Trends in the Salience of Data Collected in a Multi User Virtual Environment: an Exploratory Study

Trends in the salience of data collected in a multi user virtual environment:

An exploratory study

M. Shane Tutwiler

Tina Grotzer
Chris Dede
Diane Jass Ketelhut

A Thesis Presented to the Faculty of the Graduate School of Education of Harvard University in Partial Fulfillment of the Requirements for the Degree of Doctor of Education

2014

Dedication

To my Mom and Dad, for helping me to understand that it was okay to be curious and ask (often endless) questions as a little boy. And, to my wife and daughter, without whose love and support I would not be able to continue to seek answers to those questions as a (mostly) grown-up. I love you all.

Acknowledgements

Despite copyright and authorship claims to the contrary, dissertations are never completed in a vacuum. As I steadily worked toward completing this thesis study over the course of the last five years, I was blessed with the best possible support networks. I am eternally grateful to my wife Jessica and daughter Abigail, who were always patient and understanding with the long nights I spent analyzing, editing, and revising this work. I am also lucky to have had the support of my brilliant and passionate friends from the 2009 doctoral cohort, including Monica Ng, Adrienne Keene, Stephen Hoffman, John McLaughlin, and William Johnston. In addition, I received valuable feedback and support from members of the Causal Cognition in a Complex World laboratory at Harvard Project Zero, including Kasia Derbiszewska, Lynneth Solis, Maleka Gramling, Y. Debbie Liu, Megan Powell, Denise Bressler, and Kenyon Maree, as well as from my extended academic family from Taiwan, including Jim Chang, Heidi Hsu, Ming-Chiao Lin, Pei-Ling Lin, and particularly Zoe Kao, whose analytical guidance saved the day on more than one occasion. Finally, I would like to give special thanks to Dr. John. B. Willett for his feedback on early versions of this work.

I would be remiss, of course, to not give due thanks and credit to my committee. Dr. Diane Jass Ketelhut has been a constant source of guidance and inspiration since my EdM candidacy at the Harvard Graduate School of Education. In many ways, I have tried to emulate her exemplary work, and am deeply honored by her presence on my committee. Dr. Chris Dede has also been a long-time mentor and professional paragon. I have never had a conversation with him in which I did not walk away with a deeper understanding of the field, and a renewed passion to explore it in creative ways. I am especially indebted to my committee chair and academic advisor, Dr. Tina Grotzer. Over the course of the last five-plus years, Tina has been my mentor, advisor, collaborator, and friend. Words cannot express how inspired I am by her brilliance, passion, and dedication to conducting meaningful research that informs teaching and learning. If, over the course of my professional life as an educational researcher, I am able to be half as impactful as she has been over the course of her career, I will consider myself lucky. Thank you for your boundless patience, guidance, and trust.

Portions of this material are based upon work supported by the National Science Foundation under Grant No. 0845632. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. Portions of the research reported here were also supported by the Institute of Education Sciences, U.S. Department of Education, through Grant DRK12 1118530. The opinions expressed are those of the author and do not represent views of the Institute or the U.S. Department of Education. Figures A1 through A4 and A7 were adapted from the EcoMUVE teacher guide developed by Drs. Shari Metcalf and Amy Kamarainen, to whom I am also eternally grateful, and used with teachers in the Spring of 2011.

Table of Contents

Tables & Figures

Trends in the salience of data collected in a multi user virtual environment:

An exploratory study

M. Shane Tutwiler

Abstract

In this study, by exploring patterns in the degree of physical salience of the data the students collected, I investigated the relationship between the level of students' tendency to frame explanations in terms of complex patterns and evidence of how they attend to and select data in support of their developing understandings of causal relationships. I accomplished this by analyzing longitudinal data collected as part of a larger study of 143 7th grade students (clustered within 36 teams, 5 teachers, and 2 schools in the same Northeastern school district) as they navigated and collected data in an ecosystems-based multi-user virtual environment curriculum known as the EcoMUVE Pond module (Metcalf, Kamarainen, Tutwiler, Grotzer, Dede, 2011) .

Using individual growth modeling (Singer & Willett, 2003) I found no direct link between student pre-intervention tendency to offer explanations containing complex causal components and patterns of physical salience-driven data collection (average physical salience level, number of low physical salience data points collected, and proportion of low physical salience data points collected), though prior science content knowledge did affect the initial status and rate of change of outcomes in the average physical salience level and proportion of low physical salience data collected over time.

The findings of this study suggest two issues for consideration about the use of MUVEs to study student data collection behaviors in complex spaces. Firstly, the structure of the curriculum in which the MUVE is embedded might have a direct effect

on what types of data students choose to collect. This undercuts our ability to make inferences about student-driven decisions to collect specific types of data, and suggests that a more open-ended curricular model might be better suited to this type of inquiry. Secondly, differences between teachers' choices in how to facilitate the units likely contribute to the variance in student data collection behaviors between students with different teachers. This foreshadows external validity issues in studies that use behaviors of students within a single class to develop "detectors" of student latent traits (e.g., Baker, Corbett, Roll, Koedinger, 2008).

Chapter 1: Introduction

I broke my leg in the winter of 2008. The nature of the injury was not particularly severe. In fact, it was so mild that I did not even know it was broken. I walked around with said broken leg, unaware, for approximately one week, until a nagging pain in my back prompted a visit to the doctor. He poured over my recent medical history, examined the area of pain, and then roundly informed me that there was nothing wrong with my back. After a quick radiographic examination of my knee, he nodded sagely and gave me the bad news. The barely-noticeable break in my right leg had caused me to change my walking gate and posture, which was in turn straining muscles in my lower back.

I peered over his shoulder at the various vital signs and notes in his charts, and stood beside him as he inspected the processed radiographic film. None of it made any sense to me, but painted a very clear picture of my medical condition for him. And then, he pointed out a hair-thin dark line on the film. That was the break. I had a chance to view more x-rays and sonographs over the course of my recovery, and eventually became quite adept at identifying features therein.

A few years later I stood beside my wife as we viewed the screen during one of the half dozen sonograms she would undergo during the course of her first pregnancy. I was so easily able to attend to subtle shifts in tone that I found it very simple to perceive the lines and patterns that defined the anatomical features of our fast-growing daughter; to such a degree that the technician asked if I worked in the prenatal medical field. My wife, in contrast, could hardly make heads or tails of the live image, and benefitted greatly from being given labeled still images at the conclusion of the visit.

The above vignette highlights the interplay between the perception of and attention to specific types of data. Past experiences and prior knowledge allowed me to perceive aspects of the scans that my wife could not, and thus drew my attention to specific areas that she did not see. This anecdote also highlights the use of those data to draw causal inferences based on covariation and ordering of events. The myriad data my doctor used to establish my diagnosis were largely random blips on pages to me, and yet he was able to order and use them in a meaningful way. Ultimately, it makes clear the shift that must occur for novices to perceive potentially important data in a complex environment, a necessary precondition if one is to use said data to infer causal relationships.

Although researchers are often reminded to adhere to the maxim that “correlation does not imply causation,” the two concepts are tightly intertwined. Inferring causal relationships based on probabilistic data about the co-occurrence of events is a central aspect of the endeavor of science. Philosophers of science such as Hume (1777) have noted that the only insights humans have on possible causal relationships lies in our observations of covariation. The study of how humans might use such data has occupied psychologists for decades, spurring them to create ever-more complex models of human causal inference including a set of models based on networks of event probabilities (Shultz, 2007). For various reasons, findings from these experimental studies are difficult to generalize to more complex data collection scenarios due to constraints in the experimental paradigms. Perhaps the most important limiting factor is that they all include a relatively small set of well-defined possible causes, a luxury not often found outside of the laboratory. Research on causal inferences in more complex and authentic spaces is thus warranted.

Virtual environments are digital representations of physical spaces. These environments allow researchers to collect rich behavioral information by registering and recording the actions of people who interact with them in the form of event logs. These rich event log data provide ideal means to conduct research on learning and behavior (e.g., Ketelhut, 2007; Ketelhut, Nelson, Clarke, & Dede, 2010). The virtual world under analysis in this study is the EcoMUVE, a multi-user virtual environment developed by Drs. Chris Dede and Tina Grotzer at Harvard University. Comprised of two distinct ecosystem settings, a virtual pond and virtual forest, the EcoMUVE was designed to help support knowledge of ecosystems science concepts and complex causal patterns in middle-school students (Metcalf, Kamarainen, Tutwiler, Grotzer, Dede, 2011). Using event log data from the EcoMUVE Pond, I mapped relationships between key measures within the candidate virtual environment and specific learner traits so that they can be manipulated effectively in future studies. To do so, I studied changes in the *salience* of data collected in the virtual world over time.

What exactly is salience, however? In its most general sense, salience is how visible or noticeable a piece of data is compared to surrounding data. This visibility can be impacted by two conditions—physical properties of the item and prior knowledge by the observer. Melloni, van Leeuwen, Alink, & Muller (2012f) propose a dual-channel theory of salience, in which the salience of an object is defined by both its physical features (bottom-up salience) as well as the cognitive framing of the observer (top-down salience). Consider the following examples.

The actual physical properties of the datum may be such that it is dissimilar in some respects to its surround. Imagine a novice pilot learning to fly a new type of aircraft. There are myriad buttons on the control panel. However, a flashing button will more likely be visible to the pilot compared to its non-flashing counterparts. She is more likely to attend to and perceive this data, and use it to inform her decision making. It has a higher physical, or bottom-up, salience value.

Additionally, the observer may have prior knowledge that makes her more likely to perceive the data point. Returning to our novice pilot example, we find that as she gains more experience and learns more about the instrumentation of the plane, the novice pilot notices subtle changes in gauge levels that transfer important information about the aircraft's performance. Her prior knowledge was guiding her attention and perception of data. This is top-down salience.

My measure of a student's tendency to offer explanations containing complex causal components represents the proportion of answers students gave to a causal scenario that were framed in terms of causally complex relationships such as action at a distance or non-obvious causes (Grotzer, 2012), and is one of the many aspects of the top-down component of salience (Melloni et al., 2012). This work is among the first to explore the interaction of the two aspects of salience in an authentic data-rich environment.

This dissertation study is divided into six chapters, the first being this introduction. In the Chapter Two I highlight key theories of data-driven human causal inference, culminating in approaches based on the Causal Bayes Nets (CBN) framework (Pearl, 1988), and outline the use of data salience, in a method first proposed by Tutwiler and Grotzer (2012) and expanded by Tutwiler (2013), to impact the perceived strength of covariation that drives causal inference within the CBN framework. I then outline the use of data from a specific ecosystems science based MUVE, the EcoMUVE (Metcalf, et al., 2011), to explore the relationship between prior knowledge and preference for complex causal answers and the salience of data collected in a complex inquiry space. I then propose three research questions and outline the research plan I used to answer them. In the Chapter Three I give details of the procedures and methods used to answer these research questions. I present my findings in the Chapter Four, and discuss potential contributions, implications, and limitations of the work in Chapter Five. I conclude with final thoughts in Chapter Six.

Chapter 2: Background and Context

In this chapter I present the empirical and theoretical background of my study, while placing it in the larger contexts of causal learning and MUVE-based research. I begin by highlighting past research on data-driven human causal inferences. I then make the critical connection between the salience of data in a system and the attention and selection behaviors of observers, postulating that data more likely to be selected is also more likely to be included in updating causal hypotheses. Next, I summarize research on the MUVE being studied, and highlight past research that has effectively used MUVE-generated data to facilitate inferences from student behaviors therein. I conclude by presenting my research questions and hypotheses.

Inferring causal connections from data

In the course of our daily experiences, we are awash with information about possible cause-and-effect relationships. Events continuously occur in temporal sequences, and our brains automatically connect the dots, often (but not always) correctly inferring causal associations. The attempt to answer the question of how we are able to do so, often with very sparse evidence, has resulted in numerous theories, and models to represent them. Here, I present a summary about the development of a specific class of such models that are central to my thesis study.

The most basic model for predicting human causal inference based on covariational data (data representing the co-occurrence of events), dubbed ΔP, accounts for judgments based on the frequency of target event occurrence in the presence and absence of a common cause (Jenkins & Ward, 1965). One way to expand this simple model to include multiple candidate causes is to let the learner view training trials with each candidate cause either present or absent; a model known as probabilistic contrast (PC) (Cheng & Novick, 1990). However, neither of these models is able to distinguish pure covariation from causation to detect the presence of possible causes that are not directly observed (Novick & Cheng, 2004), which humans do quite frequently. For example, one could draw a causal connection between yellowed fingers and lung cancer given these two models when, in fact, their relationship is conditional based on unobserved common cause of both (smoking).

One method to address these shortcomings is to incorporate the concept of causal power (the strength of the causal connection) into the covariation-based predictive models. By assuming non-zero probabilities of effects occurring—or not occurring—in the presence or absence of causes (that is to say, assuming that causal processes are truly probabilistic), Cheng's (1997) Power-PC model is able to predict human causal inference using the generative or preventative power of candidate causes with invariant probabilities on observed effects. Therefore, it would distinguish between instances where one cause, such as studying for more hours, acts generatively on test scores, while losing sleep acts preventatively. In their Necessary and Sufficient (NS)-Power model, Lu, Yuille, Liljeholm, Cheng, and Holyoak (2007) further restrain the Power-PC model parameters by assuming that humans prefer simple causal models and deterministic relationships.

Glymour (1998) reframes Cheng's (1997) power-PC theory as a noisy-OR Boolean problem, in which two causes can independently produce the same effect; this can be represented visually as a directed acyclic graph (DAG, Figure 1), a graphical model in which the arrows radiate from possible causes to effects with no feedback loops.

For example, in Figure 1, below, C and D are both candidate causes of effect E and $\omega 0$ and $\omega 1$ are calculations of causal power[1]:

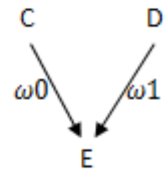


Figure 1. Example of a DAG with C and D being candidate causes of E and $\omega 0$ and $\omega 1$ representing the causal strength between candidate causes and the effect.

This framing of covariational learning as a DAG is important because, with two extra assumptions—that effects are independent conditional on their possible causes (Pearl, 1988), and that no parameterization of the DAG exists that would nullify the cause-effect relationship (Sprites, Glymour, & Scheines, 1993)—Cheng's (1997) Power-PC model represents a machine learning algorithm known as a Causal Bayes Net (CBN) (Pearl, 1988; Pearl, 2000). These CBN-based learning models are frequently used in the fields of machine learning and bioinformatics to detect causal strength and structure from sets of covariational data (e.g., Bishop, 2006; Neopolitan, 2009), and have been shown to be effective models of human learning of causal strength and structure as well (Griffiths & Tenenbaum, 2005).

