Be the Data: An Embodied Experience for Data Analytics

Xin Chen, Virginia Polytechnic Institute and State University;
Leanna L. House, Virginia Polytechnic Institute and State University;
Jessica Zeitz Self, Virginia Polytechnic Institute and State University;
Scotland C. Leman, Virginia Polytechnic Institute and State University;
Jane Robertson Evia, Virginia Polytechnic Institute and State University;
James Thomas Fry, Virginia Polytechnic Institute and State University;
Chris North, Virginia Polytechnic Institute and State University

Session Title: Learning in Immersive Environments: Science, Statistics, Assessments, and Learner Characteristics

Presentation Date 4/11/2016

Presentation Location Washington, D.C.

Descriptors: Collaboration, Computer Applications, Learning Environments

Unit: SIG-Applied Research in Immersive Environments for Learning

Paper presented at the 2016 annual meeting of the American Educational Research Association.

Abstract

With the rise of big data, it is becoming increasingly important to educate students about data analytics. In particular, many students find it challenging to understand high-dimensional data and learn relevant analytical techniques, such as dimension reduction. This paper describes an embodied approach to teach students exploring alternative 2D projections of high-dimensional data. Specifically, we present a novel approach, *Be the Data*. In a physical room, each student embodies a data point in a high-dimensional dataset. Coordinates of students' positions in the room represents the two-dimensional plane to which the high-dimensional data are projected. Students physically move in a room with respect to others to interact with alternative projections of the high-dimensional data. We presented *Be the Data* during a computer science outreach day for 7th grade students. Our results from recorded videos and surveys suggest that *Be the Data* successfully provided an effective and engaging experience for learning data analytics.

Key words: embodiment, immersive environment, collaboration, high-dimensional data

Be the Data: An embodied experience for data analytics

1. Introduction

As today's data is becoming more and more complex, there is a clear need to advance education in knowledge discovery from big data. STEM educators are called upon to attract students to learn data analytical techniques. However, it is difficult to teach data analytics to students. Students not experienced in analyzing data normally do not have necessary mathematical background to understand the underlying analytical mathematical model. And because of the lack of prior math knowledge, students tend to have an unenthusiastic attitude to learn data analysis (Valero-Mora & Ledesma, 2011). Moreover, data analysis is a complex and critical thinking process that extends beyond the comprehension of mathematical models. It includes formalizing alternative hypotheses, exploring multiple solutions, and assessing implications of discoveries (Sun, M., Bradel, L., North, C. L., & Ramakrishnan, N., 2014). Unfortunately, current approaches in teaching data analytics primarily focus on its quantitative aspects while overlook the critical analytical processes.

Our work is inspired by embodied mathematical cognition perspectives informed by Lakoff (Lakoff & Núñez, 2000). He emphasized that abstract conceptual knowledge is "*embodied*" and "*mapped within our sensory-motor system*". He further suggested to allow the understanding of mathematics "*emerges through the interaction of the mind with the world*". Immersive interface is gaining popularity as a means to engage students and help them develop an understanding of the abstract from the concrete experience (Bell, Savin-Baden, & Ward, 2011; Jackson & Fagan, 2000; Lucke, 2011; North, 2014). In addition, the immersive learning environment is attractive for its fun (Wagner, 2008). Therefore, it is particularly suitable for teaching and learning difficult abstract knowledge (Beckem & Watkins, 2012).

The goal of this research is to explore a new embodied learning approach via a novel immersive system, called *Be the Data* (Chen, Self, House, & North, 2016). In our system, each student, in a physical space, embodies a unique data point in a high-dimensional dataset. The positions of students in this physical space represents a 2D projection of the high-dimensional data. Students can explore alternative projections by physically moving themselves, and hence the corresponding data points, in the space. Therefore, students are able to pose hypotheses about data, explore data and test their hypotheses by interacting with each other in the room.

