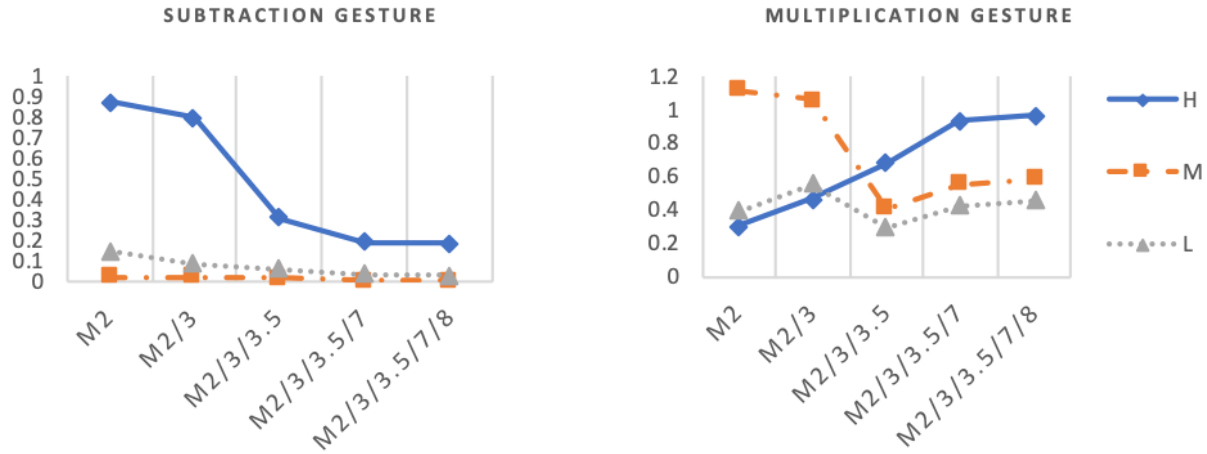
4.1. Task level vs. Subsequence level

To identify students' gestural patterns, we first explored two different levels: task level and subsequence level. As described in 3.4, non-parametric analyses examined statistically significant differences of gesture and learning variables across three clusters. Of interest, the number of frames variable showed the significant intergroup differences in every task and subsequence level. Overall gesture characteristics of each cluster were identified as "High-Frame," "Mid-Frame," and "Low-Frame" mainly using each cluster's mean rank of number of frames. Table 5 shows the characteristics of each cluster in the subsequence level analyses. In particular, we used the results of K-W analyses and post hoc tests by including selected statistically significant variables (i.e., simulation use measures, learning performance variables) identified in each segment analysis.

The task level analyses showed the participants' significant gestural behavior differences across three clusters during each task. One interesting finding is that a High-Frame cluster showed the lowest transfer change ($High_{\_mean\ rank}$ = 2.0, $Mid_{\_mean\ rank}$ = 8.0, $Low_{\_mean\ rank}$ = 8.0) during the M7 task ($\chi^2(2) = 7.748$, $p < .05$) and the lowest exponential change ($High_{\_mean\ rank}$ = 4.20, $Mid_{\_mean\ rank}$ = 9.50, $Low_{\_mean\ rank}$ = 4.75) during the M8 task ($\chi^2(2) = 6.009$, $p < .05$). The High-Frame clusters during M7 and M8 show the same tendency of longer time spent on gesture multiplication (M7: $\chi^2(2) = 55.658$, $p = .000$; M8: $\chi^2(2) = 10.515$, $p = .005$). Potentially at this point, the students who showed this gestural characteristic could be provided better scaffolding, which could, in turn, be more successfully transferred to novel contexts.