Following Glymour's (1998) lead, numerous researchers have explored the use of CBNs as normative models for human data-driven causal inferences in scenarios with sparse data. For example, young children were shown to use direct and indirect probabilistic information to make simple causal inferences (Gopnik, Sobel, Shultz, & Glymour, 2001). Sobel, Tenenbaum, & Gopnik (2004) found that four year old children showed a preference for one cause over another when more than one cause is present (an

[1] In practice, people are often trying to compare candidate events in order to determine a unitary cause of an observed effect (Sobel, Tenenbaum, & Gopnik, 2004).

ability known as backwards-blocking), though three year olds did not. This ability to make causal inferences from probabilistic data was also found to be present across domains (Schulz & Gopnik, 2004).

And yet, there were limits to how subjects used and interpreted the effect of probabilistic causes on observed outcomes. For example, despite the fact that causes did not always lead to the expected effect in every trial, young children consistently assume that causes produce effects deterministically (Schulz & Sommerville, 2006). The accuracy of children's causal inferences from probabilistic data were also higher when observations of cause and effect were relatively spatially close (Kushnir & Gopnik, 2007). Additionally, Lu, Rojas, Beckers, & Yuille (2008) found that adults preferred possible causes that appeared across multiple trials, as opposed to those that only appear in one trial in a series, even if single-event possible causes had slightly higher generative power (Cheng, 1997).

Therefore, researchers have had success using CBN-based models of human data-driven causal inference. It should be noted that, in sharp contrast to this paper, possible causes and effects were made clearly apparent to the subjects, either through the direct observation of highly-salient events, or in well-defined summary tables of data across all of these studies. This is because CBN-based models are highly sensitive to information quality—the less well defined the event occurrences are, the more data is required for the models to accurately detect the presence of a causal relationship (Bishop, 2006). This is an important trait to consider, since humans are often forced to make causal inferences with scarce amounts of relatively low-quality data.

What is the role of data salience in causal inference?

Like all algorithms, CBN-based models are highly sensitive to the quality of data used in the inference process. As outlined above, the less informative the data (the higher the entropy), the more data that is required for individuals doing the reasoning to correctly infer causal connections. This concept of the quality of data in terms of information entropy is a central tenant of Shannon's (1948) *information theory*. It has previously been proposed that, based on information theory, one way for science educators to help scaffold student learning in complex, data-rich environments is to make relevant data more salient by either directly manipulating the data, or by teaching students to recognize important patterns (Tutwiler, 2013).

This connection between the physical salience of data and attention and perception is well established in both cognitivist and behaviorist literatures outside of the CBN-based research, but has not been explored directly in connection with CBN-based models. From a neurological perspective, salience is the degree to which certain data is distinguishable from other data in a perceptual field, and is modulated in the thalamus by the pulvinar nuclei during the attentional selection process (Snow, Allen, Raphal, & Humphries, 2009). As such, salience plays a role in attentional selection in psychological theories of data-driven learning (e.g., Wickens & McCarley, 2001; Tsakanikos, 2004). In these theories, the salience of an object is defined by both its physical features (bottom-up salience) as well as by observer-specific traits such as prior knowledge (top-down salience) (Melloni et al, 2012). Information about data is then processed along these two channels, and an overall degree of salience is determined. Data with higher salience, due to weighting along either channel, tend to garner more attention and are more likely to be considered when updating mental models of causal systems, a process known as *attentional selection* (e.g., Walther, Itti, Riesenhuber, Poggio, & Koch, 2002). This relationship between salience and attentional selection of data is a critical component of this study, as well as in future studies of student learning in virtual environments.

In this study, physical salience is analogous to the bottom-up salience (Melloni et al, 2012) described above, and is defined as the degree to which data is *conspicuous* in relation to the surrounding environment (Itti, Koch, & Nieber, 1998), a definition which has been operationalized and used in the Salience, Effort, Expectancy, and Value (SEEV) model of attention capture (Wickens, Helleberg, Goh, Xu, & Horrey ,2001; Wickens, Goh, Horrey, Helleberg, & Talleur, 2003; Wickens, McCarley, & Steelman-Allen, 2009) and guided Clarke (2009) in her exploration of the systematicity of data collection behaviors in River City, a MUVE developed by Dr. Chris Dede and colleagues at Harvard University to help support student understanding of the spread of diseases in a historical context (Ketelhut et al., 2010). Clarke (2009) found that students became more systematic in their data collection over successive visits to the River City MUVE. Her measure of systematicity accounted for the physical salience of the data each student collected over time. It is this aspect of her work upon which I draw most heavily here.

Student data collection in a complex virtual environment: The EcoMUVE

In this study I explore student data collection in a data-rich MUVE: the EcoMUVE pond module. As introduced above, the EcoMUVE is a multi-user virtual environment designed to help support middle-school student understanding of ecosystems science topics and complex causal patterns. In the EcoMUVE, students take on the roles of specific types of data-collection specialists (related to flora/fauna population, water chemistry, weather, personal testimony) and work in teams of four through a jigsaw pedagogical model (e.g., Colosi & Zales, 1998; De Baz, 2001) to solve a scientific mystery in one of two ecological settings: pond or forest (Metcalf et al, 2011). To do so, students must interact with the virtual environment through the use of specific data-collection tools, make careful observations about changes within the system over time, and interact with residents of the virtual world. They then share data from these interactions in order to formulate a possible hypothesis about the cause of the scientific mystery at hand, which they share with their classmates.

The present study is embedded within the larger design-based research agendas of the EcoMUVE Project Team, and is intended to provide guidance for future research involving this specific virtual environment. In the broader project work, two different modules were developed to test the feasibility of using specific affordances of virtual environments (Dede, 2009) to impact students' ecosystems content knowledge and recognition of complex causal patterns (Grotzer, Basca, & Donis, 2002). In the first module, EcoMUVE Pond, students exploring a virtual ecosystem centered on a pond observe that all of the large fish die on a given date. They must then travel back and forth through time (a specific affordance of the technology) within the MUVE, collecting data to explain the fish die-off. In the second module, EcoMUVE Forest, students travel back and forth through time to explore the impact on the ecosystem created by populations of wolves and deer on various forested islands.

Results of preliminary research indicated that use of the EcoMUVE in middle-school classrooms was feasible (Metcalf, Kamarainen, Grotzer, & Dede, 2012), and that students who used the EcoMUVE modules demonstrated positive shifts in measures of content knowledge (Metcalf, Kamarainen, Tutwiler, Grotzer, & Dede, 2013), science attitudes (Kamarainen, Metcalf, Grotzer, & Dede, 2012; Chen & Metcalf, 2013), and complex causal understanding (Tutwiler, Grotzer, Kamarainen, Metcalf, & Dede, 2011; Grotzer, Kamarainen, Tutwiler, Metcalf, & Dede, 2013). Prior research has also established that the initial use and rate of use of tools to collect data vary over the course of the student use of the EcoMUVE Pond module and are related to student prior knowledge of ecosystems science content and understanding of complex causal models (Tutwiler, Grotzer, Kamarainen, Metcalf, & Dede, 2013).

What of the *salience* of the data collected, however? The EcoMUVE Pond module is a data-rich virtual environment, with said data representing a wide range of physical salience. For example, a student in the role of water chemist may notice that the clarity of the water in the pond changes over time and might collect turbidity measurements across multiple days. Compared to the shift in turbidity measurements, the changing water clarity is a more easily recognizable feature of the pond, though they essentially both measure the same thing. In short, *the shift in water clarity is a more salient feature.*

In future research, we intend to vary the level of salience of data within the EcoMUVE Pond curriculum to measure its impact on student data collection and causal inferences. First, however, we need to understand the initial value and rate-of-change of the salience of data collected by students in the EcoMUVE Pond curriculum in its current state, and its relationship to student prior ecosystem content knowledge, attitudes about science, and knowledge of complex causal patterns.

Inferring student behavior from event log data

Traditional observational techniques, such as video-taped sessions or direct researcher observation, would not allow for the detailed exploration of change in the behavior of hundreds of students over time. Modern technologies allow us to overcome these limitations, however. For example, student actions in MUVEs can be meticulously observed and stored in longitudinal event log format. These records can then be analyzed via a host of exploratory and hypothesis-testing methods.

Researchers have utilized event log data to explore behavior in various MUVEs. For example in Quest Atlantis, a MUVE based on social and environmental citizenship for middle-school aged students, event log data have been parsed to explore student navigation and chat (Borner & Penumarthay, 2003; Penumarthay & Borner, 2006). In Whyville, a MUVE focused on upper-elementary to middle school science and social science learning, researchers used event log data to track user reaction to a virtual epidemic (Kafai, Quientero, & Feldon, 2010). Neither of these sets of analyses fully utilized the longitudinal nature of the event log data to explore and test hypotheses about student behavior in MUVEs, however.

The use of event log data to explore student behavior in the River City MUVE has been particularly fruitful. Using the River City MUVE as a case study, Dukas (2009) outlined design heuristics to maximize the use of MUVE-based event log data in exploratory and confirmatory research. Combining longitudinal event log data with pre-and-post intervention assessment data, Nelson (2007) found that students who used individual reflective guidance systems more frequently showed larger learning gains. In a similar vein of inquiry, Ketelhut (2007) found that students in the River City MUVE increased their scientific data gathering behaviors with each visit to the River City virtual world. Additionally, middle-school students who used the River City MUVE were found to become more systematic in their scientific behavior over time (Clarke, 2009). It is to this latter body of work that this dissertation study contributes by, as outlined below, formatting the event log data as a person-period data set and fitting individual growth models (Singer & Willett, 2003), similar to those used by Ketelhut (2007) and Clarke (2009), to specifically look at the relationship between the top-down and bottom-up salience of data collected by students exploring and learning in a MUVE.

Research Questions

In this chapter, I have reviewed theories on data-driven causal inferences in humans and highlighted their sensitivity to various properties of the data, specifically salience. I then framed how data salience drives attentional selection of data. Next, I outlined how event log data from MUVEs have been used to explore student data collection behaviors in past research. In this study, I contribute to the literature by exploring the relationship between aspects of students' prior knowledge and their attentional selection of data over time, based on the physical salience levels of the data they collect. Understanding this relationship is crucial, as it will inform curriculum designers and researchers as to what users attend to in the MUVE, but also potentially in similar real-world settings as well. This should allow content creators to better scaffold student learning in complex environments such as science labs, field trips, and MUVEs.

To discover potential methodological issues inherent to future research on the impact of data salience on causal inferences in a high fidelity virtual environment, I conducted an exploratory study of the ways in which students' prior preference for complex causal explanations is related to the physical salience of the data they collect in one such virtual environment, the EcoMUVE Pond. To do so, I fit models predicting three outcomes based on the level of physical salience of data collected by students across a number of visits to the EcoMUVE pond module in the spring of 2011. I then use these measures, as well as measures of pre-intervention preference for complex causal explanations, prior knowledge, and indicators of which teacher each student had, to answer the research questions below. For each research question, I offer an a priori hypothesis about the nature of the relationship between student pre-intervention tendencies to offer complex causal explanations and each outcome based on the extant research.

RQ1: Do students with a higher pre-intervention tendency to offer explanations containing complex causal components demonstrate more focused attentional selection of data by reducing the salience of the data they collect during each visit to the virtual world more rapidly than their peers with a lower pre-intervention tendency to offer explanations containing complex causal components?

Based on prior research on student data collection in the EcoMUVE (Tutwiler et al, 2013) and River City MUVE (Clarke, 2009), as well as the goal of the EcoMUVE curriculum, I hypothesize that students in the EcoMUVE will focus their attention on less physically salient data over time. I further hypothesize that students with a greater tendency to use more complex causal explanations will initially collect less physically salient data, and experience a steeper decline in the average salience of the data they collect over time in the EcoMUVE.

RQ2: Do students with a higher pre-intervention tendency to offer explanations containing complex causal components demonstrate more focused attentional selection of data by decreasing the number of collected data points that have low physical salience each visit to the virtual world less rapidly than their peers with lower pre-intervention tendency to offer explanations containing complex causal components?

In general, students collect less data each time they use the EcoMUVE (Grotzer et al, 2013). However, students with a greater tendency to use more complex causal explanations will collect comparatively more data points classified as having low physical salience over time compared to their peers with less complex causal understanding.

RQ3: Do students with a higher pre-intervention tendency to offer explanations containing complex causal components demonstrate more focused attentional selection of data by increasing the proportion of data they collect with low physical salience each visit to the virtual world more rapidly than their peers with a lower pre-intervention tendency to offer explanations containing complex causal components?

Students with a tendency to use more complex causal explanations will initially have a higher proportion of low physical salience data, and increase the proportion of low physical salience data they collect over time more rapidly than their peers with less complex causal understanding.

Chapter 3: Research Design

Site

Data for this study were collected in the Spring of 2011 from 143 7th-grade students (nested within five teachers) in a two schools in a suburban school district in the North Eastern United States. The student population of this district is approximately 61% White Non-Hispanic, 6% Asian, 23% Latino/Hispanic, and 7% Black/African American, with approximately 21% of students qualifying for Special Education status, 36% of students classified as First Language Not English, 13% of students classified as English Language Learners, and 39% of students qualifying for free or reduced price lunches. Teachers for this sample were recruited by the research team, and written consent was obtained from each student included in the dataset.

Dataset

Three types of data were analyzed for this study: time-stamped event log data from students' actions within the EcoMUVE pond module, pre-intervention assessment of prior knowledge and complex causal understanding, and demographic data collected prior to the EcoMUVE pond intervention. The unit of analysis for my study was at the student level. Taken as a set, these data allowed me to answer my research question as such: The average salience level of data collected, number of low-salience data points, ratio of low salience to total data collected per visit, and the record of student visits to the EcoMUVE pond module were all calculated from records in the event log. The question predictors were developed based on the pre-intervention assessment data, and the control variables were derived from demographic data and pre-intervention assessment data.

Sample

The sample for this study was comprised of 143 7th grade students, 53% of which are female. Student responses to the pre-intervention causal survey, described below, yielded a mean score of 0.58(SD=0.44) with a minimum of 0 and maximum of 1.0. I expect to be able to detect an effect size of 0.33 standard deviation units at a statistical power of 0.80 at standard levels (.05) of Type I error.

Instruments

To create a measure of the physical salience level of data gathered in the EcoMUVE Pond module, I used a method first employed by Clarke (2009). I developed and applied a rubric of salience for the various types of data in the EcoMUVE pond scenario based on visibility and location in the MUVE (Appendix B). For example, data on the size of the population of bass in the pond may be given a salience score of 2 out of 4 because it is in the pond (where students collect most data), but requires the student to look under the water to collect the information, whereas collecting bacteria population data using the submarine tool would score a salience of 1, since it requires the additional use of the submarine tool and multiple layers of magnification before using the "population tool" to collected the desired data.