Be the Data is an extension of a point and click software called Andromeda that we developed previously to enable professional analysts to explore high-dimensional data (Self, Self, House, Leman, & North, 2014). Using a visualization technique called Weight Multidimensional Scaling (WMDS) (Kruskal & Wish, 1978), Andromeda responds to user feedback that is communicated by clicking and dragging data points and creates multiple two-dimensional visualizations of the same dataset. Andromeda shields users from the technicalities of WMDS so that users may focus on exploring and learning from data. However, the point and click platform of Andromeda has educational limitations in that users must have advanced cognitive skills to conceptualize an abstract mapping from the data to the visualizations for interpreting them. To help students with the conceptualization, we employed *Be the Data* (Chen et al., 2016).). It combines the physical and virtual worlds so that students may experience the mapping on a personal level and construct their understanding about data analytics.

By applying novel immersive technology, this study is one of few pioneer outreach to young children in learning data analytics - an essential skill in STEM. In this paper, we present *Be the Data* in detail and answer the following research questions:

1. Did students learn WMDS associated concepts, including dimension reduction, relative distance, variable, and data exploration?
2. Did students produce a positive attitude towards *Be the Data*?
3. What learning strategies students took to learn about data and analytical process?

2. Literature Review

2.1 Immersive Interface

With the technological breakthrough in virtual reality and mixed reality, the immersive interface is developed as the combination of real and virtual that immerses users in real-time natural interactions, enhancing the sense of users actually being within it (Azuma, 1995; Bailenson, Yee, Blascovich, Beall, Lundblad, & Jin, 2008). By extending interaction beyond the traditional desktop and into the real environment, the immersive environment opens up tremendous opportunities to facilitate learning through embodied interaction: creating, manipulating, and sharing meaning through interacting in a physical and social environment (Dourish, 2004).

2.2 Embodied Learning

Embodied interaction has been increasingly applied in the education arena, especially for teaching and learning in abstract problem domains. Moso tangibles are a set of interactive physical artifacts for children to learn abstract music concepts (i.e., pitch, volume, and tempo) (Bakker, van den Hoven, & Antle, 2011). With an ongoing tone, students could point a pitch artifact upward or downward to make the pitch higher or lower, squeeze a volume artifact wildly or quietly to make the volume louder softer, and shake a tempo artifact fast or slowly to make the tempo faster or slower. The study found that while not every child was able to verbally express their understanding of the targeted concepts, all of them could reproduce sound examples with required pitch, volume and tempo by interacting with the artifacts. It indicated that children can understand these abstract concepts in terms of their familiar movement related concepts (i.e., low/high, quietly/wildly, slow/fast). It also indicated that students were able to reason about these concepts using proper movement rather than words.

Embodied learning is particularly applicable to mathematical education because the understanding of mathematical concepts is rooted in embodied interaction (Lakoff & N´u˜nez, 2000). Since childhood, people are encouraged to play with building blocks to comprehend integers, use fingers to learn how to count, and slid beads on an abacus to gain a symbolic understanding of "addition" and "deduction" concepts. Howison and his colleagues (Howison, Trninic, Reinholz, & Abrahamson, 2011) studied how to apply body movement to help young students understand proportional equivalence (e.g., $\frac{2}{3} = \frac{4}{6}$). In their study, students were asked to position two arms above a desk so that the distance between the arms and the desk could make and maintain a certain proportion. For example, to maintain a $\frac{1}{2}$ ratio, for every 1 units of distance on the left arm, it's 2 units of distance on the right arm (e.g., $\frac{1}{2} = \frac{2}{4} = \frac{3}{6}$). With a natural perception of relative distance between two arms and desk, all the participants succeeded in getting correct proportions. While arm-movement has not yet been exploited in teaching and learning mathematics, this study provides empirical evidence that this embodied interaction is able to evoke basic arithmetic operations to understand proportional equivalence.

Embodiment is important to cognitive activities as our mind and body are deeply integrated. People unconsciously applied their bodily experience, such as movement (Ball, North, & Bowman, 2007), gesturing (Bakker et al., 2011), and spatial relationship (Howison et al., 2011; Martin & Schwartz, 2005) to facilitate their cognitive process. Embodiment provides learners with a concrete physical metaphor (e.g., up-down, near-far, slow-fast) to reason about the targeted abstract concepts. After the concepts become grounded in embodied action, learners are able to apply the new interpretation to re-purpose many environments to solve similar problems in other contexts (Martin & Schwartz, 2005).