The subsequence level analyses also identified three clusters in each aggregated segment. As shown in Table 5, within each subsequence segment, all High-Frame clusters show the longest time spent on gesture subtraction, while all Mid-Frame clusters show the shortest time spent on gesture subtraction, as evaluated using K-W non-parametric testing. The two gestures, subtraction and division, indicate a student made an adjustment of the gestures they have added, realizing previous gestures were made incorrectly. To complete the tasks in the optimal way, these two gestures are not required at all. The presence of these patterns in the data suggests that students may have had difficulty employing the correct sequence of gestures that were required. Figure 4 shows the average time spent trend (i.e., the average time spent per a single gesture) on two gesture types: gesture multiplication and gesture subtraction. One interpretation may be that students initially had trouble grasping the system, as the average time spent on Gesture Subtraction decreases over the segments (see Figure 4a). However, we also recognize that the High-Frame groups show significantly more time spent on gesture subtraction during

the first two tasks than the other groups which may signify the overall difficulty these students had and may require additional support in situ.



(a) Average Time Spent Cluster Trends on Gesture Subtraction

(b) Average Time Spent Cluster Trends on Gesture Multiplication

Figure 4. Average time spent trends on gesture

As shown in Figure 4b, High-Frame clusters show an increasing trend, while the Mid- and Low-Frame clusters show a decreasing trend. Particularly, the High-Frame cluster show the lowest time spent on multiplication gestures for the first two subsequence segments: M2 and M2/3.5. The High-Frame clusters then overtook the other two clusters when the later tasks' data (i.e., M3.5, M7) were added. This increasing average time spent on multiplication gesture may indicate this is a point of tension for the student. During the last two segments, compared to the Mid-Frame clusters, the students in the High-Frame clusters also appeared to achieve lower normalized gain on transfer knowledge, indicating they were unable to transfer their knowledge to new contexts.

In Figure 5, the normalized volume per gesture shows the similar trends across all clusters; that is, the increasing trend on the M2/3 and M2/3/3.5/7/8 subsequences. When we included M3 in the subsequence (i.e., the M2/3 segment), the average volume made for each gesture increased. This was echoed when M8 data was included in the full sequence. The High-Frame clusters showed consistently the lowest average of volume over the entire task progression. Task M8 (from M7) is the easiest task that requires students to go over the similar thinking process they practiced during M2 and M3. Additionally, the Mid-Frame group showed the sharpest increasing trend of average volume from M7 to M8. Given the highest transfer knowledge gain Mid-Frame groups showed, the tendency of increasing volume during each following task of M2 or M7 may indicate students' confidence on their gestures.