Guidelines derived from the procedures used by Clarke (2009) were used to develop a basic coding scheme based on location and visibility of the data. Two independent raters[2] and I then applied this rubric to subsets (approximately 25% each rater per round) of the various types of data students are able to collect in the EcoMUVE

---

[2] The first rater was a graduate student in the educational technology program of a private university in New York . She was familiar with coding data in virtual worlds based on her course work with Professor Ryan Baker. The second rater was a graduate student in the educational technology program at a private university in Connecticut.

pond module. I discussed any points of departure between our scoring independently with each rater, and iteratively updated my rubric after each round. Once a high degree of agreement was met (>75%) between each of the raters and myself regarding the physical salience levels of the data in those subsets, I revised the rubric and used the final coding scheme to rate the degree of physical salience of the data collected by students in the EcoMUVE pond module.

With the physical salience accounted for, I explored the relationship between student tendency to offer explanations containing complex causal components and prior content knowledge, hypothesized to be critical components of top-down salience (Melloni et al, 2012), in relation to the level, amount, and proportion of physical salience collected. One such component was the degree to which each student had a tendency to offer explanations with complex causal patterns (Grotzer, 2012) that underlay the observed events in the EcoMUVE pond module. The other was the level of prior content knowledge possessed by each student (Melloni et al, 2012).

Student pre-intervention tendency to offer explanations containing complex causal components was assessed via the c (Grotzer & Derbiszewska, 2011), designed to assess the degree to which students reason about an ecological scenario using complex causal components (such as non-obvious causes, across spatial gaps, change over time, and changes in population levels) by coding responses to prompts about these complex causal patterns. A high or score on the complex causal components (low on simple components) or balanced score between them (as the ecosystem scientists argued for) is considered more sophisticated. The assessment is closely aligned with the EcoMUVE

Pond Module. Scoring of student responses by independent raters yielded a high inter-rater reliability (over .89 Cohen's Kappa).

Student pre-intervention content knowledge was measured using a 19 question mixed-format assessment comprised of 14 multiple choice and 5 open-response items (Metcalf & Kamarainen, 2011). Student responses to the multiple choice questions were marked right or wrong, and multiple assessors coded their open-ended responses (over .80 Cohen's Kappa). The instrument, as a whole, had adequate reliability (Cronbach's standardized alpha = 0.75).

Procedures

Student use of the EcoMUVE.

The EcoMUVE Pond unit is a ten day curriculum designed to be used with middle-school science students. A full account of the daily activities each student undertook is given in Appendix A. Students' tendencies to offer complex causal explanations, their prior content knowledge, and their demographic information were collected prior to the first day of the intervention. Students then visited the EcoMUVE (that is to say, logged-in to the world and collected data during a given class period) approximately six of the next ten class periods, exploring and collecting data within the virtual world.

Data for this study were collected on each student's second, third, and fourth visit to the EcoMUVE Pond. Students were introduced to the fish die-off on the second visit (the second day of the unit), and collected data within the MUVE on that visit as well as on the remaining four visits to the EcoMUVE (the fourth, fifth, seventh, and eighth days of the unit.) Prior to their third visit (day five of the unit), students were assigned to teams and chose roles to guide their data collection within the MUVE (see Appendix A for details.)

Data from the first, fourth, fifth, or sixth visits to the EcoMUVE were not analyzed. As outlined in Appendix A, the first visit (on the first day of the unit) occurred before the students were made aware of the fish die-off event; they were tasked with collecting photographs of flora and fauna within the virtual world to help them learn about the different areas and how to navigate therein. The data from this visit would not be appropriate to answer my research questions, which assume students are collecting data in service of unraveling the causal-and-event relationships that resulted in the fish die-off event. Observations of student use of the EcoMUVE in similar classroom environments showed that later in the curriculum (visits four through six) students often shared information by talking in groups outside of the MUVE, or shared information across groups. This erodes the validity of any inferences about student-level data collection behaviors, and resulted in my decision to exclude those visits from this analysis.

Measures

Given the longitudinal nature of the outcomes, the values of all measures were recorded in a person-period dataset (Singer & Willett, 2003). Time was recorded from each student's first visit to the EcoMUVE pond module until their last visit. For example, if a student used the EcoMUVE three times over the course of one school week (five days), there are three rows of data for that particular student. The outcome predictors all represent different measurements of bottom-up salience, whilst the question predictor and covariates (vector of teacher fixed effects notwithstanding) represent top-down contributions to salience (Melloni et al, 2012). Descriptive statistics of the measures are given in Table 1.

Outcomes.

$AVG\_SALIENCE_{ij}$ is a continuous variable that represents the average level of salience of data collected by a student *j* on a given visit *i* to the EcoMUVE pond module. It was constructed by summing across the salience scores (ranging from 1 to 4) of each piece of data collected that day and dividing by the number of discrete data collection events for that student on the particular day.

NUM_LOWij is a continuous variable that represent the number of data points collected by student *j* on visit *i* that have low values (1 or 2) of physical salience. To meet distributional assumptions, this variable was natural-log transformed.

$PROP\_LOW_{ij}$ is a continuous variable[3] that represents the proportion of data collected by student *j* on visit *i* that are classified as having low physical salience (values of 1 or 2). It was constructed by summing the number of low-salience data points collected each day and dividing it by the total number of data points collected. To meet distributional assumptions, the square-root of this variable was arcsine transformed (Kirchner, 2001).

Question predictors.

To answer my research questions, I first modeled each of the outcome variables on the predictor $VISIT_{ij}$, an ordinal variable (0, 1, 2) which represents the second through

[3] Though proportions are strictly bound by 0 and 1, proper transformation allows for tentative inferences to be made based on regression-based analysis of this outcome (Johnston, 2014).

fourth visit to the EcoMUVE by student *j* on day *i.* In my individual growth models, the intercepts correspond to a student's second visit to the EcoMUVE.

I then included a measure of tendency to give complex causal explanations ($CAUSAL\_PRE_j$), derived from the pre-intervention assessment as described in the Instruments and Procedures sections above. I calculated the proportion of complex explanations by creating a ratio of complex (non-obvious and distant) to total (obvious, local, non-obvious, and distant) count scores on the Causal Dynamics Assessment to produce the $CAUSAL\_PRE_j$ measure. This measure was then centered on the sample mean to aid in interpretability of the associated regression coefficients (Singer & Willett, 2003).

Controls.

Based on prior research (Metcalf et al, 2013), I included a vector of dummy control variables to account for between-teacher variability ($TEACHER_j$)[4], again withholding one from the model for comparison. In addition, I included a dummy variable set to 1 if the student was female and 0 if they were male ($FEMALE_j$), based on findings of the impact of gender on student data collection in past MUVE research (Ketelhut, 2007). Finally, to control for the possible effects of prior knowledge of ecosystems science concepts, an important top-down salience component per Melloni et

[4] Exploratory intra-class correlation analyses across all three outcomes in this study resulted in my excluding random effects or vectors of fixed-effects to control for possible autocorrelation due to clustering by team and class period.

I also excluded controls for the roles that students took on in the EcoMUVE. This is due to the fact that they were assigned those roles after having already made two visits to the EcoMUVE and their influence is only reflected in one day of data for this study. Roles were also chosen by students, and not exogenously assigned by the teacher. An exploratory analysis showed that they were not statistically significantly related to the outcomes, and were thus excluded from the analysis.

al. (2012), on data gathering behaviors. I included the total score, centered on the sample mean, from the pre-intervention content knowledge assessment described in the Instruments and Procedures sections ($PRE\_CONTENT_j$). These controls are represented in the models below as the student-level vector Z_j.

Table 1. Descriptive statistics

	Mean	SD	Min	Max	Obs.
AVG_SALIENCE	2.46	0.87	1	4	389
NUM_LOW[a]	2.48	1.14	0.69	5.07	389
PROP_LOW[b]	0.78	0.44	0	1.57	389
VISIT	1.02	0.81	0	2	389
PRE_CAUSAL[c]	0.00	0.15	-0.41	0.34	389
PRE_CONTENT[c]	0.00	4.66	-11.85	11.14	389
FEMALE	0.53	0.50	0	1	389
TEACHER1	0.23	0.42	0	1	389
TEACHER2	0.22	0.41	0	1	389
TEACHER3	0.25	0.44	0	1	389
TEACHER4	0.12	0.33	0	1	389
TEACHER5	0.18	0.39	0	1	389

a Natural log transformed
b Arcsine-square root transformed
c Centered on sample mean

Data Analytic Plan

Cleaning and shaping of event log data.

As with any complex quantitative study, cleaning and shaping of data were major parts of the overall effort undertaken in conducting this study. In general, the cleaning and shaping process consisted of four major steps: 1) Exporting data from the EcoMUVE back-end to Microsoft Excel via SQL, 2) cleaning and mining the data in Microsoft Excel using VBASIC and Pivot Table functions, 3) evaluating the cleaned data with the physical salience rubric described above, and 4) shaping the data into the required format for longitudinal analysis.

The first two steps were relatively straightforward. Student data records, previously anonymized and organized by an assigned student id number, were downloaded from the server in which they were stored via a secure web portal. This was accomplished primarily through the use of SQL scripts. This produced approximately sixty-thousand rows of data that had to be further cleaned and mined prior to analysis. To accomplish this, using VBASIC I deleted any events that were canceled prior to student completion. For example, if a student took a photograph of a duck in the pond, but chose not to store the information in the virtual field guide, that data entry was removed. I then screened out any events that were not related to data collection, such as logging-in to or out of the MUVE.

Next, I used the finalized physical salience rubric (Appendix B) to calculate the physical salience scores of the various data that students could collect in the EcoMUVE Pond. The data were ordered by their location and visibility and coded per the rubric. After the salience codes were added, student pre-intervention demographic, causal, and content-knowledge scores were merged. Finally, the data were arranged into a person-period format (Singer & Willett, 2003) in order to support the subsequent analyses. This required me to sort and group all of the observations first by student ID number, and then by date (converted to the VISIT variable described above.)

Data analysis.

To answer my research questions, I fitted a series of multi-level individual growth models (Singer & Willett, 2003). The same structural (regression coefficients) and stochastic (residuals) components were used across all three models. In this section, I present the model used to answer the first research question, and explain how the other two were answered using similarly fitted models.

RQ1: Do students with a higher pre-intervention tendency to offer explanations containing complex causal components demonstrate more focused attentional selection of data by reducing the salience of the data they collect during each visit to the virtual world more rapidly than their peers with a lower pre-intervention tendency to offer explanations containing complex causal components?

To answer this question, I first propose the following level-1 trajectory for the salience of data collected by student *j* on visit *i*:

$$\mathrm{AVG\_SALIENCE_{ij}} = \pi_{0i} + \pi_{1i}(\mathrm{VISIT_{ij}}) + \varepsilon_{ij}$$

$$\text{where } \varepsilon \sim N(0, \sigma_\varepsilon^2)$$

In this hypothesized level-1 population model the intercept parameter, π_{0i}, represents the initial average salience of data collected while the slope parameter, π_{1i}, represents the rate of change of salience of data collected per visit to the EcoMUVE pond module. The level-1 residual term, ε_{ij}, which denotes within-person differences in the outcome, is hypothesized to be normally distributed with a mean of zero and variance σ_ε^2.

To answer the first research question, I also propose the following level-2 model:

$$\pi_{0i} = \gamma_{00} + \gamma_{01}\mathrm{CAUSAL\_PRE_j} + \delta_0 Z_j + \xi_{0j}$$

$$\pi_{1i} = \gamma_{10} + \gamma_{11}\mathrm{CAUSAL\_PRE_j} + \delta_1 Z_j + \xi_{1j}$$

$$\text{where} \begin{bmatrix} \xi_{0j} \\ \xi_{1j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix}\right)$$

In this hypothesized level-2 population model the intercept parameters, γ_{00} and γ_{10}, represent the population average initial values of the level and rate of change, respectfully, of the salience of data collected during each visit of a hypothetical male student with no pre-intervention tendency to offer explanations containing complex causal components and no prior ecosystems content knowledge in the classroom of the teacher who filled the comparison role in the EcoMUVE pond module. The slope parameters γ_{01} and γ_{11} represent the effect of student pre-intervention tendency to offer explanations containing complex causal components on the initial level and rate of change, respectively, of the salience of data collected during each visit of a hypothetical male student with no prior content knowledge in the classroom of the teacher who filled the comparison role in the EcoMUVE pond module. The parameter δ represents the effect of a vector of control variables accounting for the effect prior content knowledge, gender, and the student's teacher on the initial level (γ_{02}- γ_{07}) and rate of change (γ_{12}- γ_{17}), respectively, of the salience of data collected during each student's visit to the EcoMUVE pond module. It is assumed that the level-2 residuals, ξ_{0j} and ξ_{1j} are randomly drawn from a bivariate normal distribution with respective variances of σ_0^2 and σ_1^2, and a covariance of σ_{10}.

To answer my first research question, I examined the parameters γ_{01} and γ_{11}. For example, if γ_{01} and γ_{11} are negative and statistically significant, then I can claim that students with a higher pre-intervention tendency to offer explanations containing complex causal components will initially collect less salient data as well as demonstrate a more rapid focusing of their attention on less-salient data compared to their peers with a lower pre-intervention tendency to offer explanations containing complex causal components, on average in the population and controlling for all other variables in the model.

The preceding two-level model can also be represented as the following hypothesized population mixed model:

$$\begin{aligned}\mathrm{AVG\_SALIENCE}_{ij} &\\ = \gamma_{00} + \gamma_{10}\left(\mathrm{VISIT}_{ij}\right) + \delta_0 \mathrm{Z}_j + \gamma_{01}\mathrm{CAUSAL\_PRE}_j + \gamma_{11}(\mathrm{CAUSAL\_PRE}_j &\\ * \mathrm{VISIT}_{ij}) + \delta_1(\mathrm{Z}_j * \mathrm{VISIT}_{ij}) + \{\varepsilon_{ij} + \xi_{0j} + \xi_{1j} * \mathrm{VISIT}_{ij}\}\end{aligned}$$

To answer my second and third research questions, I replaced $\mathrm{AVG\_SALIENCE}_{ij}$ with the outcomes $\mathrm{NUM\_LOW}_{ij}$ (log transformed) and $\mathrm{PROP\_LOW}_{ij}$ (arcsine-square root transformed) respectively. I then examined I examined the parameters γ_{01} and γ_{11} in each model, making inferences as outlined above.