2.3 Collaborative Data Analysis

Data analysis is a complex thinking process. Collaborative work can be extremely beneficial for data analytics tasks (Heer & Agrawala, 2008; Isenberg, Fisher, Paul, Morris, Inkpen, & Czerwinski, 2012; Keel, 2006). In an exploratory study of a digital tabletop to support collaborative visual analytics tasks (Isenberg et al., 2012), researchers found that task success was highly correlated with the time spent working collaboratively. From the perspective of communities of practice (Wegner, 1998), learning data analysis benefits from a collaborative environment where students converse, interact, and collaborate to negotiate alternative hypotheses, judgment, and solutions. The interactive nature of immersive interface lends itself to constructivist approaches to instruction, position students in the center of the learning process, and promote collaboration among students (Bergero, Hargreaves, & Nichols, 2012; Chen, Fang, & Lockee, 2015; Dede, 2009; Lucke, 2011; Peer & Giachritsis, 2012).

3. Our System: *Be the Data*

Be the Data system was developed to take advantages of embodied learning and collaborative environment. Table 1 shows a summary of design considerations and corresponding system features.

We exploited a unique new physical space, the multi-media Cube at our residing university. The Cube is comprised of advanced interactive technologies for physical-virtual cross-overs, including a large display, a tracking system, and a software package for direct manipulation of virtual high-dimensional data models (Figure 1a).