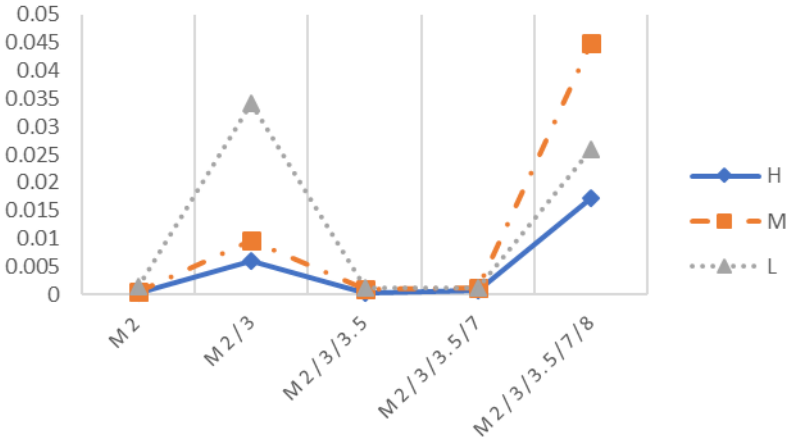


Figure 5. Average normalized volume trend

Table 5. Cluster characteristics: Gesture and learning - subsequence level

Tasks included	# Granularity data[1]	Selected key variables[2]	High-Frame Cluster	Mid-Frame Cluster	Low-Frame Cluster	Post hoc results[5]
M2	Total = 64 High = 20 Mid = 21 Low = 20	Frames ($\chi^2(2) = 25.121$, $p = .000^{**}$)	MR[4] = 45.33	MR = 36.55	MR = 17.65	**High-Low; **Mid-Low
		Timespent_S[3] ($\chi^2(2) = 18.017$, $p = .000^{**}$)	MR = 43.25	MR = 26.19	MR = 28.91	**High-Mid; **High-Low
		Timespent_M ($\chi^2(2) = 17.228$, $p = .000^{**}$)	MR = 23.55	MR = 44.95	MR = 28.91	**High-Mid; **Mid-Low
M2/ M3	Total = 85 High = 25 Mid = 30 Low = 30	Frames ($\chi^2(2) = 25.121$, $p = .000^{**}$)	MR=63.73	MR = 36.55	MR = 17.65	**High-Low; **Mid-Low
		Timespent_S[3] ($\chi^2(2) = 18.017$, $p = .000^{**}$)	MR = 43.25	MR = 26.19	MR = 28.91	**High-Mid; **High-Low
		Timespent_M ($\chi^2(2) = 17.228$, $p = .000^{**}$)	MR = 23.55	MR = 44.95	MR = 28.91	**High-Mid; **Mid-Low
M2/ M3/ M3.5	Total = 258 High = 57 Mid = 147 Low = 52	Frames ($\chi^2(2) = 107.164$, $p = .000^{**}$)	MR = 202.11	MR = 129.02	MR = 54.13	**High-Mid; **High-Low
		Volume_Average ($\chi^2(2) = 17.482$, $p = .000^{**}$)	MR = 106.7	MR = 146.11	MR = 106.88	**High-Mid; **Mid-Low
		Speed_Average ($\chi^2(2) = 10.973$, $p = .004^{**}$)	MR = 131.82	MR = 139.11	MR = 99.42	**Mid-Low
		Timespent_S ($\chi^2(2) = 20.970$, $p = .000^{**}$)	MR = 146.32	MR = 123.48	MR = 128.33	**High-Mid; *High-Low
M2/ M3/ M3.5/ M7	Total = 344 High = 98 Mid = 138 Low = 108	Frames ($\chi^2(2) = 177.092$, $p = .000^{**}$)	MR = 268.41	MR = 173.74	MR = 83.88	**High-Mid; **Mid-Low; **High-Low
		Volume_Average ($\chi^2(2) = 10.113$, $p = .006^{**}$)	MR = 155.46	MR = 193.13	MR = 161.6	*High-Mid; *Mid-Low
		Speed_Average ($\chi^2(2) = 9.315$, $p = .010^{*}$)	MR=183.58	MR=183.4	MR = 148.52	*Mid-Low; *High-Low
		Timespent_S ($\chi^2(2) = 17.125$, $p = .000^{**}$)	MR = 185.27	MR = 165.19	MR = 170.26	**High-Mid; *High-Low
		Timespent_M ($\chi^2(2) = 15.129$, $p = .001^{**}$)	MR = 202.1	MR = 167.29	MR = 152.3	**High-Mid; *High-Low
		Transfer Change ($\chi^2(2) = 8.437$, $p = .015^{*}$)	MR = 2.00	MR = 8.80	MR = 7.00	*High-Mid
M2/ M3/ M3.5/ M7/ M8	Total = 365 High = 101 Mid = 145 Low = 119	Frames ($\chi^2(2) = 11.192$, $p = .000^{**}$)	MR = 287.16	MR = 187.42	MR = 89.21	**High-Mid; **Mid-Low; **High-Low
		Volume_Average ($\chi^2(2) = 7.660$, $p = .022^{*}$)	MR = 165.39	MR = 201.16	MR = 175.82	*High-Mid
		Timespent_S ($\chi^2(2) = 17.874$, p	MR = 196.38	MR = 175.7	MR = 180.53	**High-Mid; **High-Low

	=.000**)				
	Timespent_M ($\chi^2(2) = 16.575$, p =.000**)	MR = 215.51	MR = 178.53	MR = 160.84	*High-Mid; **High-Low
	Transfer Change ($\chi^2(2) = 8.437$, p =.015*)	MR = 2.00	MR = 8.80	MR = 7.00	*High-Mid

Note. [1]Number of gesture granularity data included in the analysis. [2]Key selected significant variables from non-parametric analyses of each cluster. [3]Total time spent on gesture (A: Addition, S: Subtraction, M: Multiplication, D: Division). [4]Mean Rank. [5] Post hoc test results using Bonferroni correction. *indicates p-value < .05. **indicates p-value < .01.