Chapter 4: Findings

In this section I first report the shift in trends of physical salience of data collected in the second, third, and fourth visits to the EcoMUVE Pond unit. I then present the results, organized by research question, of fitting the aforementioned individual growth models. In each section I will first refer to a table with fitted taxonomies of regression models. Column 1 of each table includes the fixed-effects measures of initial value (Rows 1-9) and rate of change (Rows 10-16) of each model. Variance components are given in Rows 17-19, while a range of fit statistics are given in Rows 20-22. Column 2 of each table presents the coefficient symbols given in the population-level model above. In each table, I follow the same pattern of model building, moving from an unconditional growth model (Column 3) with time as the only predictor, adding key control variables (Column 4), adding the question predictor (CAUSAL_PRE_j) (Column 5), and then presenting a final parsimonious model for further evaluation (Column 6). No statistically significant interactions were detected. Finally, I highlight key findings via figures with plots of prototypical students. Further insights on these findings are given in the discussion section that follows.

Trends in Physical Salience over Time

Figure 2 is a panel of plots summarizing the data collected across all three measured student visits to the EcoMUVE. The vertical axes are counts of data collected, and the horizontal axes are physical salience levels of the data. Note that, across all three observed visits to the EcoMUVE Pond virtual environment, a majority of the data collected were coded as low physical salience (mostly level 1). Looking at Panel A, I see that students collected just over seven thousand pieces of data on their second visit to the

EcoMUVE. This rises to just over eight thousand collected pieces of data on visit 3 (Figure 2, Panel B), before dropping sharply to just over five thousand pieces on visit 4 (Figure 2, Panel C). I also note that students collected a higher number of very high physical salience (level 4) data on the second and third visits relative to the amount of moderately physically salient data (level 3).

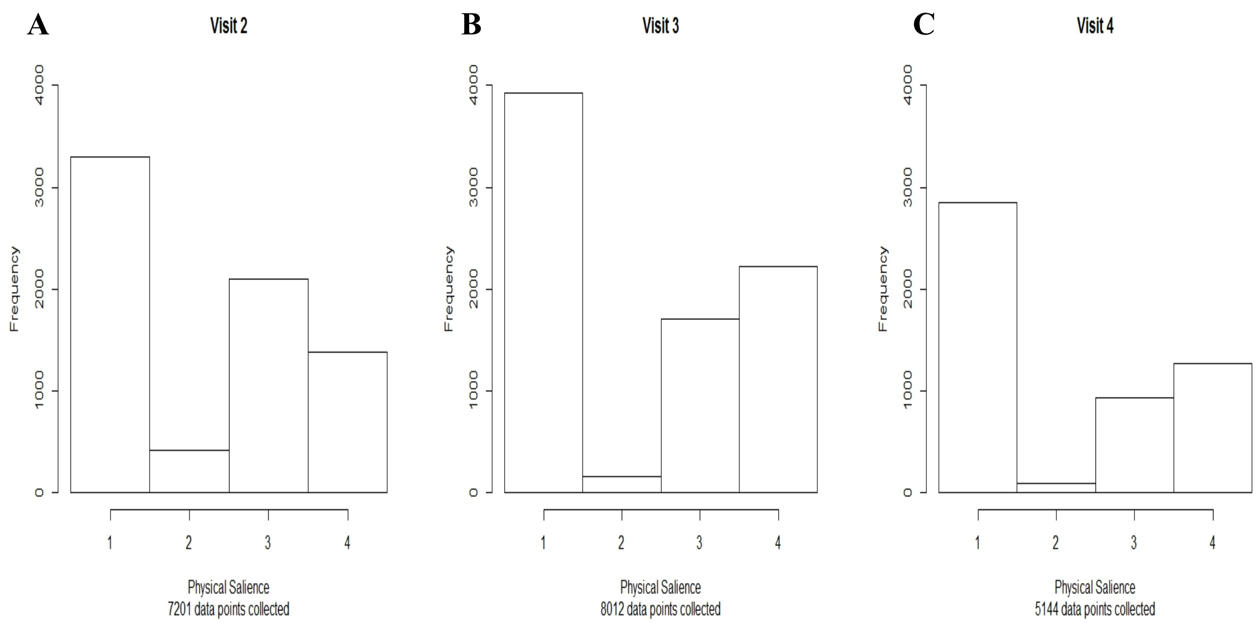


Figure 2. Panel of plots showing the frequency of data collected by students (n=143) at each level of physical salience across the second (Panel A), third (Panel B), and fourth (Panel C) visits to the EcoMUVE pond unit.

RQ1: Average physical salience collected per visit to the EcoMUVE

Viewing Table 2, Model 1, note that without controlling for the effect of prior knowledge, gender, or teacher, the average level of physical salience of data collected by students on their second visit to the EcoMUVE is 2.34 (γ_{00} , $p<.001$), and increases by about 0.12 (γ_{01}, $p<.05$) per visit over the next two days of the curriculum.

Controlling for these factors (Table 2, Model 2), on average in the population every one point difference in pre-intervention content equates to an average physical salience score that is 0.02 (γ_{02} , $p<.05$) points lower on Visit 2. In addition, the rate of change of the level of physical salience of data collected does not differ based on prior content knowledge ($\gamma 12=0.01$, n.s.). The vector of fixed-effects accounting for teacher impact on the initial level ($\gamma 04$- $\gamma 07$) and rate of change ($\gamma 14$- $\gamma 17$) are both statistically significant ($p<.05$). Student gender was not statistically significantly related to the initial value ($\gamma 03= -0.10$, n.s.) or the rate of change ($\gamma 13= 0.16$, n.s.) in average physical salience level of data collected over time, controlling for prior knowledge fand teacher.

Adding the question predictor (PRE_CAUSAL) in order to answer my first research question (Table 2, Model 3), I observe that student causal knowledge prior to the intervention, as measured by the proportion of complex responses given on the assessment, does not have a statistically significant effect on the starting value ($\gamma 01$, n.s.) or rate of change ($\gamma 11$, n.s.) of the average physical salience level of data collected, controlling for all other factors in the model. That is, *students with higher pre-intervention tendency to offer explanations containing complex causal components do not exhibit any more or less rapid change in the physical salience of data they choose to collect over time compared to their peers with lower pre-intervention tendency to offer explanations containing complex causal components, controlling for prior content knowledge, gender, and the effects of their teacher.*

The final parsimonious model (Table 2, Model 4), then, represents the starting point and change in average physical salience levels of data collected by 7^{th} grade students over time in the EcoMUVE pond module. I highlight two important trends from this model in the fitted plots of prototypical students that follow. Figure 2 illustrates the effect of prior knowledge on the physical salience level of data collected over time for two randomly selected students (i.e., each measure in the vector of teacher fixed effects was set to 0.5), one of whom has high prior content knowledge (two standard deviations above the sample mean) and one of whom has low prior knowledge (two standard deviations below the sample mean). Figure 3, on the other hand, highlights differences in the average physical salience levels of data collected over time by prototypical students (with prior knowledge at the sample mean) across the five teachers in the sample.

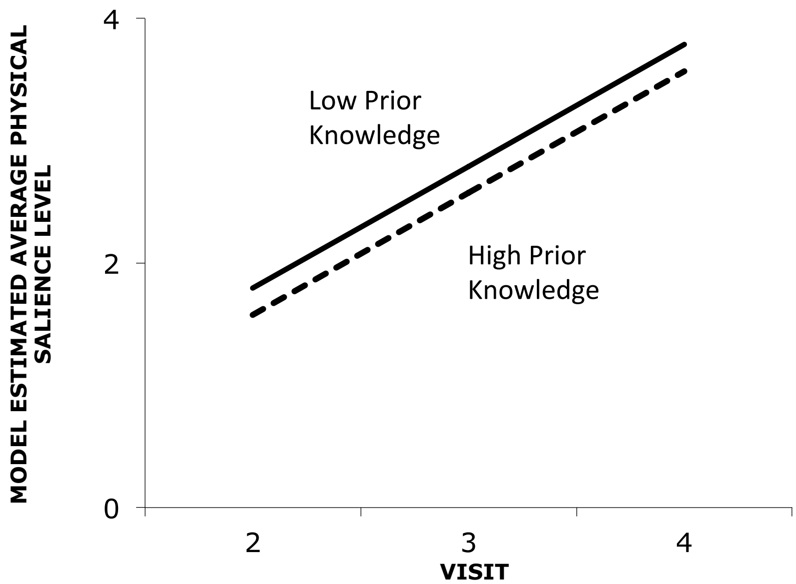


Figure 3. Average physical salience level of data collected by randomly selected students as a function of visit by students with high and low prior knowledge (n=143).

Observing Figure 3, note that there is a marginal initial difference ($p<.10$) between randomly selected students with high and low prior knowledge. Students with high prior knowledge initially collect data with an average physical salience value of 1.58, whereas randomly selected students with low prior knowledge initially collect data with a physical salience value of 1.79. This gap is then consistent over time, with high prior-knowledge students collecting data with lower average physical salience values than their peers with lower prior knowledge.

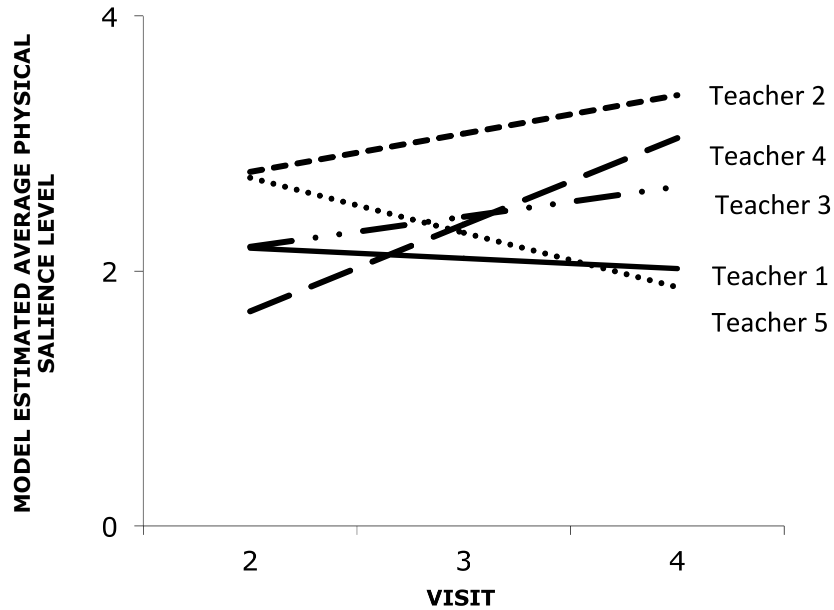


Figure 4. Average physical salience level of data collected by prototypical students (prior content knowledge at sample mean) as a function of visit, by teacher (n=143).

Inspection of Figure 4 yields another side of the story, however. If one considers prototypical students (with prior content knowledge at the sample mean) from the class of each teacher, then a wide range of variance between teachers in the trends of physical salience levels of collected data becomes apparent. For example, note that the initial value of average physical salience was indistinguishable for teachers 2 and 5 on Visit 2. However, their longitudinal trends in data collection were strikingly different, with teacher 2's prototypical students collecting data that was more physically salient and teacher 5's prototypical students collecting data that was less salient. As such, teacher 2's prototypical students collected the data with the highest average physical salience on Visit 4, whereas teacher 5's prototypical students collected the least salient data, a difference that was statistically significant ($p<.05$). I explore potential sources of this variance in the discussion section that follows.

Table 2. Results of fitting individual growth models of the change in average physical salience of data collected over time by middle school students in a multi-user virtual environment (n_teach=5; n_student=143; n_observations=389)

		Model 1. Unconditional Growth Model	Model 2. Model 1 +Controls	Model 3. Model 2 +Predictor	Model 4. Final Model
Fixed Effects					
INTERCEPT	γ_{00}	2.34*** (0.06)	2.80*** (0.15)	2.78*** (0.16)	2.73*** (0.14)
VISIT	γ_{10}	0.12* (0.06)	-0.54*** (0.14)	-0.53*** (0.14)	-0.43*** (0.12)
PRE_CAUSAL†	γ_{01}			0.22 (0.33)	
PRE_CONTENT†	γ_{02}		-0.02* (0.01)	-0.02* (0.01)	-0.01~ (0.01)
FEMALE	γ_{03}		-0.10 (0.11)	-0.09 (0.10)	
TEACHER1	γ_{04}		-0.56** (0.18)	-0.55** (0.18)	-0.55** (0.18)
TEACHER2	γ_{05}		0.03 (0.18)	-0.04 (0.18)	0.05 (0.18)
TEACHER3	γ_{06}		-0.55** (0.17)	-0.54** (0.17)	-0.54** (0.19)
TEACHER4	γ_{07}		-1.10*** (0.21)	-1.10*** (0.21)	-1.05*** (0.20)
PRE_CAUSAL*VISIT	γ_{11}			-0.10 (0.33)	
PRE_CONTENT*VISIT	γ_{12}		0.01 (0.01)	0.01 (0.01)	
FEMALE*VISIT	γ_{13}		0.16 (0.10)	0.16 (0.10)	
TEACHER1*VISIT	γ_{14}		0.37* (0.16)	0.36* (0.17)	0.35* (0.16)
TEACHER2*VISIT	γ_{15}		0.76*** (0.16)	0.76*** (0.15)	0.73*** (0.15)
TEACHER3*VISIT	γ_{16}		0.69*** (0.16)	0.68*** (0.16)	0.66*** (0.16)
TEACHER4*VISIT	γ_{17}		1.19*** (0.19)	1.19*** (0.19)	1.11*** (0.19)
Variance Components					
Level1 Residual	σ_{ε}^2	0.6064***	0.5436***	0.5423***	0.5513***
Level 2 Initial	σ_0^2	-0.06701~	-0.1211***	-0.1210***	-0.1263***
Level 2 Rate of Change	σ_1^2	0.1202***	0.06648**	0.06686**	0.06996***

Goodness of fit				
-2LL	966.0	837.0	836.4	841.4
AIC	976.0	871.0	874.4	869.4
BIC	990.8	921.4	930.7	910.9

† Centered on the sample mean
~p<.1; *p<.05; **p<.01; ***p<.001

RQ2: Number of low physical salience data points collected per visit to the EcoMUVE

Viewing Table 3, Model 1, note that without controlling for the effect of prior knowledge, gender, or teacher, the (log transformed) number of low physical salience data points collected by students on their second visit to the EcoMUVE is 2.74 ($\gamma 00$, $p<.001$), and decreases by about 0.26 ($\gamma 01$, $p<.05$) per visit over the next two days of the curriculum.

How do the student's prior knowledge, gender, and teacher affect this trend? Note while looking at Model 2 of Table 3 that neither prior knowledge nor gender have a statistically significant relationship with initial value ($\gamma 02= 0.02$, n.s.; $\gamma 03= 0.10$, n.s.) or change over time ($\gamma 12= -0.02$, n.s.; $\gamma 13= -0.18$, n.s.) in the (log transformed) number of low physical salience data points collected over time in the EcoMUVE. However, as with the trend in average level of physical salience, the vectors of dummy variables predicting the effect of teachers on initial value ($\gamma 04$- $\gamma 07$) and rate of change ($\gamma 14$- $\gamma 17$) were statistically significant ($p<.05$).