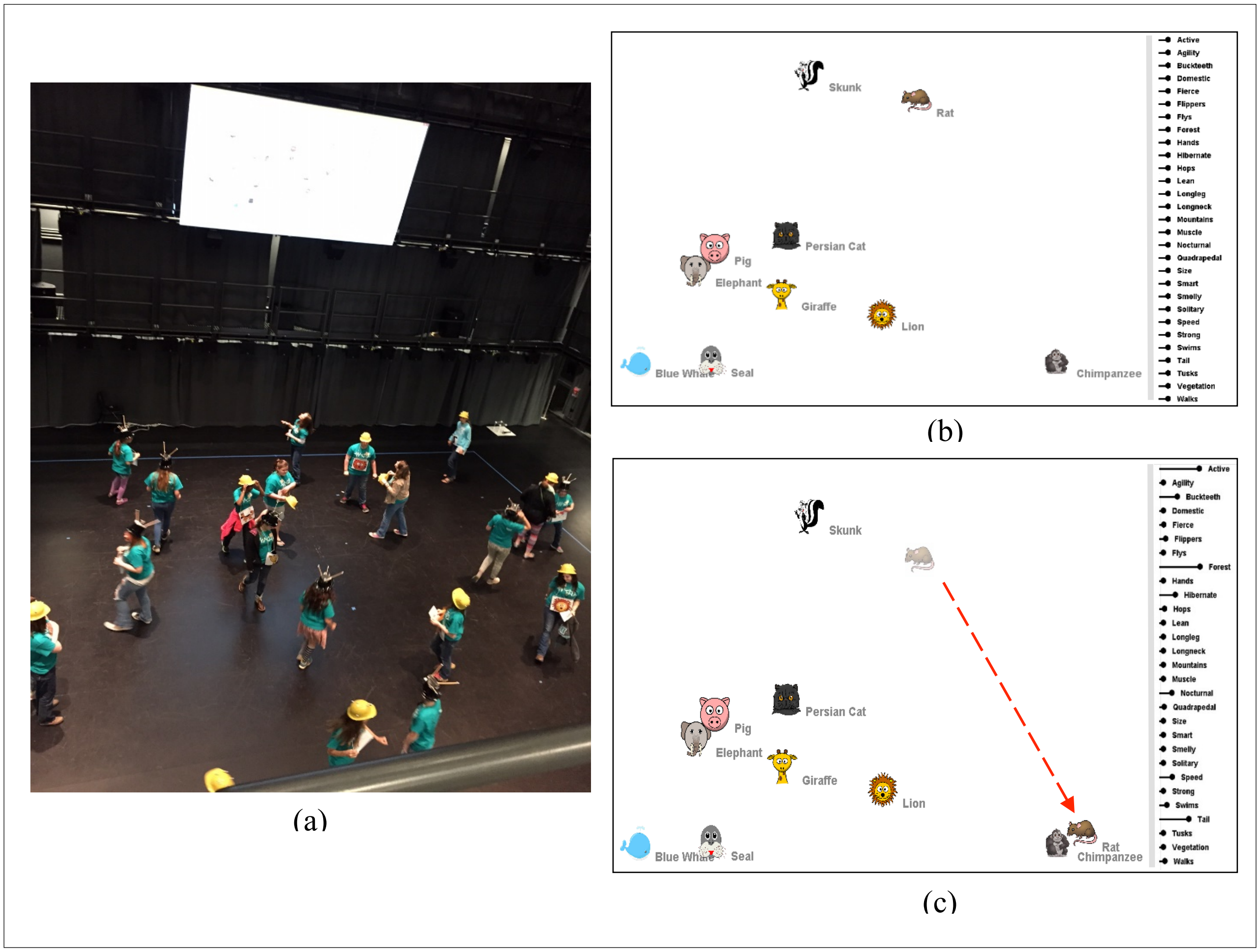

Figure 1. (a) In *Be the Data*, students become the data points and a birds-eye views of their locations in the room are displayed above them. (b) A clear image of the overhead display. (c) When students move, they are changing which dimensions to emphasize. *Be the Data* communicates this by change the lengths of bars next to the dimensions.

For *Be the Data*, students enter the Cube and embody virtual data points by wearing trackable hats for motion detection. For example, if we consider a high-dimensional dataset about animals (Figure 2), each student is an animal and her location is displayed or visualized above her (Figure 1a). To interpret the visualization, near implies relatively similar and far implies relatively different in the variables that are emphasized (i.e., those that have more weight than others). Relative weights are shown alongside the visualization on the right (Figure 1b). Variables with long bars have more weight in the visualization then variables with short bars. As animals move in the room, weights on the variables change, as demonstrated in Figures 1b and 1c. The rat moves closer to the chimpanzee in Figure 1c, thus the rat is now considered more similar to a chimpanzee than the remaining animals in the variables that are now emphasized, which are Active, Forest, Hibernate, Nocturnal, and Tail.

Crucially, students do not need to know their exact coordinates or weight values to interpret the visualization. They simply need to assess the relative distance (i.e., who is near or far from them and the relative weights (i.e. which weight bar is longer or shorter). Thus, the students continuously identify where they are in the room and the meaning of proximity to others with respect to up-weighted variables.

Table 1.

Design considerations matched with system features

Design Considerations	**Supportive System Features**
Body movement as a type of embodied interaction.	• User's whole body as the input to the system; • Immerse user as a movable data point; • Real time visual feedback from the system;
Familiar metaphors: 1) "Near is similar; Far is different" to map the abstract WMDS algorithm. 2) "Longer is more; Shorter is less" to map the abstract variable weights.	• Physical embodied interaction; • Interactive visualization;
Collaborative data analysis.	• Simultaneously tracking multiple users; • A large shared display; • Co-located in a physical space.

Name	Walks	Vegetatior	Tusks	Tail	Swims	Strong	Spots	Speed	Solitary	Smelly
Persian Cat	65.69	6.25	0	66.8	6.25	12.58	6.25	26.98	36.6	7.86
Horse	55.58	51.05	0	70.42	0	69.13	15.8	81.68	16.78	33.07
Blue Whale	0	0	0	26.42	71.82	55.26	23.75	21.42	25.99	13.75
Skunk	64.86	44.38	0	83.33	8.33	3.12	1.25	30.21	47.85	100

Figure 2. A screenshot of a portion of the animal dataset (20 animals X 31 variables). It shows 4 animals (the 1^{st} column) and 10 variables (from the 2^{nd} column to the last column).