4.2. All tasks: Membership transfer

To examine how each student shifted cluster memberships over five different tasks, we completed another analysis in which clustering took place where each task was evaluated for each person. Three clusters were selected as the optimal number of clusters. Similarly, the number of frame feature shows statistically significant differences across three clusters; therefore, we also used the mean ranks of number of frames to label each cluster. The non-parametric tests and post hoc analyses showed some interesting statistically significant variables across the three clusters: time spent on gesture addition ($\chi^2(2) = 38.597$, $p = .000$) and time spent on gesture subtraction ($\chi^2(2) = 7.020$, $p = .030$). For example, High-Frame cluster shows the longest time spent on gesture addition tendency ($High_{\_mean\ rank} = 212.95$), while Mid-Frame cluster shows the shortest time spent tendency ($Mid_{\_mean\ rank} = 114.5$). Such patterns of High-Frame cluster may indicate students were less strategic or struggling, since the use of gesture addition is indicative of less understanding of non-linear relationships between the amplitudes of seismic waves and the magnitude.

Table 6. Membership transfer at all tasks level

Pseudonym	M2	M3	M3.5	M7	M8	[2]Key learning gains
Alyssa	H[1]	H	H	H	M	Low Exponential/Conceptual/Transfer
Blair	H	H	H	H	M	Low Transfer
David	H	M	H	L	M	None
Gabby	H	M	H	L	L	Low Conceptual; High Exponential
George	L	L	L	L	M	None
Louise	L	L	L	L	L	High Exponential/Transfer
Matthew	L	M	L	L	M	Low Exponential
Mindy	L	M	L	L	M	None
Rhiannon	L	M	H	L	L	High Exponential/Transfer
Rosalind	L	M	L	L	M	High Exponential
Steven	M	M	H	H	M	Low Conceptual/Exponential/Transfer
Tabitha	M	M	L	L	M	Low Conceptual; High Transfer

Note. [1]This is an abbreviation of each cluster (H: High-Frame, M: Mid-Frame, L: Low-Frame). [2]This is a summary of each student's learning gain in each category of pre-/post-test scores. Low indicates the 25% of participants who had the least learning gains, where High represents the 25% participants who made greater learning gains.

Table 6 shows each participant's clustering membership transfer over the five tasks and their key learning gain. Notably, the majority of students were assigned to a High-Frame cluster during M3.5, the most challenging task. After the completion of the first two tasks (M2, M3), we expect students to no longer stay in High-Frame cluster during the second set of the similar practice (M7, M8), where they were able to apply what they learned from the earlier tasks. Therefore, the students who exhibit the patterns staying in High-Frame cluster (i.e., Alyssa, Blair) or switching to High-Frame cluster (i.e., Steven) seemed unable to figure out how to use the system until the end of the simulation session. These students' lower learning performance indicated that such patterns may be an indicator of struggles. The participants who tended to stay in Low-Frame cluster toward the end of the simulation showed better learning gains in both exponential and transfer knowledge (i.e., Louise, Rhiannon). Understanding such cluster membership transfer reveals the trend of each student's gesture use across five different tasks. This may suggest ideal gesture behavior that leads to a positive learning outcome, which needs further research to verify the relationship.

5. Discussion

Recent research has highlighted students' challenges in understanding complex and abstract STEM concepts and the role that embodiment can play in overcoming these challenges (e.g., Duijzer et al., 2019; Stieff et al., 2016). ELASTIC^{3}S is an XR environment that was designed to facilitate crosscutting concept knowledge-building (scale, proportion, and quantity) by having students develop gestures to express ideas about linear/non-linear growth in different science domains. In this study, we were particularly interested in understanding embodied interaction data collected from an immersive XR platform as students were engaged in learning about non-linear relationships in one specific science domain (i.e., earthquakes). This study employed the DTW clustering method to explore and better understand students' embodied learning processes based on the fine-grained time-series gesture data recorded from the Microsoft Kinect.

