The Dissertation Committee for Emily Katherine Gaertner Certifies that this is the approved version of the following dissertation:

A Computer Model for Learning to Teach: Proposed Categorizations and Demonstrated Effects

Committee:

Walter M. Stroup, Supervisor

James P. Barufaldi

Leema K. Berland

Susan B. Empson

P. Uri Treisman

A Computer Model for Learning to Teach: Proposed Categorizations and Demonstrated Effects

by

Emily Katherine Gaertner, B.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2013

Dedication

When through fiery trials thy pathways shall lie,

My grace, all sufficient, shall be thy supply;

The flame shall not hurt thee; I only design

Thy dross to consume, and thy gold to refine.

To the one who first distinguished gold from dross.

Acknowledgements

Acknowledgements pages usually consist of a list of names that the reader has never heard of, associated with tedious-sounding tasks that the reader doesn't care about, and capped off by the obvious: this author owes a lot to her parents, her significant other, and her dog/computer/supplier of caffeine/etc. Well, I don't drink coffee and my dogs are more hindrances than aids to writing. And while I do owe an indescribable debt to many individuals, it is perhaps more real to recognize the diversity of support than to pretend to a monotonous positivity:

Thank you for the times you doubted me. It was a valuable spur.

Thank you for the demands you made upon me. I re-engaged with life.

Thank you for your unshakeable confidence in my success. Your beacon shone when I wandered in the dark.

Thank you for removing logistical obstacles from my path. Sometimes, it is all just too much to juggle.

Above all, thank you for valuing the endeavor, as well as the outcome.

A Computer Model for Learning to Teach: Proposed Categorizations and Demonstrated Effects

Emily Katherine Gaertner, Ph.D.

The University of Texas at Austin, 2013

Supervisor: Walter M. Stroup

With the proliferation of new technological alternatives to the traditional classroom, it becomes increasingly important understand the role that innovative technologies play in learning. Computer environments for learning to teach have the potential to be innovative tools that improve the skill and effectiveness of pre-service and in-service teachers. There is a tacit sense in such environments that "realism" is best created through, and associated with, a kind of pictorial literalism. I designed a computer model (the *Direct Instruction tool*) that, though simple, appears realistic to many users and thus contradicts that sense of literalism. I also propose a theoretical classification of computer representations based on the relationship (or lack thereof) between perceived usefulness or relevance and realism. In this study, I investigate two questions: 1) What are the kinds of claims or insights that respondents generate in relation to using the *DI tool* to organize their experiences? 2) How do the functionalities of the *DI tool* fit with or support what respondents see as meaningful? Results indicate that a model can be seen as relevant and useful even if it is not internally consistent. Two major themes that were meaningful to study participants were the simultaneously positive and negative role of "difficulty" in the classroom, and the balance between past performance and future potential. The *DI tool* seems to promote a shared focus on these themes despite the

diversity of past educational experiences among study participants. Responses to this model suggest that extremely abstracted representations of teaching are able to influence the claims and insights of users, affording a glimpse into the internal realities of pre-service teachers. This in turn creates an opportunity to articulate these alternative realities without judgment, describe them with respect, and make them an object of consideration rather than a hidden force. The results of this study contribute to a theory of computer environments for learning to teach that can shape the effective use of these tools in the present, as well as accommodate new models that may be developed as technologies change in the future.

Table of Contents

Chapter 1: Introduction

With the proliferation of new technological alternatives to the traditional classroom, it becomes increasingly important understand the role that innovative technologies play in learning. Some proposed educational technologies constitute true innovation, allowing students and teachers to engage in patterns of exploration or discourse that were previously unavailable. Many technologies, however, change the form of a learning activity without altering the underlying functionality. For example, ITunes U and Kahn Academy essentially reproduce the model of "transmission" of information from teacher to learner. Even massive open online courses, though they revolutionize the scale at which students can access a particular instructor, are organized around a lecture model of teaching and learning. Modern computing tools in most homes, and the interconnectedness of the internet, have changed the context of the listening, but the fundamental presumed learning mode remains the same.

Not all technological change is incremental or merely substitutive, though. Some technologies are substantively innovative. Within educational technology in particular, one can distinguish between three types of technological change:

- Substitution maintains existing relationships and patterns of discourse, but shifts some actions into the digital realm (for example, a video recording instead of a live lecture).
- Delegation introduces a triangle between the teacher, the learner and the technology, but the technology plays a secondary role (for example, conditional feedback from an online multiple-choice quiz).
- Innovation equalizes the triangle; computer modeling, for example, can serve as a "way of thinking, a means of expression, and subject of investigations" (Riley,

2007). The innovative technology is integral to the activity that includes it; there exists either no previous analog, or only one that is extremely simplistic relative to the new capabilities.

The evaluation of innovation, however, still depends on local values and desired outcomes. Each of these types of change can give rise to a large number of possible new technological tools for teaching and learning. Not all such tools are equally beneficial.