4. Methods

We implemented four 55-minute educational workshops for STEM outreach activities hosted at our institution. Sixty-two 7^{th} grade students from southwestern Virginia participated in the event. All the participants were new to the MDS algorithm as the 7^{th} curriculum had not covered data analytics related concepts. In groups of 15-20, students explored the animal dataset (Figure 2) using our system.

The workshop started with a short introduction on what is high dimensional data by using the example shown in Figure 2. The instructor further explained how to use the system. After the introduction, students performed group-based analytical tasks to learn from the data. Students were encouraged to use the system to answer any question about this animal data. When they failed to come up with their own questions, the instructor suggested one for them to solve.

To assess the success of the workshop, we collected video data and post-workshop surveys. Digital videos were captured to document the workshops' execution. These videos were deemed to be discreet enough to limit visibility to the participants, but still allowed researchers to investigate class activities. We had two authors watched the replay and identified collaborative strategies students used. The surveys had two parts: multiple choice and open-ended questions. We used simple summary statistics to analyze the quantitative data from the surveys. When applicable, we report the percent of students with normal approximations of the standard deviation (*sd*) in parentheses. We had two authors qualitatively analyzed free responses to open-ended questions

5. Results

5.1 Did students learn WMDS associated concepts, including dimension reduction, relative distance, variable, and data exploration?

Sixty of the 62 participants ($N = 60$) completed the post surveys. To assess their understanding of dimension reduction, we asked how many dimensions are being displayed in a dimension reduction plot, and 78% ($sd = 5\%$) of the students answered correctly. To assess their understanding of relative distance, we asked them to interpret similarities of data points in a dimension reduction visualization based on relative distances, and 92% ($sd = 4\%$) of students answered correctly. However, only 42% ($sd = 6\%$) answered correctly the question that asked explicitly for an example of a variable. This arguably low percent could be reflective of the workshop or the need for some minor improvements in terminology or phrasing in the survey, as many students used the term "variable" correctly in a sentence. For example, in response to "What did you learn?" students said, "There are certain variables that are weighted more heavily", "How to work with multiple variables", and "variables are more than numbers".

The concept of data exploration is open ended. It is a process that typically relies on numerical and visual summaries of data to discover features and patterns. We feel that many students developed their own idea of data exploration (e.g., "The same data can be used to answer many different questions", "The same data can be organized in different ways", "I learned how to compare data", "There are a lot of different ways to group stuff").

5.2 Did students produce a positive attitude towards *Be the Data*?

In 60 responses to the question about students' experiences, 32 students said they liked moving around, 25 students said they liked the animal dataset, and 15 said they liked working in teams. One student wrote "Data can be fun, not always tedious and boring."

Enhanced engagement was also observed during the workshop. Children excelled when they were provided opportunities to move. They laughed, whistled, and made other exclamations throughout the workshop. Some students gave 'high-fives' when they collaboratively solved the given tasks. One of the middle school teachers watched his class and commented, "I have never seen my students being so engaged."

5.3 What learning strategies students took to learn about data and analytical process?

Video recordings of our workshops suggest that students used collaborative learning strategies to develop their understanding of data and data analytics skills. We identified five key components of students' learning experience that evolved as the instructor's guidance decreased (Figure 3).

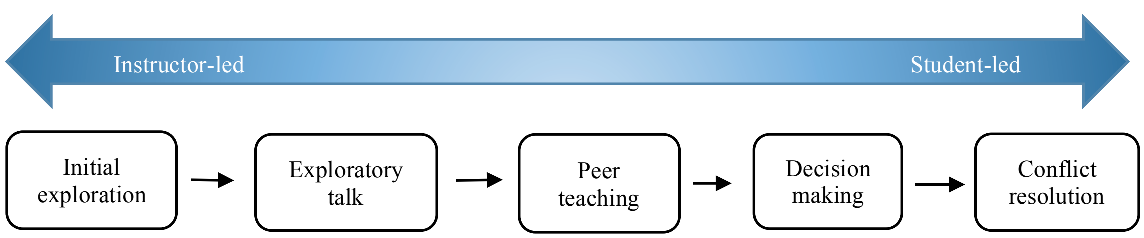

Figure 3. Key components of the immersive learning with *Be the Data*.