To answer my second research question, I added the question predictor ($CAUSAL\_PRE_j$) to Model 2 (Table 3, Model 3). Notef that student pre-intervention causal understanding is not statistically significantly related to the starting value ($\gamma 01=$ -

0.61, n.s.) or change over time ($\gamma 12= -0.02$, n.s.) in the (log-transformed) number of low physical salience data points collected by students in the EcoMUVE, when controlling for prior knowledge, gender, and teacher. In other words, *students with higher pre-intervention tendency to offer explanations containing complex causal components do not reduce the number of low salience data points collected over time any less rapidly than their peers with lower pre-intervention tendency to offer explanations containing complex causal components.*

The final model (Table 3, Model 4), then, reflects the fact that student pre-intervention tendency to offer explanations containing complex causal components, content knowledge, and gender have no effect on trends in the amount of (log-transformed) low physical salience data they collect over time, when controlling for the effects of their teacher. I highlight this via two plots of prototypical students. Figure 43 shows the trend in number of low salience data points collected by a randomly chosen student of either gender and of any level of prior causal and content knowledge. Figure 5 shows how this trend differs across prototypical students with different teachers.

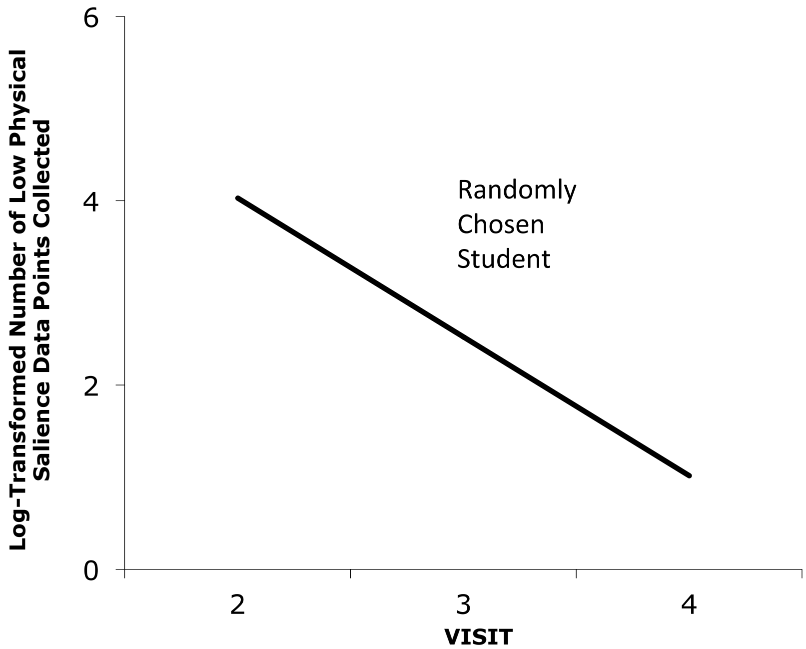


Figure 5. Number of (log transformed) low physical salience data points collected by a randomly selected student as a function of visit (n=143).

Inspecting Figure 5, Note that, regardless of prior knowledge or gender, a student chosen at random initially collects approximately 4 (log-transformed) low physical salience data points on Visit 2. However, she or he only collects 1 (log-transformed) low physical salience data point on Visit 4. This represents a statistically significant ($p<.001$) change across visits.

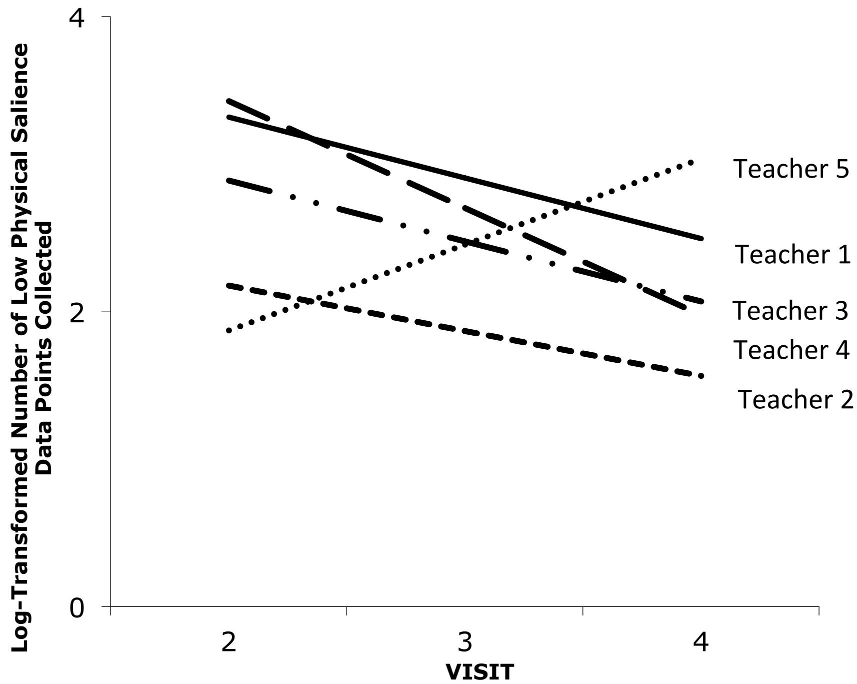


Figure 6. Number of (log transformed) low physical salience data points collected by prototypical students as a function of visit, by teacher (n=143).

Evaluating the trend from a different perspective, Note in Figure 6 that prototypical students in teacher 5's classes collected more (log-transformed) low physical salience data points over time, as opposed to their peers in other teachers' classes. For example, teacher 2 and teacher 5's students both collect about 2 (log-transformed) low physical salience data points on Visit 2, a difference that is not statistically significantly different. However, by Visit 4 teacher 2's students collected about 1 (log-transformed) low physical salience data point, whereas teacher 5's students collected about 3 (log-transformed) low physical salience data points, a difference that is statistically significant ($p<.05$).

Table 3. Results of individual growth models of the log-transformed number of low physical salience data (level 1 or 2) collected over time by middle school students in a multi-user virtual environment (n_teach=5; n_student=143;n_observations=389)

		Model 1. Unconditional Growth Model	Model 2. Model 1 +Controls	Model 3. Model 2 +Predictor	Model 4. Final Model
Fixed Effects					
INTERCEPT	γ_{00}	2.74*** (0.09)	1.79*** (0.23)	1.82*** (0.23)	1.87*** (0.21)
VISIT	γ_{10}	-0.26*** (0.07)	0.70*** (0.18)	0.68*** (0.18)	0.58*** (0.16)
PRE_CAUSAL†	γ_{01}			-0.61 (0.49)	
PRE_CONTENT†	γ_{02}		0.02 (0.02)	0.03 (0.02)	
FEMALE	γ_{03}		0.10 (0.15)	0.08 (0.16)	
TEACHER1	γ_{04}		1.46*** (0.27)	1.44*** (0.27)	1.45*** (0.26)
TEACHER2	γ_{05}		0.32 (0.26)	0.37 (0.26)	0.30 (0.26)
TEACHER3	γ_{06}		1.04*** (0.26)	1.01*** (0.26)	1.01*** (0.26)
TEACHER4	γ_{07}		1.65*** (0.31)	1.67*** (0.31)	1.55*** (0.30)
PRE_CAUSAL*VISIT	γ_{11}			0.66 (0.43)	
PRE_CONTENT*VISIT	γ_{12}		-0.02 (0.01)	-0.02 (0.01)	
FEMALE*VISIT	γ_{13}		-0.18 (0.14)	-0.15 (0.14)	
TEACHER1xVISIT	γ_{14}		-1.01*** (0.15)	-1.00*** (0.22)	-0.99*** (0.22)
TEACHER2xVISIT	γ_{15}		-0.92*** (0.22)	-0.92*** (0.21)	-0.89*** (0.22)
TEACHER3xVISIT	γ_{16}		-1.02*** (0.21)	-1.00*** (0.22)	-0.99*** (0.21)
TEACHER4xVISIT	γ_{17}		-1.40*** (0.26)	-1.43*** (0.26)	-1.30*** (0.25)
Variance Components					
Level1 Residual	σ_ε^2	1.2718***	0.9988***	1.0793***	1.1000***
Level 2 Initial	σ_0^2	-0.09770	-0.1676**	-0.1641**	-0.1740**
Level 2 Rate of Change	σ_1^2	0.04643	0.04880	0.04725	0.05193

Goodness of fit				
-2LL	1189.2	1095.1	1092.7	1099.1
AIC	1199.2	1129.1	1130.7	1125.1
BIC	1214.0	1179.4	1187.0	1163.6

† Centered on the sample mean
p<.1; *p<.05; **p<.01; ***p<.001

RQ3: Proportion of low physical salience data collected per visit to the EcoMUVE

Viewing Table 4, Model 1, Note that without controlling for the effect of prior knowledge, gender, or teacher in an *unconditional growth model*, the (arcsine square root transformed) proportion of low physical salience data points collected by students on their second visit to the EcoMUVE is 0.84 ($\gamma 00$, $p<.001$), and decreases by about 0.06 ($\gamma 01$, $p<.05$) per visit over the next two days of the curriculum.

When I add the controls for the effects of prior content knowledge, gender, and teacher to the unconditional growth model (Table 4, Model 2), Note that prior content knowledge has a statistically significant effect on the initial value ($\gamma 02= 0.02$, $p<.01$) and a marginally significant effect on the rate of change ($\gamma 12= -.01$, $p<.10$) of the (arcsine square root transformed) proportion of low physical salience data points collected by students over Visits 2 through 4 in the EcoMUVE. As in the previous models, gender has no effect on the initial value ($\gamma 03= 0.06$, n.s.) or rate of change ($\gamma 13= -0.09$, p<.n.s.) of the (arcsine square root transformed) proportion of low physical salience data points collected by students in the EcoMUVE over Visits 2 through 4. Also, as in previous models, the vector of teacher fixed effects on the initial status and rate of change on the (arcsine square root transformed) proportion of low physical salience data points collected by students in the EcoMUVE over Visits 2 through 4 are statistically significant ($p<.05$).

Adding the question predictor (PRE_CAUSAL) to Model 2 (Table 4, Model 3) Note that student pre-intervention tendency to offer explanations containing complex causal components is not statistically significantly related to the starting value ($\gamma01=$ 0.09, n.s.) or change over time ($\gamma12=$ -0.02, n.s.) in the (arcsine square root transformed) proportion of low physical salience data points collected during Visits 2 through 4 of the EcoMUVE. In other words, *students with higher pre-intervention tendency to offer explanations containing complex causal components do not increase the proportion of low physical salience data they collect any more or less rapidly over time than their peers with lower pre-intervention tendency to offer explanations containing complex causal components.*

The final parsimonious model is given in Table 4, Model 4, and shows that the initial value and rate of change of the (arcsine square root transformed) proportion of low physical salience data points collected by students over time in the EcoMUVE are affected by their prior content knowledge, when controlling for the effect of their teachers. I highlight two trends from this model via fitted plots of prototypical students. In Figure 6, I show the effect of prior content knowledge on data collection over time. In Figure 7, I highlight the differential trends between students with different teachers.

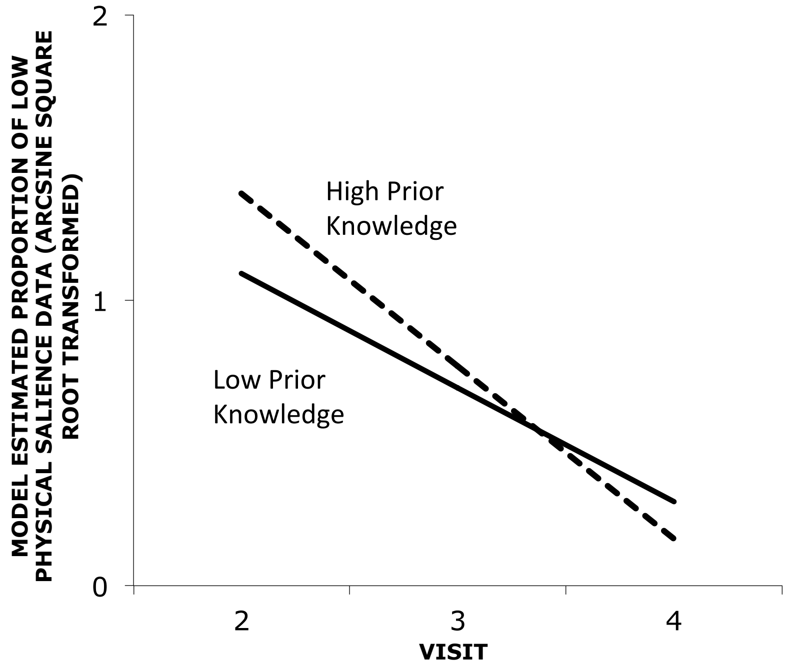


Figure 7. Proportion of low physical salience (arcsine square root transformed) data collected by randomly selected students as a function of visit by students with high and low prior knowledge (n=143).

Inspecting Figure 7, note that a prototypical student chosen at random from the teachers with high prior content knowledge (two standard deviations above the sample mean) initially collect a statistically significantly (p<.05) high proportion (arcsine square root transformed) of low physical salience data (1.37) than a prototypical student chosen at random with lower prior content knowledge (two standard deviations below the sample mean) on Visit 2 (1.09). Both students collect a smaller proportion of low physical salience data over time, but the rate of decrease is greater for the high prior knowledge student. As such, the high prior knowledge student collects a marginally (p<.10) smaller (arcsine square root transformed) proportion (0.16) of low physical salience data than their low prior knowledge peer (0.29) on Visit 4.