It is important to explore different data granularities as they can reveal meaningful patterns in different contexts (e.g., Shen & Chi, 2017). To best describe the embodied learning experiences in this simulation, we first explored different types of data granularity: frame vs. gesture. The frame granularity did not tell us as much information as we were interested in the gesture as a whole, rather than a single frame; therefore, the gesture granularity yielded more interpretable results. We further explored different levels of analyses: (1) task, (2) subsequence, and (3) all tasks, to discover the best data windowing techniques for sequential patterns of embodied learning. This highlights the importance of data exploration to best tell the story in a certain context, and contributes on the literature of understanding embodied learning process within technology-enhanced learning environments (e.g., Price et al., 2016), especially by applying different analytical applications of fine-grained time-series embodied interaction data.

In ELASTIC^{3}S, students complete a series of five tasks with the goal of reaching a specific value on the Richter scale. Moving through the tasks, both students' gesture use and their conceptual understandings are advanced as they become more familiar with the system. The subsequence level analyses revealed more meaningful temporal patterns than task level analyses, as their captured behavior in a given task was placed in context with previous tasks. Further, any changes that occurred at each segment were certainly influenced by performance the time before.

The results indicate that certain gestural interaction trends of students' simulation use identified in Mid-Frame clusters (e.g., a decreasing trend of time spent on gesture multiplication, relatively lower time spent on gesture subtraction over the entire segments, relatively higher average volume over the last three segments) seemed to be associated with higher learning gains. This suggests their increasing trend of volume toward the later segments may indicate the students' higher confidence of using gesture as they become more familiar with the system and the exponential growth concepts, which needs further research to verify such relationship. The characteristics of a High-Frame cluster including the relationship with the learning gain measures showed that students exhibiting such characteristics (e.g., an increasing trend of time spent on gesture multiplication, relatively lower average volume over the last three segments) seemed less likely to learn the core ideas presented in the simulation. The results of the M2/3/3.5/7 segment, showed significant transfer knowledge gain differences between clusters. This may be an indicator of the needs of guidance for some students who showed the characteristics of High-Frame cluster based on their simulation use up to M7. For example, the system can track each individual's subsequence data aggregated from M2 to M7. If a student shows relatively high speed, longer time spent on gesture subtraction or gesture multiplication, or smaller volume, the system may provide certain prompts so that students can reflect on what they have done up to the current task. In this way, students will be able to receive more practice. This is aligned with the literature that highlights the importance of capturing the subtle change of students' multimodal interaction to understand their learning process and further yield positive learning performance (Jewitt, 2006; Ochoa, 2017). It is noteworthy that our sample size is relatively small (n=12) for making predictions that generalize to other groups of students. A larger sample is needed to validate that such patterns are representative of the behaviors of a larger group.

The subsequence level analyses overall tracked different gestural interaction patterns during each task within the context of past tasks and gestures, suggesting which task the participants may have struggled in completing. We further examined each individual student's cluster membership transfer (all tasks level) indicating an individual's sequential pattern of simulation use. The students who switched to Low-Frame cluster toward to the last task tended to achieve more learning. However, the students who exhibited the tendency to backtrack (i.e., the longest time spent on gesture subtraction) and struggled with understanding non-linear contexts (i.e., the longest time spent on gesture addition) showed lower conceptual gain. This suggests the clustering transfer trend might be an indicator of more support or practice being needed. These findings are descriptive in nature, and require further investigation for causal inferences between gestural patterns and learning performance.

Research (e.g., Abrahamson & Lindgren, 2014; DeSutter & Stieff, 2017) have identified components (e.g., activities, materials, and facilitation) of embodied learning environments and design principles that can be applied to the learning environment design. In particular, facilitation should be an integral part of the embodied learning environment, in which learning is facilitated by situated and timely feedback. Since body cueing is one of the ways embodied interactions are prompted around learning content, it is a way to integrate students' understanding of new knowledge with an embodied simulation (Lindgren, 2014). Therefore, it is critical to track students' interactions with the learning environment and to design effective cues that facilitate productive whole-body movements (Black et al., 2012; Lindgren & Johnson-Glenberg, 2013).