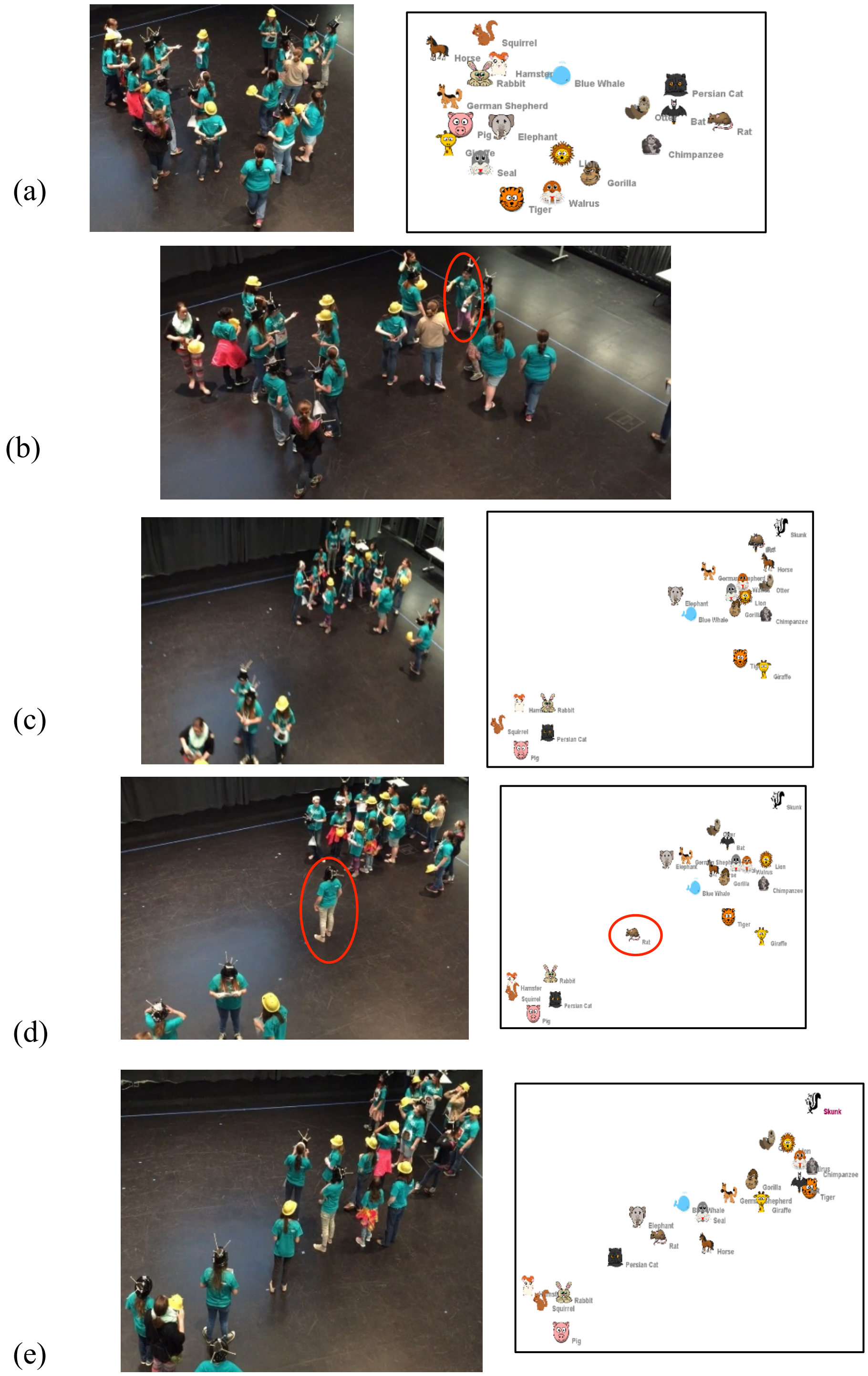

Figure 4. Collaborative strategies for learning high-dimensional data: (a) Exploratory talk. (b) Peer teaching as a student led the group. (c) Students moved to form two clusters. (d) Conflicts arose. (e) Conflict resolution.