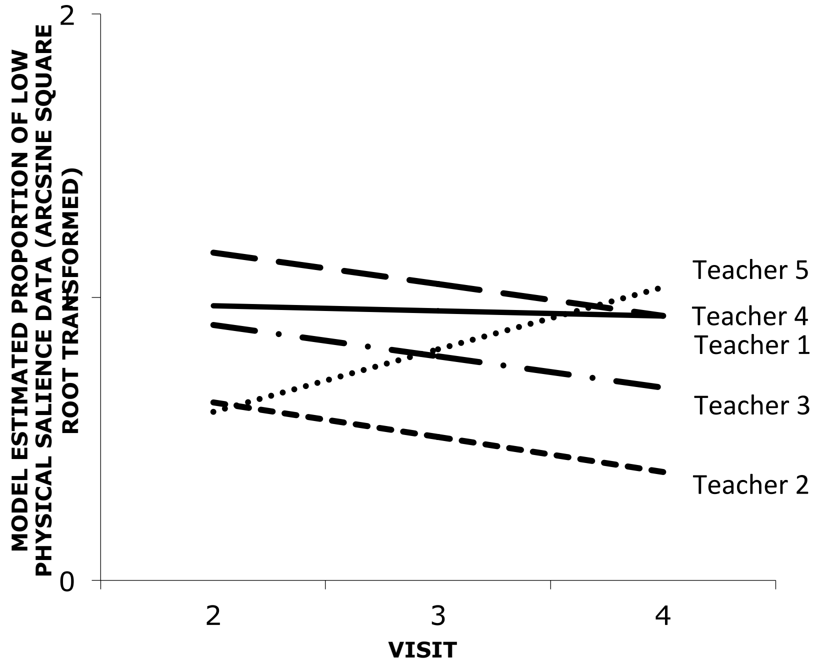


Figure 8. Proportion of (arcsine square root transformed) low physical salience data points collected by prototypical students as a function of visit, by teacher (n=143).

Note in Figure 8 that, much as in previous analyses, a prototypical student (prior knowledge at the sample mean) in teacher 5's class is indistinguishable from a prototypical student in teacher 2's class, when being evaluated on the (arcsine square root transformed) proportion of data they collect on Visit 2 to the EcoMUVE. However, the change over time for the student in teacher 5's class is the opposite of their peer in teacher 2's class (as well as their peers taught by all of the other teachers). By Visit 4, the prototypical student in Teacher 5's class collects a relatively large (1.04) proportion (arcsine square root transformed) of low physical salience data, whereas the student in teacher 2's class's (arcsine square root transformed) proportion is approximately 0.38, a difference that is statistically significant ($p<.05$).

Table 4. Results of individual growth models of the proportion low physical salience data (level 1 or 2, arcsine transformation of the square root) collected over time by middle school students in a multi-user virtual environment (n_teach=5; n_student=143;n_observations=389)

		Model 1. Unconditional Growth Model	Model 2. Model 1 +Controls	Model 3. Model 2 +Predictor	Model 4. Final Model
Fixed Effects					
INTERCEPT	γ_{00}	0.84*** (0.03)	0.56*** (0.08)	0.56*** (0.08)	0.59*** (0.07)
VISIT	γ_{10}	-0.06* (0.03)	0.28*** (0.08)	0.27*** (0.07)	0.22*** (0.06)
PRE_CAUSAL†	γ_{01}			0.09 (0.17)	
PRE_CONTENT†	γ_{02}		0.02** (0.01)	0.02** (0.01)	0.02** (0.01)
FEMALE	γ_{03}		0.06 (0.05)	0.06 (0.05)	
TEACHER1	γ_{04}		0.38*** (0.09)	0.38*** (0.09)	0.37*** (0.09)
TEACHER2	γ_{05}		0.04 (0.09)	0.04 (0.09)	0.03 (0.09)
TEACHER3	γ_{06}		0.31*** (0.09)	0.31*** (0.09)	0.31*** (0.09)
TEACHER4	γ_{07}		0.57*** (0.11)	0.58*** (0.11)	0.56*** (0.11)
PRE_CAUSAL*VISIT	γ_{11}			0.07 (0.17)	
PRE_CONTENT*VISIT	γ_{12}		-0.01~ (0.01)	-0.01~ (0.01)	-0.01~ (0.01)
FEMALE*VISIT	γ_{13}		-0.09 (0.05)	-0.09 (0.05)	
TEACHER1xVISIT	γ_{14}		-0.25** (0.09)	-0.24** (0.09)	-0.24** (0.09)
TEACHER2xVISIT	γ_{15}		-0.35*** (0.09)	-0.35*** (0.08)	-0.34*** (0.09)
TEACHER3xVISIT	γ_{16}		-0.34*** (0.08)	-0.34*** (0.08)	-0.33*** (0.08)
TEACHER4xVISIT	γ_{17}		-0.55*** (0.10)	-0.55*** (0.10)	-0.53*** (0.10)
Variance Components					
Level1 Residual	σ_ε^2	0.1609***	0.1430***	0.1428***	0.1445***
Level 2 Initial	σ_0^2	-0.01616	-0.02980**	-0.02975**	-0.03090**

			*		*
Level 2 Rate of Change	σ_1^2	0.02960***	0.02366***	0.02368***	0.02436***
Goodness of fit					
-2LL		447.9	348.9	348.6	351.7
AIC		457.9	382.9	386.6	381.7
BIC		472.7	433.3	442.9	426.1

† Centered on the sample mean
~p<.1; *p<.05; **p<.01; ***p<.001

Chapter 5: Discussion

My intent in this study was to explore trends in the average levels of physical salience of data selected by students over time in the EcoMUVE pond scenario, and to examine what, if any, effect their pre-intervention tendency to offer explanations containing complex causal components had on those trends. In this section, I describe the overall trends in physical salience of data, unpack the non-relationship between prior causal complexity of causal explanation and trends in the physical salience of data collected, as well as highlight additional key findings of interest. I conclude by outlining limitations of the study.

Overall Trends

Recall that I framed my research questions and related hypotheses in terms of clear expectations of patterns of data collection behaviors. Specifically, I anticipated that the average level of physical salience of data collected by students in the EcoMUVE would decline from visits two through four and that the number of low salience data points collected by each student would decrease over time, thus resulting in an increase in the proportion of low physical salience data points collected. In fact, two of those three trends were exactly the opposite in my sample and fitted models. I highlight the observed trends, de-transformed (in their original units) in Panels A-C of Figure 8. Considering a prototypical student (with sample-average prior content knowledge) chosen at random over Visits 2-4 in the EcoMUVE pond: Panel A shows the trend in average physical salience level; Panel B shows the trend in number of low physical salience data points collected; and Panel C shows the trend in the proportion of data collected that was rated as having low physical salience.

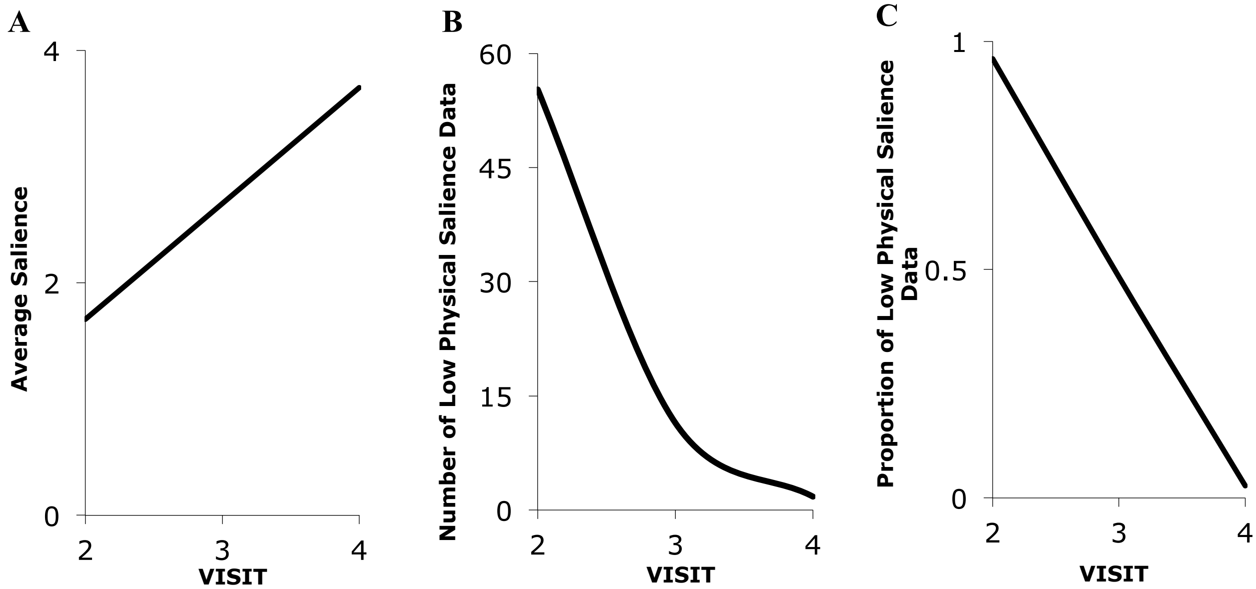


Figure 9. Panel of plots showing the fitted trend of collection of low physical salience data by a randomly chosen prototypical student (prior content knowledge at sample mean) in terms of: A) average salience level of data points, B) number of low physical salience data points, C) proportion of low physical salience data points.

Observing Figure 9, note that the trend in the *absolute* number of low physical salience data points collected by a prototypical student chosen at random declines, as expected based on prior research in student data collection behaviors in MUVEs (Ketelhut, 2007, Grotzer et al, 2013). My assumption that students would collect less physically salient data over time runs completely counter to the modeled pattern in Panel A, however. The positive trend in average salience, together with the reduction in number of low physical salience data collected over time, leads directly to the scenario in Panel C, in which the proportion of low physical salience data collected by a prototypical student chosen at random drops from over 0.90 to nearly 0 by visit 4.

There are a couple of possible reasons that the average salience level of data collected by students actually increased over time. Firstly, the student's avatar is initially placed next to the pond during each visit to the EcoMUVE; this is also truewhenever she changes dates using the calendar tool or exits the submarine tool used to view microscopic organisms in the pond. This might act to initially concentrate their data collection in areas spatially local to the pond. Such areas have a higher concentration of opportunities to collect less-salient data, as defined by the rubric in this study. Research is currently underway to explore trends in data collections patterns in relation to distance from the pond in the EcoMUVE, and should give further context to the data collection patterns observed in this study.

Secondly, the EcoMUVE pond curriculum itself gave instruction on how to use some of the sampling tools early in the lesson sequence. If students were following the lesson plan closely (Appendix A), there would be a higher probability that they would use these tools on the first couple of days they were in the virtual world, resulting in lower average physical salience on that day. Future research tracking student data collection in the EcoMUVE should broaden the scope to more days, while controlling for the types of offline conversations that forced me to look at the three day window used in this study.

Tendency to Offer Explanations Containing Complex Causal Components

I hypothesized that students with higher pre-intervention use of explanations with complex causal relationships would show more focused attentional selection of data in three ways: they would collect data with lower average physical salience scores than their peers with lower understanding, initially and over time; they would initially collect more and reduce the number of low salience data points collected less rapidly over time; and they would initially collect more and increase the proportion of low physical salience data points collected more rapidly over time. Student prior complex causal explanations, as measured, had none of these effects on the data collection behaviors of interest. Why might this be?

As discussed in relation to the overall data collection trends, one reason may be that the curriculum was more prescriptive than envisioned. Students are given free time to collect data as they see fit as they learn to use the tools in the EcoMUVE over the first few days of the curriculum. However, teachers may operationalize this section of the unit in a more stringent manner, directing students to collect certain types of data or use the tools in a prescribed way for longer than intended by EcoMUVEs designers.

Alternatively, the students themselves may have interpreted the curriculum to be more prescriptive, and may have focused their data collection behaviors accordingly. It may also be that the design of the EcoMUVE itself, as discussed above, biased students towards collecting certain types of data over time. Any, or all, of these scenarios would result in a possible severing of the correlation between prior causal explanations and salience-related data collection behaviors. One interesting strand of research currently underway will overlay the analytical methods from this paper with observational data from student interactions with the EcoMUVE, each other, and their teacher. This mixed-methods approach should provide some of the context required to better understand the possible relationship between complex causal explanations and attentional focus based on data salience.

Another possibility is that the connection between the tendency to offer explanations containing complex causal components and the physical salience of data collected by students is sensitive to the operationalization of the question predictor. Recall that the frequency data gathered from the Causal Dynamics Assessment (CDA) was used to create a measure based on the proportion of complex causal responses to total responses. This choice of measure was based on construction of similar measures in a previous study of pre-post student response patterns on the CDA (Grotzer et al, 2013), and due to the favorable distributional properties of the resulting measure. Was this truly the best way to capture tendency to detect complex causal relationships in the current study, however? For example, exploratory analyses with a measure of complex causal understanding framed as a ratio of complex to simple responses resulted in a marginal ($p<.10$) effect of complex causal understanding on average salience and proportion of low physical salience data collected. Similar models using raw count data as the predictor were similar in magnitude and direction, though not significant. Further research should be conducted on the construction and validation of measures specifically intended to measure causal understanding as a top-down salience component before definitive inferences are made regarding the connection between causal understanding and attentional focus of data based on physical salience properties.

Content Knowledge

In contrast to the pre-intervention tendency to offer explanations containing complex causal components, prior content knowledge was related to the initial value of average data salience and the initial value and rate of change of the proportion of low salience data collected over time. This is in line with Ketelhut's (2007) study of student data collection behaviors in the *River City* MUVE. Having higher prior content knowledge might indicate that the student is aware of specific types and locations of critical factors in the virtual environment itself (such as the presence of bacteria in water). This would account for why a student chosen at random with higher prior knowledge initially collects data with lower physical salience compared to peers with lower content knowledge, on average in the population, and continues to do so throughout the course of the unit, even as the average salience levels of collected data rise over time. As previously mentioned, a mixed-methods research study triangulating student understanding through the quantitative methods here, as well as qualitative observational and interview protocols, would make this link more explicit.

Teacher

Perhaps one of the most intriguing findings from this study was the degree of heterogeneity between the fitted trends of prototypical students of different teachers across all three outcomes. Revisiting Figures 4, 6, and 8, Note that the trends exhibited by prototypical students of teachers 1 through 4 are generally of the same magnitude and direction, even if the starting points vary substantially. A prototypical student in teacher 5's classroom, on the other hand, initially starts out similar to students in teacher 2's classroom, but exhibits a trend that runs completely counter to students of the rest of the teachers. This extreme variability leads to a critical question: how much of this variance is due to the actions of each teacher, rather than mental models inherent to their particular students?

One way to potentially explore this question would be to compare key pre-intervention student and teacher-generated variables between students in the classes of teachers 2 and 5, to look for similarities and differences. An ANOVA with follow-up Tukey's Honest Difference analysis found no statistically significant difference between students of teachers 2 and 5 on their pre-intervention content knowledge or causal understanding. Nor were there any differences in how well their teachers expected them to do, or their teachers' estimates of pre-intervention reading level. It should also be noted that both teachers were reported by Metcalf et al. (2013) to have implemented the EcoMUVE curriculum to the same relatively high degree of fidelity (that is to say, they were both observed by researchers to have followed the teacher guide reasonably well.) As such, researchers conducting future quantitative studies of the effect of the EcoMUVE (or any other MUVE) on changes in student behavior, attitude, or understanding should take care to collect as much information as possible about the students and teachers under observation, and take care to collect as much supplemental qualitative data as possible (such as audio and video recording during implementation) to give deeper context to observed differences between students grouped by teacher (or team, class period, school, or other meaningful grouping). Recent qualitative exploration (Courter et al, 2014) of student movement over time in the EcoMUVE pond scenario is an exemplar of the level of overlapping data required for this fuller contextual understanding.