Overall, certain temporal patterns using different simulation use measures can be used for early detection of students' struggles throughout the process of exploring embodied activities within an XR learning environment such as ELASTIC^{3}S. Advanced adaptive learning technologies may help to engage the learners with personalized prompts based on kinematic markers to enhance students' cognitive activities in the process of learning. Most of the current studies in the area of embodied learning have been conducted in a laboratory environment (e.g., Lindgren & Johnson-Glenberg, 2013), where a researcher or teacher is present throughout the learning activities and provides a participant with just-in-time guidance whenever needed. Exploring potential features driven from multimodal data is critical in the future implementation of embodied learning environment in either a formal or informal setting (e.g., Ochoa, 2017).

6. Limitations and future work

This study has limitations that should be addressed in future research. First, the number of participants is notably small (n = 12). However, the Kinect logfiles contain a massive gesture dataset for each participant. This yielded the adequate amount of data (see Table 4) that represents valid and reliable behaviors of the participants, which allowed for the analyses we conducted in this study. In future work, we plan to include a larger sample to validate (1) whether the captured patterns are representative of the behaviors of a larger group and (2) the relationships between the patterns and learning performance. Second, it should be noted that we did not collect the participants' academic background such as fields of study or previous experiences in STEM, while we recruited the participants from a general educational psychology course where the majority of students' fields of study was not STEM.

Third, we identified the cluster patterns including volume and speed features, which suggested their affective states such as confidence of using gesture rather than cognitive states. This needs further research to verify such a relationship. This also suggests interesting lines of follow-up inquiry on the causal relationship between kinematic features or gesture pattern trajectories and affective or cognitive states. Lastly, we extracted the simulation use variables by considering all joint information in one frame. There is a lot more nuance to 3-dimensional metrics especially in spatial positions. Therefore, future studies should further explore other features, such as focusing on one joint for all frames or all joints and all frames, as each new feature will return different metrics and representations. We hope the volume features we used in this study are a good starting point for further exploration. The analytical approach used in this study indicates the potential of kinematic features as key indicators of the quality of learner perceptions and comprehension, and the potential need for gestural interaction guidance, which can further support their learning in other domains.

7. Conclusion

Previous studies have shown that the embodied learning simulation, ELASTIC^{3}S, has an overall a positive impact on students' understanding of content objectives and the crosscutting concept of non-linear growth. Those studies highlighted the critical features of embodied simulations that facilitate student reasoning. Technological advances in gesture recognition allow for the creation of XR environments that can track and respond to students' gestures in real time. This study therefore investigated how the gestures learners perform in these environments relate to their subsequent learning, and how understanding the features of productive gestures can be applied to future embodied XR learning environment design. We applied multivariate Dynamic Time Warping for clustering to identify gestural patterns in ELASTIC^{3}S as evidence for understanding learning processes. Our findings showed that identified patterns of simulation use were indicative of students' comprehension and struggles with learning target ideas.

The main contribution of this paper is to apply an underutilized analytical approach to understand students' gestural interactions with embodied XR learning environments. Specifically, this study contributes to the early

work of detecting sequential gesture patterns that represent students' embodied learning experiences by applying different data granularities (frame vs. gesture) and different levels of analysis (task, subsequence, and all tasks). Different levels of analyses applied in this study highlights the importance of considering various ways of structuring data, which can reveal more meaningful patterns of simulation use and serves as evidence of a more positive embodied learning experience. The results of this study pave the way for further research on the design of XR embodied simulation environments that provide real-time guidance and promote a more powerful and adaptive learning experience.

We proposed potential kinematic features for personalized feedback that may facilitate students' productive whole-body movements and learning. It is worth noting that future research including lager samples is needed to validate the present findings. We believe such analytical applications explored in this paper provide guidance for researchers to replicate or adapt when dealing with fine-grained time-series gesture data within XR embodied environments.

References

Abrahamson, D., & Lindgren, R. (2014). Embodiment and embodied design. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (2nd ed., pp. 358–376). Cambridge, UK: Cambridge University Press.