Initial exploration. Initial exploration occurred when students experienced the system for the first time and learned how to use it. Under the guidance of the instructor, students moved around the Cube and watched the large display to understand how the visualization adjusted based upon their movement.

Exploratory talk. As students learned more about the system, one student said that she could use it to “compare different things”. At times, students moved intentionally closer to (or away from) others to assess changes in variable weights.

Peer teaching. As students got to know the system, they started to work on their specific questions. For example, to answer the question “What makes some animals good to eat?”, they first came together to discuss (Figure 4a). Several students offered their own opinions for group consideration. Some students suggested looking at variables related to tasty in the dataset. Some students discussed the relationship between Muscle, Vegetation, Lean, etc. Co-construction of shared understanding was manifested in these conversations. As students heard multiple perspectives from others, their own ideas were sometimes challenged. A few students gradually took the dominant role and started to direct the movement of other students. For example, student A (as noted in Figure 4b) raised her hand and said loudly, “So less edible animals move here, more edible animals move there.” Data clusters began to emerge.

Decision making. One unique feature of our system is that every participant’s input matters. Therefore, students had to work together to find a solution. While some students expressed uncertainty about student A’s suggestion, remaining students confirmed her ideas. After negotiation, all students accepted Student A’s suggestion to move into two groups: edible and non-edible (Figure 4c).

Conflict resolution. Sometimes, there were conflicting student opinions. For example, one student felt that she did not belong to either of the groups. Therefore, she stood between the two (Figure 4d). Her explanation was although rat was normally not edible, it might be a dish in certain cultures. More justification and alternative hypotheses were collaboratively made. Real-time feedback from the system allowed constant opportunities for students to test their guesses, elaborate, and reflect on their thoughts. Finally, all students arrived at a desired agreed-upon answer (Figure 4e). They found that buckteeth, domestic, and smelly were variables that mostly mattered.

Discussion

We aim to explore students' embodied learning experience with *Be the Data*. The present study provides evidence that *Be The Data* offered an effective learning experience associated with play and fun.

Be the Data serves as an intuitive medium for students to understand abstract analytical methods. Our system capitalizes upon humans' familiarity with spatial organization in differing similarity versus difference. The embodied near-far metaphor is familiar to students as young as 7^{th} grade and matches the conceptualization of underlying mathematical model. Therefore, after embodying a data point, students were good at conjecturing various relationships of data despite their inexperience with the mathematical methods.

Be the Data facilitates a collaborative environment to engage students. The interactive feature of the system invited students to work together in a student-centric environment. They played an active role in collaboratively producing the final product that reflects the ideas and opinions of the group as a whole. Critical thinking demonstrated in their collaborative negotiation is an essential skill for data analysis.

Our findings suggest that *Be the Data* is worth investigating further. However, there is opportunity for improvement. The findings are based on an observational study. Future work requires more rigorous analyses of *Be the Data* to quantify what students gain from interacting with *Be The Data*. Also, we wish to examine the impact of different levels of embodiment and specific teaching approaches applicable to *Be the Data*.

Acknowledgements

This research was partially supported by NSF grants IIS-1447416, IIS-1218346, and DUE-1141096; and a seed grant from Virginia Tech's Institute for Creativity, Arts, and Technology (ICAT).

References

Azuma, R. (1995). A survey of augmented reality. *Presence, 6*(4), 355-385.

Bailenson, J. N., Yee, N., Blascovich, J., Beall, A. C., Lundblad, N., & Jin, M. (2008). The use of immersive virtual reality in the learning sciences: digital transformations of teachers, students, and social context. *The Journal of the Learning Sciences, 17*(1), 102-141.

Bakker, S., van den Hoven, E., & Antle, A. N. (2011). MoSo tangibles: evaluating embodied learning. In *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction* (pp. 85-92). ACM.