Limitations

My study, though rigorous, has multiple limitations. Chief amongst them is the strong assumption that student behaviors in the EcoMUVE represent authentic data collection events in the virtual world in the service of developing an understanding of the complex causal chain of events that resulted in the fish die off. When students used various data collection tools, such as the camera, they were given the option to save the data, or cancel the data collection event. I took care to only include in my analyses data that were purposefully saved. However, that does not account for the fact that students can notice changing aspects of the world, such as the color of the pond water, that does not generate usable data for this study. As such, my results may be biased downward, since only a sampling the possible variance in physical salience of data collected is represented here. Future replications of this study that identify and quantify further facets of physical salience in the EcoMUVE would be useful in addressing the questions raised by this study.

An additional limitation is the possible lack of validity of my novel outcome variables. It could be argued that my physical salience construct lacks evidence of construct validity, thus calling into question the inferences made in this study. While it is true that no separate validation study has been conducted, I endeavored to bolster the validity of inferences based on the outcomes of this exploratory study in two ways. Firstly, I took care to operationalize the concept of physical (or bottom-up) salience as closely as possible with past research in the field of engineering psychology (e.g., Wickens et al, 2003; Wickens, et al, 2009), in which instruments on airplane display and control panels were assigned different salience values based on their visibility, location, and importance for a given routine. Further, I applied Clarke's (2009) rubric of data salience in MUVEs as close to verbatim as possible, making some changes to account for differences between the EcoMUVE and River City virtual environments. Finally, the EcoMUVE was designed with the visibility of certain types of data in mind, and I followed this design schema during the coding process. That being said, I strongly recommend further studies to validate the bottom-up and top-down constructs of data salience in MUVEs further.

Chapter 6: Final Thoughts

I undertook this exploratory study to shed light on the relationship between student pre-intervention tendency to use complex causal explanations and the attentional selection of data based on the physical salience properties of said data. In the process, I generated further questions and developed key methodological insights relating to the study of student behavior and learning in MUVEs. As in past research (e.g., Ketelhut, 2007; Clarke, 2009) I was able to leverage the longitudinal nature of data generated by student use of the MUVE to model patterns of their behaviors and draw inferences accordingly. I found that pre-intervention tendency to use complex causal explanations, as measured here, does not directly relate to changes in the attentional selection of data in the EcoMUVE, likely due to the prescriptive nature of the EcoMUVE Pond curriculum used for the study. This is important, since activation and use of that knowledge should be a critical component of determining what types of data students select. Adjustments to the curriculum, perhaps allowing for more open-ended exploration of the world, might allow for this.

Just as importantly, however, this study has made salient (no pun intended) two factors that should simultaneously guide future MUVE-related research and provide a lens through which to evaluate the findings of past work. Though I believe that purely quantitative explorations such as this provide critical insights and help to identify trends, throughout the discussion of the findings and limitations of the current study I repeatedly highlight the need for additional types of data to triangulate and fully understand the linkages between the observed behaviors and latent traits. I echo Dukas' (2009) suggestion that single types of MUVE-based data, no matter how multi-dimensional, are often insufficient to make such inferences. I thus strongly encourage future research of MUVEs and other complex technology-enhanced curricula (such as Augmented Reality-based units) utilize mixed methodologies such as in Bressler & Bodzin (2013) and Bressler (2014).

In addition to the forward-looking need for more mixed-methods research, my findings also provide a critical lens through which to view current and past quantitative studies seeking to infer latent student traits from complex behavioral data. In particular, the degree of heterogeneity in initial value and rates of change of the physical salience of data collected over time in the EcoMUVE would make it very difficult to make generalizable inferences using automatic data-mining routines (detectors) designed to recognize patterns of behaviors as indicators of important constructs such as boredom or engagement (e.g., Baker et al., 2008). Even in study such as this, where a relatively small number of teachers are clustered into two very similar schools in the same district, the degree of between-teacher variability in trends would make it nearly impossible to generalize the use of a detector trained in one teacher's class to assess students of another teacher. This is especially true if the teacher's students' behaviors deviated wildly from those of other teachers' over time, such as teacher 5's students in the current study. As such, the common assumption that detectors, and the inferences drawn from their use, are generalizable within schools or districts with similar socio-economic profiles is suspect and should be tested and validated with further research.

Ultimately, this study highlights the need for continued research in complex technology-supported learning environments such as MUVEs and Individual Virtual Environments (IVEs). Both MUVEs and IVEs demonstrate great potential as powerful spaces for learning and assessment (Nelson, Ketelhut, & Schifter, 2010). But this power comes with a responsibility to bring to bear as many complementary analytical techniques as possible when exploring their potential as transformative pedagogical tools.

Bibliography

Baker, R.S.J.d., Corbett, A.T., Roll, I., Koedinger, K.R. (2008) Developing a Generalizable Detector of When Students Game the System. *User Modeling and User-Adapted Interaction,* 18(3): 287-314.

Bishop, C.M., (2006). *Pattern Recognition and Machine Learning*. New York, NY: Springer Science+Business Media, LLC.

Börner, K., & Penumarthy, S. (2003). Social Diffusion Patterns in Three-Dimensional Virtual Worlds. *Information Visualization*. 2(3),182-198.

Bressler, D.M. (2014). Is it all in the game? Flow experiences and scientific practices during an INPLACE mobile game. *Unpublished Doctoral Dissertation, Lehigh University.*

Bressler, D.M., & Bodzin, A.M. (2013). A mixed methods assessment of students' flowexperiences during a mobile augmented reality science game. *Journal of Computer Assisted Learning. doi: 10.1111/jcal.12008*

Chen, J. & Metcalf, S.J. (2013). Motivation and Beliefs about the Nature of Scientific Knowledge Within an Immersive Virtual Ecosystems Environment. *American Educational Research Association (AERA) Conference*, San Francisco, CA, Spring 2013.

Cheng, P.W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, 104(2): 367-405.

Cheng, P.W., Novick. L.R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology*, 58(4), 545-567.

Clarke, J. (2009). Exploring the complexity of inquiry learning in an open-ended problem space. *Unpublished doctoral dissertation, Harvard Graduate School of Education.*

Colosi, J. C., & Zales, C. R. (1998). Jigsaw Cooperative Learning Improves Biology Lab Courses. *Bioscience, 48*(2), 118-124.

Courter, C.J, Grotzer, T.A., Derbiszewska, K.M., Powell, M., Tutwiler, M.S., Kamrarainen, A., Metcalf, S., Dede, C. (2014). A case study contrasting students' explorations behaviors of a complex causal scenario in a virtual world: Mapping proximity and goal-directed behaviors in EcoMUVE. *National Association for the Research of Science Teaching (NARST) Conference,* Pittsburgh, PA, Spring 2014.

De Baz, T. (2001). The Effectiveness of the Jigsaw Cooperative Learning on Students' Achievement and Attitudes toward Science. *Science Education International, 12*(4), 6-11.

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

Dede, C., Nelson, B., Ketelhut, D. J., Clarke, J., & Bowman, C. (2004). Design-based research strategies for studying situated learning in a multi-user virtual environment. *Paper presented at the International Conference on Learning Sciences,* Mahweh, NJ.

Dukas, G. (2009). Characterizing student navigation in educational multiuser virtual environments: A case study using data from the River City project. *Unpublished Doctoral Dissertation, Harvard Graduate School of Education.*

Glymour, C. (1998). Learning Causes: Psychological Explanations of Causal Explanation. *Minds and Machines.* 8(1):39-60.

Gopnik, A., Sobel, D.M., Schulz, L E., Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, 37(5): 620-629.

Griffiths, T.L., Tenenbaum, J.B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, 51(4):334-384.

Grotzer, T.A. (2012). *Learning causality in a complex world: Understandings of consequence.* Lanham, MD: Rowman & Littlefield.

Grotzer, T.A., Basca, B., & Donis, K. (2002). *Causal patterns in ecosystems: Lessons to infuse into ecosystems units.* Cambridge, MA: President and Fellows of Harvard College.

Grotzer, T.A. & Derbiszewska, K. (2011). *Causal Dynamics Assessment.* Cambridge, MA: President and Fellows of Harvard College.

Grotzer, T.A., Kamarainen, A., Tutwiler, M.S, Metcalf, S, & Dede, C. (2013) Learning to reason about ecosystems dynamics over time: The challenges of an event-based causal focus. *BioScience*. 63(4):288-296.

Hume, D. (1777). Enquiries Concerning the Human Understanding, and Concerning The Principles of Morals. L.A.Selby-Bigge (Ed) 1902.

Itti, L., Koch, C., & Niebur, E. (1998, November). A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Learning*, 20(11).

Jenkins, H.M., Ward, M.C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs: General and Applied*, 79(1), 1-17.

Johnston, W. (2014) Personal communication on 11 April 2014.

Kafai, Y., Queintero, M., Feldon, D. (2010) Investigating the 'Why' in Whypox: Causal and Systematic Explorations of a Virtual Epidemic. *Games & Culture*. 5(1):116-135.

Kamarainen, A.M., Metcalf, S., Grotzer, T. & Dede, C. (2012). EcoMUVE: Shifts in Affective Beliefs and Values About Science Through Learning Experiences in Immersive Virtual Environments. *American Educational Research Association (AERA) Conference*, Vancouver, BC, Canada. April, 2012.

Ketelhut, D.J. (2007). The Impact of Student Self-efficacy on Scientific Inquiry Skills: An Exploratory Investigation in River City, a Multi-user Virtual Environment. *Journal of Science Education & Technology,* 16(1), 99-111.

Ketelhut, D., Nelson, B., Clarke, J., Dede, C. (2010). A Multi-user virtual environment for building and assessing higher order inquiry skills in science. *British Journal of Educational Technology*, 41(1):56-68.

Kirchner, J. (2001). *Data Analysis Toolkit #3: Tools for Transforming Data*. Retrieved from http://seismo.berkeley.edu/~kirchner/eps_120/Toolkits/Toolkit_03.pdf on 01 April 2014.

Kushnir, T., Gopnik, A. (2007). Conditional probability versus spatial contiguity in causal learning: Preschoolers use new contingency evidence to overcome prior spatial assumptions. *Developmental Psychology,* 43(1):186-196.

Lu, H.J., Rojas, R, Beckers, T., & Yuille, A.L. (2008). Sequential causal learning in humans and rats. *Proceedings of the 30th Annual Conference of the Cognitive Science Society*. July 2008.

Lu, H., Yuille, A., Liljehold, M., Cheng, P.W., Holyoak, K.J. (2007). Bayesian Models of Causal Judgment: A Comparison. In D. S. McNamara & G. Trafton (Eds.), *Proceedings of the Twenty-ninth Annual Conference of the Cognitive Science Society* (pp. 1241-1246). Austin, TX: Cognitive Science Society.

Melloni, L., van Leeuwen, S. Alink, A., Muller, N.G. (2012). Interaction between bottom-up saliency and top-down control: how saliency maps are created in the human brain. *Cerebral cortex, 22(12):2943-2952.*

Metcalf, S. & Kamarainen, A.M. (2011). *EcoMVE content knowledge assessment.* Cambridge, MA: President and Fellows of Harvard College.

Metcalf, S., Kamarainen, A.M., Grotzer, T. & Dede, C. (2012). Teacher Perceptions of the Practicality and Effectiveness of Immersive Ecological Simulations and Classroom Curricula*. American Educational Research Association (AERA) Conference, Vancouver, BC, Canada.* April, 2012.

Metcalf, S., Kamarainen, A., Tutwiler, M.S., Grotzer, T., & Dede, C. (2011). Ecosystem science learning via multi-user virtual environments. *International Journal of Gaming and Computer-Mediated Simulations*, 3(1) 86-90.

Metcalf, S., Tutwiler, M.S., Kamarainen, A., Grotzer, T., & Dede, C. (2013) Multi User Virtual Environments to Promote Middle School Ecosystem Science Learning and Attitudes about Science. *AERA, San Francisco, CA*, Spring 2013.

Nelson, B. (2007). Exploring the use of individualized, reflective guidance in an educational multi-user virtual environment. *The Journal of Science Education and Technology* 16(1): 83-97.

Nelson, B., Ketelhut, D.J., and Schifter, C. (2010). Exploring Cognitive Load in Immersive Educational Games: The SAVE Science Project. *International* Journal *for Gaming and Computer Mediated Simulations* 2 (1): 31-39.

Neopolitan, R.E. (2009). *Probabilistic Methods for Bioinformatics with an Introduction to Bayesian Networks.* Morgan Kaufmann Publishers, Boston, MA.

Novick, L.R., Cheng, P.W. (2004). Assessing interactive causal influence. *Psychological Review,* 111(2):455-85.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufman, San Mateo, Ca.

Pearl, J. (2000). *Causality.* Cambridge University Press, New York, NY.

Penumarthy, S. and Börner, K. (2006). Analysis and Visualization of Social Diffusion Patterns in Three-Dimensional Virtual Worlds. In Schroeder, Ralph & Axelsson, Ann-Sofie (Eds.), *Computer Supported Cooperative Work: Avatars at Work and Play: Collaboration and Interaction in Shared Virtual Environments*, Springer-Verlag, Vol. 34, chapter 3, pp. 39-61.

Perkins, D.N. & Grotzer, T.A. (2000, April). Models and moves: Focusing on dimensions of causal complexity to achieve deeper scientific understanding. American Educational Research Association (AERA) Conference, New Orleans, LA.

Spirtes, P, Glymour, C., Scheines, R. (1993). *Causation, prediction, and search*. Springer-Verlag, New York, NY.

Sobel, D.M., Tenenbaum, J.B., Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science* 28(3): 303–333.

Schulz, L.E., Gopnik, A. (2004). Causal Learning Across Domains. *Developmental Psychology*, 40(2):162-176.

Schulz, L.E., Sommerville, J. (2006). God Does Not Play Dice: Causal Determinism and Preschoolers' Causal Inferences. *Child Development*, 77(2): 427–442.

Shannon, C.E. (1948). "A Mathematical Theory of Communication", *Bell System Technical Journal,* 27: 379–423 & 623–656, July & October, 1948.

Shultz, T. (2007). The Bayesian revolution approaches psychological development. *Developmental Scienc*e 10(3), 357-364.

Singer, J.D., & Willett, J.B., (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence.* New York, NY: Oxford University Press.