Alameh, S., Morphew, J., Mathayas, N., & Lindgren, R. (2016). Exploring the relationship between gesture and student reasoning regarding linear and exponential growth. In the *Proceedings of the 12th International Conference of the Learning Sciences* (pp. 1006–1009). Singapore: International Society of the Learning Sciences.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*(1), 617–645. doi:10.1146/annurev.psych.59.103006.093639

Birchfield, D., Thornburg, H., Megowan-Romanowicz, M. C., Hatton, S., Mechtley, B., Dolgov, I., & Burleson, W. (2008). Embodiment, multimodality, and composition: Convergent themes across HCI and education for mixed-reality learning environments. *Advances in Human-Computer Interaction*. doi:10.1155/2008/874563

Black, J., Segal, A., Vitale, J., & Fadjo, C. (2012). Embodied cognition and learning environment design. In D. Jonassen, & S. Land (Eds.), *Theoretical foundations of student-centered learning environments* (2nd ed., pp. 198–223). New York, NY: Routledge.

Crowder, E. M. (1996). Gestures at work in sense-making science talk. *Journal of the Learning Sciences, 5*(3), 173–208.

DeSutter, D., & Stieff, M. (2017). Teaching students to think spatially through embodied actions: Design principles for learning environments in science, technology, engineering, and mathematics. *Cognitive Research: Principles and Implications, 2*(1), 22. doi:10.1186/s41235-016-0039-y

Duijzer, C., Van den Heuvel-Panhuizen, M., Veldhuis, M., Doorman, M., & Leseman, P. (2019). Embodied learning environments for graphing motion: A Systematic literature review. *Educational Psychology Review, 31*(3), 597-629.

Gallagher, S. (2006). *How the body shapes the mind*. New York, NY: Clarendon Press.

Gallagher, S., & Lindgren, R. (2015). Enactive metaphors: Learning through full-body engagement. *Educational Psychology Review, 27*(3), 391–404.

Genolini, C., & Falissard, B. (2011). Kml: A Package to cluster longitudinal data. *Computational Methods Programs Biomed, 104*(3), 112-121.

Glenberg, A. M. (2008). Embodiment for education. In P. Calvo & A. Gomila (Eds.), *Handbook of Cognitive Science* (pp. 355–372). doi:10.1016/B978-0-08-046616-3.00018-9

Glenberg, A. M. (2010). Embodiment as a unifying perspective for psychology. *Wires Cognitive Science, 1*(4), 586–596.

Goldin-Meadow, S. (2011). Learning through gesture. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(6), 595-607.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A Six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64–74. doi:10.1119/1.18809

Han, I., & Black, J. B. (2011). Incorporating haptic feedback in simulation for learning physics. *Computers & Education, 57*(4), 2281–2290.

Jewitt, C. (2006). *Technology, literacy and learning: A Multimodal approach*. New York, NY: Routledge.

Johnson-Glenberg, M. C., Birchfield, D., Savvides, P., & Megowan-Romanowicz, C. (2011). Semi-virtual embodied learning-real world stem assessment. In L. Annetta & S. C. Bronack (Eds.), *Serious educational game assessment: Practical*

methods and models for educational games, simulations and virtual worlds (pp. 241–257). Rotterdam, The Netherlands: Sense Publishers.

Johnson-Glenberg, M. C., Birchfield, D. A., Tolentino, L., & Koziupa, T. (2014). Collaborative embodied learning in mixed reality motion-capture environments: Two science studies. *Journal of Educational Psychology, 106*(1), 86–104.

Johnson-Glenberg, M. C., & Megowan-Romanowicz, C. (2017). Embodied science and mixed reality: How gesture and motion capture affect physics education. *Cognitive Research: Principles and Implications, 2*(1), 24. doi:10.1186/s41235-017-0060-9

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.

Kang, J., Lindgren, R., & Planey, J. (2018). Exploring emergent features of student interaction within an embodied science learning simulation. *Multimodal Technologies and Interaction, 2*(3), 39.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association, 47*(260), 583–621.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Li, Y., Chen, H., & Wu, Z. (2010). Dynamic time warping distance method for similarity test of multipoint ground motion field. *Mathematical Problems in Engineering*. doi:10.1155/2010/749517

Lin, J., Williamson, S., Borne, K. D., & DeBarr, D. (2012). Pattern recognition in time series. In *Advances in Machine Learning and Data Mining for Astronomy* (pp. 617–645). New York, NY: CRC Press.