Ball, R., North, C., & Bowman, D. A. (2007). Move to improve: promoting physical navigation to increase user performance with large displays. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 191-200). ACM.

Beckem, J., & Watkins, M. (2012). Bringing life to learning: immersive experiential learning simulations for online and blended courses. *Journal of Asynchronous Learning Networks, 16*(5), 61-70.

Bell, F., Savin-Baden, M., & Ward, R. (2011). Learning and teaching in immersive virtual worlds. *Research in Learning Technology, 16*(3).

Bergero, C., Hargreaves, L., & Nichols, A. (2012). Collaborative healthcare immersive learning dynamic: transitioning to simulation-based learning. *Clinical Nurse Specialist, 26*(1), 42-47.

Chen, X., Fang, Y., & Lockee, B. (2015). Integrative review of social presence in distance education: issues and challenges. *Educational Research and Reviews, 10*(13), 1796-1806.

Chen, X., Self, J. Z., House, L., & North, C. (2016). Be the data: a new approach for immersive analytics. In *Virtual Reality Workshop on Immersive Analytics.* (To appear)

Dede, C. (2009). Immersive interfaces for engagement and learning. *Science, 323*(5910), 66-69.

Dourish, P. (2004). *Dourish, Paul. Where the action is: the foundations of embodied interaction*. Cambridge, MA: MIT press.

Heer, J., & Agrawala, M. (2008). Design considerations for collaborative visual analytics. *Information visualization, 7*(1), 49-62.

Howison, M., Trninic, D., Reinholz, D., & Abrahamson, D. (2011). The Mathematical Imagery Trainer: from embodied interaction to conceptual learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1989-1998). ACM.

Isenberg, P., Fisher, D., Paul, S., Morris, M. R., Inkpen, K., & Czerwinski, M. (2012). Co-located collaborative visual analytics around a tabletop display. *Transactions on Visualization and Computer Graphics, 18*(5), 689-702.

Jackson, R. L., & Fagan, E. (2000). Collaboration and learning within immersive virtual reality. In *Proceedings of the third international conference on Collaborative Virtual Environments* (pp. 83-92). ACM.

Keel, P. E. (2006). Collaborative visual analytics: Inferring from the spatial organization and collaborative use of information. In *Proceedings of VIS Conference on Visual Analytics Science And Technology,* (pp. 137-144). IEEE.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional Scaling* (Vol. 11). Beverly Hills, Calif: Sage.

Lakoff, G., & Núñez, R. E. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being*. New York: Basic books.

Lucke, U. (2011). Immersive learning related to theory and practice of blended learning. In *World Conference on Educational Multimedia, Hypermedia and Telecommunications* (Vol. 2011, No. 1, pp. 3484-3487).

Martin, T., & Schwartz, D. L. (2005). Physically distributed learning: Adapting and reinterpreting physical environments in the development of fraction concepts. *Cognitive science, 29*(4), 587-625.

North, M. (2014). Distance learning immersive environments: Sense of presence exploration. *Journal of Visual Languages and Computing, 25*(6), 1040-1041.

Peer, A., & Giachritsis, C. D. (2012). *Immersive multimodal interactive presence*. New York: London: Springer.

Self, J. Z., Self, N., House, L., Leman, S., & North, C. (2014). Improving students' cognitive dimensionality through education with object-level interaction. Retrieved from http://people.cs.vt.edu/~jazeitz/jzself_vast2015_tech_report.pdf

Sun, M., Bradel, L., North, C. L., & Ramakrishnan, N. (2014, April). The role of interactive biclusters in sensemaking. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 1559-1562). ACM.

Valero-Mora, P. M., & Ledesma, R. D. (2011). Using interactive graphics to teach multivariate data analysis to psychology students. *Journal of Statistics Education, 19*(1), 1-19.

Wagner, C. J. o. I. S. E., 19(3), 263. (2008). Learning experience with virtual worlds. *Journal of Information Systems Education, 19*(3), 263-266.

Wegner, E. (1998). *Communities of practice: learning, meaning, and identity*. Cambridge, England: Cambridge University Press.