Snow JC, Allen HA, Rafal RD, Humphreys GW (March 2009). "Impaired attentional selection following lesions to human pulvinar: evidence for homology between human and monkey". *Proc. Natl. Acad. Sci. U.S.A.* 106 (10): 4054–9. doi:10.1073/pnas.0810086106. PMC 2656203. PMID 19237580.

Tsakanikos, E. (2004). Latent inhibition, visual pop-out and schizotypy: is disruption of latent inhibition due to enhanced stimulus salience? *Personality and Individual Differences*, 37, 1347-1358.

Tutwiler, M.S. (2013). The Interaction of Causal Bayes Nets and Information Theories in a Multi-User Virtual Environment. *Unpublished Qualifying Paper. Harvard Graduate School of Education.*

Tutwiler, M.S., Grotzer, T. (2012) Irreducible Complexity: How Do Causal Bayes Nets Theories of Human Causal Inference Inform the Design of a Virtual Ecosystem?

International Conference on Learning Sciences, Sydney, Australia, Summer 2012.

Tutwiler, M.S., Grotzer, T.A., Kamarainen, A., Metcalf, S. Dede, C. (2011). Exploring student understanding of complex causality in an ecosystems-based multi-user virtual environment *Conference on Computer-Supported Collaborative Learning*, Hong Kong, Summer 2011.

Tutwiler, M.S., Grotzer, T., Kamarainen, A., Metcalf, S., & Dede, C. (2013) Exploring Virtual Worlds: Causal Understanding and Data-Collection Behaviors in an Ecosystems-Based Multi-User Virtual Environment. *National Association for the Research of Science Teaching (NARST) Conference*, San Juan, Puerto Rico, Spring 2013.

Walther, D., Itti, L., Reisenhuber, M., Poggio, T., & Koch., C. (2002). Attentional selection for object recognition—a gentle way. *Proc., 2nd workshop on biologically motivated computer vision* BMCV200: 472–479.

Wickens, C. D., Goh, J., Helleberg, J., Horrey, W. J., & Talleur, D. A. (2003). Attentional models of multitask pilot performance using advanced display technology. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 45(3), 360-380.

Wickens, C. D., Helleberg, J., Goh, J., Xu, X., & Horrey, W. J. (2001). *Pilot task management: Testing an attentional expected value model of visual scanning*. Savoy, University of Illinois Institute of Aviation.

Wickens, C. D., & McCarley, J. S. (2001). Attention-situation awareness (A-SA) model of pilot error (*Final Technical Report ARL-01-13/NASA-01-6*). Savoy, IL: University of Illinois, Aviation Research Lab.

Wickens, C., McCarley, J., & Steelman-Allen, K. (2009, October). NT-SEEV: A model of attention capture and noticing on the flight deck. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 53, No. 12, pp. 769-773). SAGE Publications.

Appendix A: The EcoMUVE Pond Unit

The EcoMUVE Pond Unit

The EcoMUVE Pond unit is a multi-day middle school curriculum designed by Drs. Tina Grotzer and Chris Dede at the Harvard Graduate School of Education. The purpose of the unit is to concurrently teach students aspects of ecosystems science and complex causal understanding via interactions within a Multi-User Virtual Environment (MUVE). In order to solve a mystery (the death of the large fish in a pond), students work together in teams of four and take on the roles of various types of experts in order to navigate and collect data in the virtual world. While in the virtual world, the students, through their virtual embodiments known as avatars, interact with virtual agents (Figure A1), various types of virtual flora and fauna (Figure A2), and each other.

The Virtual Ecosystem

The ecosystem in the EcoMUVE Pond unit centers on a small pond flanked by a golf course and housing development (Figure A3). The pond is populated by numerous macroscopic and microscopic species of plants and animals. The area immediately surrounding the pond also contains numerous plants and animals. Over time, students are meant to notice shifts in the populations of the large-scale flora and fauna, as well as varying levels of microscopic biotic and abiotic factors.

In addition to the changing populations of animals and plants, students will also encounter various virtual agents, or non-player-characters (NPCs). These NPCs give students information of varying degrees of quality and helpfulness. The students must use judge the validity of the given testimony and weigh it against other data that they have collected in order to fully understand the fish die-off event.

Tool Use and Data Collection

In order to collect the necessary data, students use scientific tools across various points of time in the virtual world. For example, students are able to measure the turbidity of the water in the pond (Figure A4) or move back and forth through time using a calendar tool (Figure A5). The data that students collect are then stored in a table (Figure A6) that can be shared with their teammates in order to develop plots to assess trends over time (Figure A7).

Each student experiences the EcoMUVE pond unit slightly differently. Some students meticulously collect all of the possible data on each given calendar day within the world, while others strictly follow the suggested data collection points associated with their chosen role. At the end of the unit, the teams come together as a class again and share their causal hypothesis, supported with evidence, via a concept map.

Daily Schedule

In addition to exploring and collecting data in the virtual world, students work together during class-based activities as well. These activities support student learning of key content within the EcoMUVE, and are interspersed throughout the two-week curriculum. What follows is the ten-day schedule that teachers in the current study were asked to follow while using the EcoMUVE pond curriculum with their students.

Day 0. Pre-intervention surveys of student complex causal framing, content knowledge, and science attitudes were administered. Teachers completed surveys of student reading level and their expectation that student would succeed at solving the scientific mystery within EcoMUVE.

Day 1. Students are introduced to the EcoMUVE Pond. They are allowed to explore the virtual world, and are asked to take pictures of as many organisms as possible.

Day 2. Students are introduced to the data collection tools in the EcoMUVE and are asked to explore the virtual world across different calendar days, making observations as they do. Student attention is direction toward the fish die-off on July 28th. The students are tasked with finding out why the fish died.

Day 3. Students are split into multiple groups within their classroom and travel to different stations to learn about different water measurement tools.

Day 4. Students form teams of four and discuss ideas about what caused the fish to die. Each student takes on a role (microscopic specialist, private investigator, water chemist, meteorologist) and returns to the EcoMUVE to collect more data.

- The microscopic specialist collects data using the submarine tool over time
- The private investigator talks to virtual residents in the EcoMUVE Pond unit over time
- The water chemist uses the appropriate tools to take measurements in the pond (e.g., turbidity and pH) over time
- The meteorologist gathers information about the weather each day in the EcoMUVE Pond

Day 5. Students continue to collect data. They are encouraged to coordinate efforts via the online chat function and share data within their team.

Day 6. Students are presented a case study about an ecosystem collapse scenario, and are asked to work within their teams to build concept maps to explain what is happening within the EcoMUVE that might result in the fish die-off.

Day 7. Misconceptions from team concept maps are addressed, the importance of evidence, data, and graphs is discussed, and students are allowed to collect more data within the EcoMUVE. Students are asked to revise their concept maps highlighting the reasons for proposed connections and data sources that support them.

Day 8. Students finish data collection within the EcoMUVE and prepare a presentation outlining their hypothesis about the cause(s) of the fish die-off event.

Day 9. Each team presents their explanation and supporting evidence. As a class, students evaluate which evidence was most important and discuss why the actual cause was so difficult to determine. Students are asked to reflect on other examples of complex causal patterns such as those found in the EcoMUVE.

Day 10. Students complete the post-intervention assessments of causal framing, content knowledge, and attitude.

Figure A1. Ranger Susan, a virtual agent in the EcoMUVE Pond unit.

Figure A2. A predacious diving beetle as seen from the submarine tool in the EcoMUVE Pond unit.

Figure A3. The EcoMUVE Pond with tools displayed at the bottom of the screen.

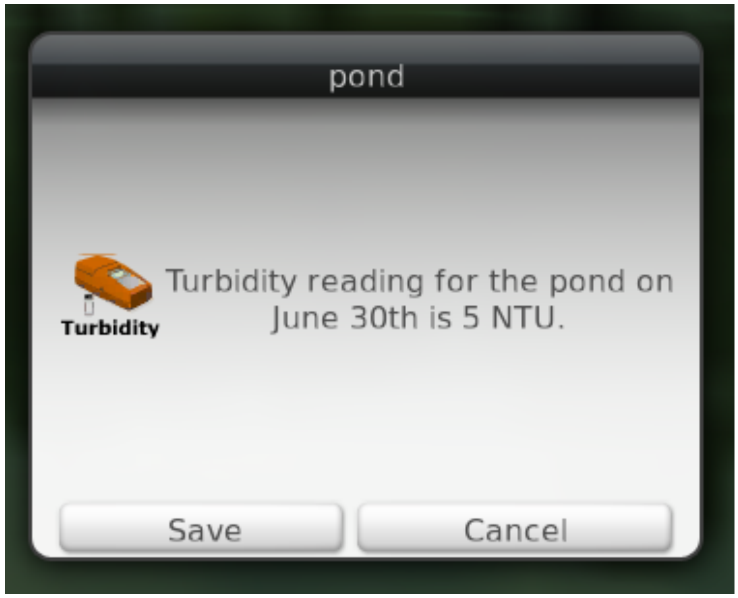
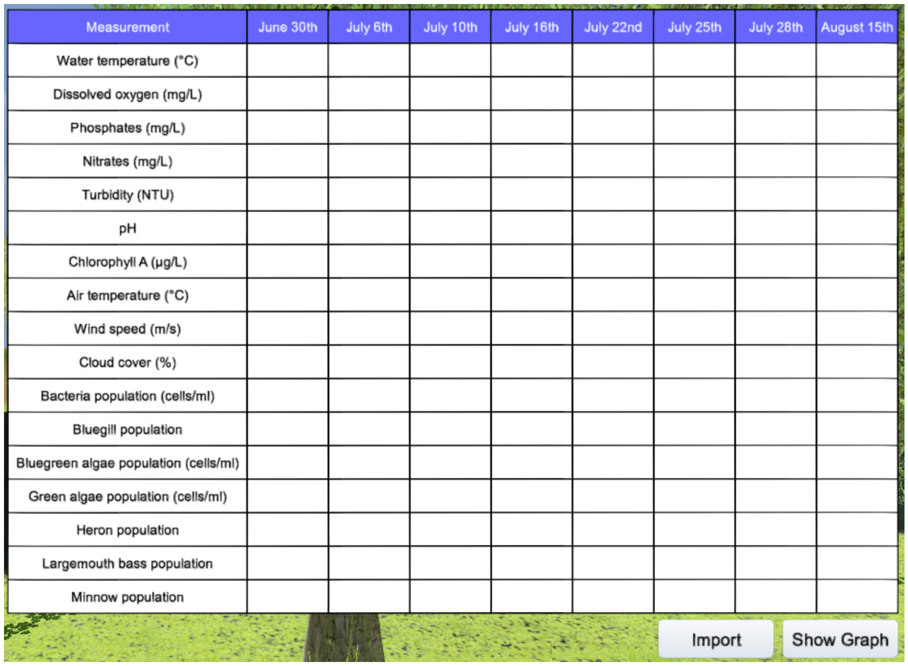


Figure A4. Result of a turbidity reading in the EcoMUVE Pond unit.

Figure A5. The EcoMUVE calendar tool.



Measurement	June 30th	July 6th	July 10th	July 16th	July 22nd	July 25th	July 28th	August 15th
Water temperature (°C)								
Dissolved oxygen (mg/L)								
Phosphates (mg/L)								
Nitrates (mg/L)								
Turbidity (NTU)								
pH								
Chlorophyll A (µg/L)								
Air temperature (°C)								
Wind speed (m/s)								
Cloud cover (%)								
Bacteria population (cells/ml)								
Bluegill population								
Bluegreen algae population (cells/ml)								
Green algae population (cells/ml)								
Heron population								
Largemouth bass population								
Minnow population								

Figure A6. The EcoMUVE Pond data table.

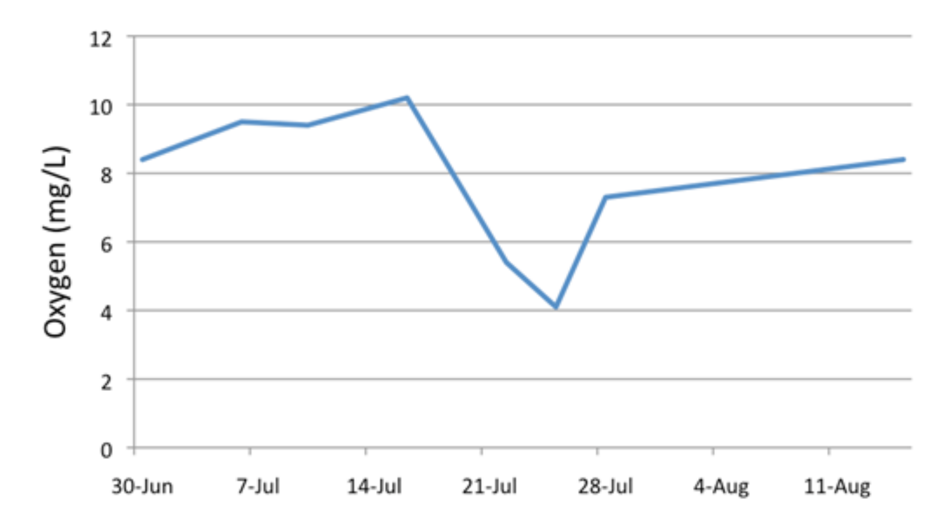


Figure A7. Change in dissolved oxygen over time within the EcoMUVE Pond unit.

Appendix B: Physical Salience Scoring Rubric

	1	2	3	4
Location	Deep under water (requires submarine tool)	Spatially distant (e.g., near golf course or housing development)	Near the pond, but more than a few steps away (e.g., near the drain pipe)	Spatially local (within a few steps of the pond)
Visibility	Not visible without a data collection tool (e.g., pH reading).	Under water (macroscopic) or hidden by fauna (e.g., fox)	Clearly visible up close (e.g., fertilizer bags)	Clearly visible at a distance (e.g., ducks swimming on the pond)

Data points were rated along these the dimensions of location and visibility, with those scores then summed and averaged to yield the physical salience score for each possible unit of datum that students could collect. Some measurements, such as water quality measurements and weather measurements, were not given location scores due to that facet not contributing to their physical salience.

VITA

M. Shane Tutwiler

1996-2002	Petty Officer First Class United States Navy, USA	
2004-06	Temple University Philadelphia, Pennsylvania	B.S. May 2006
2006-07	Graduate School of Education Harvard University	Ed.M. May 2007
2007-09	Science Faculty Avon Old Farms School, Connecticut	
2009-14	Doctoral Candidate Graduate School of Education Harvard University	
2009-13	Teaching Fellow Graduate School of Education Harvard University	
2009-14	Research Assistant Graduate School of Education Harvard University	
2011-12	Mathematics Faculty Taipei American School, Taiwan	
2013-14	Adjunct Professor School of Education University of St. Joseph	
2014-15	Instructor in Education Graduate School of Education Harvard University	