Lindgren, R., & Johnson-Glenberg, M. (2013). Emboldened by embodiment: Six precepts for research on embodied learning and mixed reality. *Educational Researcher*. doi:10.3102/0013189X13511661

Lindgren, R. (2014). Getting into the cue: Embracing technology-facilitated body movements as a starting point for learning. In *Learning Technologies and the Body* (pp. 51–66). doi:10.4324/9781315772639-9

Lindgren, R., Tscholl, M., Wang, S., & Johnson, E. (2016). Enhancing learning and engagement through embodied interaction within a mixed reality simulation. *Computers & Education, 95*, 174–187.

Mathayas, N., Brown, D. E., Wallon, R. C., & Lindgren, R. (2019). Representational gesturing as an epistemic tool for the development of mechanistic explanatory models. *Science Education, 103*(4), 1047–1079.

Mezari, A., & Maglogiannis, I. (2017). Gesture recognition using symbolic aggregate approximation and dynamic time warping on motion data. In *the Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare* (pp. 342–347). Barcelona, Spain: ACM. doi:10.1145/3154862.3154927

Nathan, M. J., & Walkington, C. (2017). Grounded and embodied mathematical cognition: Promoting mathematical insight and proof using action and language. *Cognitive Research: Principles and Implications, 2*(1), 9. doi:10.1186/s41235-016-0040-5

Noice, H., & Noice, T. (2006). What studies of actors and acting can tell us about memory and cognitive functioning. *Current Directions in Psychological Science, 15*(1), 14–18.

Ochoa, X. (2017). Multimodal learning analytics. In *The Handbook of Learning Analytics* (pp. 129–141). doi:10.18608/hla17

Price, S., Jewitt, C., & Sakr, M. (2016). Embodied experiences of place: a study of history learning with mobile technologies. *Journal of Computer Assisted Learning, 32*(4), 345–359.

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.

Segal, A., Tversky, B., & Black, J. (2014). Conceptually congruent actions can promote thought. *Journal of Applied Research in Memory and Cognition, 3*(3), 124–130.

Shapiro, L. (2019). *Embodied cognition*. New York, NY: Routledge.

Shen, S., & Chi, M. (2017, June). Clustering student sequential trajectories using dynamic time warping. In *the Proceedings of 10th the International Conference on Educational Data Mining (EDM)* (pp. 266–271), Wuhan, China: International Educational Data Mining Society.

Skulmowski, A., & Rey, G. D. (2018). Embodied learning: introducing a taxonomy based on bodily engagement and task integration. *Cognitive research: principles and implications, 3*(1), 6. doi:10.1186/s41235-018-0092-9

Smith, C., King, B., & Gonzalez, D. (2016). Using multimodal learning analytics to identify patterns of interactions in a body-based mathematics activity. *Journal of Interactive Learning Research, 27*(4), 355–379.

Stieff, M., Lira, M. E., & Scopelitis, S. A. (2016). Gesture supports spatial thinking in STEM. *Cognition and Instruction, 34*(2), 80–99.

Tretter, T. R., Jones, M. G., Andre, T., Negishi, A., & Minogue, J. (2006). Conceptual boundaries and distances: Students' and experts' concepts of the scale of scientific phenomena. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 43*(3), 282–319.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic bulletin & review, 9*(4), 625–636.

Appendix A: Sample questions from pre- and post-tests

Conceptual Knowledge
- What causes earthquakes?

Exponential Knowledge
- What happens to the amount of damage in a town if the earthquake goes from a magnitude of 7.2 to a magnitude of 8.2?

Transfer Knowledge
- I want you to imagine that you McDonald's and Burger King are going to start opening restaurants in China. McDonald's plans to open 3,000 restaurants every year for 12 years. Burger King is going to start with 1 restaurant and then triple the number of restaurants every year for 12 years. Which restaurant chain do you think will have the most restaurants in 12 years?