Computer environments for learning to teach have the potential to be innovative tools that improve the skill and effectiveness of pre-service and in-service teachers. These computer programs are especially difficult to define, categorize, or understand, however. Very few examples of these types of computer environments exist, due in part to widespread skepticism that machine-based interactions can adequately copy human-to-human interactions. As I demonstrate with this work, thinking of a computer environment for learning as a "copy" in the first place prevents us from taking full advantage of the revolutionary possibilities of new technologies. The goal of such software should be to get users to explore claims about teaching and learning, rather than to create a surface "likeness."

The results of this study contribute to a theory of computer environments for learning to teach that can shape the effective use of these tools in the present, as well as accommodate new models that may be developed as technologies change in the future. In particular, the analytical focus around these types of models needs to be on making sense of teaching – an engagement with the logic of teaching – not the visual similarities that might matter more in game design per se. What makes a model "about teaching" becomes especially important as it implicates moving beyond one-to-one correspondence with picture-like aspects of a classroom; we set aside the challenge of programming hair

to move realistically in favor of the challenge of capturing essential elements of classroom events.

MODELING HUMAN-TO-HUMAN INTERACTIONS

What I had not realized is that extremely short exposures to a relatively simple computer program could induce powerful delusional thinking in quite normal people.

(Weizenbaum, 1976, p.7).

This thesis engages questions related to how users' responses reveal what kind of "similarity to reality" matters for a simulation of instruction. The reflective, revealing nature of users' responses to the computer tool in this study, which relate to what aspects of reality can be observed to matter to users, invites comparison to another computer simulation. The program *ELIZA* (Weizenbaum, 1966), written in the mid-1960's as part of research into computer recognition and use of human language, generated surprising responses from the people who experienced it. One version of *ELIZA* was intended as a near-parody of non-directive therapy associated with Carl Rogers (Rogers, 1951). *ELIZA* used simple matching of word-strings with pre-programmed responses to "conduct" the simulated therapy session. The Rogerian approach to therapy was chosen for simulation because it required no database of outside knowledge. Notably, there was no attempt to simulate a therapy environment beyond the typed input from the human user and the typewritten output from the program (Weizenbaum, 1976). And in reality, the program does not respond to the meaning of the input. Instead, the presence or absence of certain key phrases trigger formulaic responses that themselves contain phrases that may have their origin in the user-generated input. Weizenbaum was primarily exploring issues related to the then-nascent field of Artificial Intelligence, and not at all interested in developing a usable or "realistic" therapeutic tool.

Despite the highly contrived and constrained nature of the simulation, Weizenbaum found that users of the program quickly became emotionally involved with the "therapeutic" experience - even those users who were part of the research team. Users treated the interactions as meaningful, even knowing that the program had no semantic processing capabilities. So engaged were the users that some complained that recording transcripts of their chats with the program would violate their privacy and would be inappropriately intrusive.

A second outcome was that some formally trained, professional therapists – people who might be expected to know better – saw so much consistency with Rogerian therapy that they predicted that *ELIZA*-like programs would take over therapy on a broad scale. This was particularly surprising because it implied that human therapists seemed to be making the judgment that their profession is no more effective than a remarkably simple computer program. It also implies a disturbingly mechanistic view of patients – as equally simple to help – that was in fact borne out by the responses of users (Weizenbaum, 1976, p.6).

The *Direct Instruction tool* (*DI tool*), a computer program that I designed, may be another example of, to echo Weizenbaum's words in the epigram, a "delusion-inducing" program. Similar to Weizenbaum's near-parody of therapy, the *DI tool* was developed as something like a parody of direct instruction. Unspecified bits of information are broadcast to learners, and a simple numerical comparison determines whether each learner "gets" a particular bit of information. One additional rule updates the learner's numerical state based on their success or failure for each broadcast bit. The simulated teacher is ostensibly the source of information, but plays no additional role.

Yet, the simulation did not serve in ways consistent with the original parodying intent of the *DI tool*. Instead, users embraced the idea that the model could and perhaps

should be used to guide instructional decision-making, despite its simplicity. This thesis is an attempt to characterize the apparent credibility and utility this environment is observed to have.

Part of the reason such an investigation might matter – beyond being another instance of mismatched intent and actual use – is that we live in a world of mixed digital and analog elements. Most of us are much more comfortable with digital tools than the users of *ELIZA* were in the mid-1960's. If *ELIZA* could induce a misplaced warrant in people not immersed in a digital world, the possibilities for similar "delusion" now are much greater. Partial digital copies of the real world are ubiquitous, and their very familiarity makes them difficult to parse and critique. Understanding some of the ways people can be drawn in by even simple, virtually real, environments is important because it allows us to either harness or protect ourselves against this distorting process. And on the positive side, we might begin to better understand what could matter in developing computer environments that are intended to function in particular ways. For example, one tension that has already arisen in the design of computer environments for learning to teach is the relative importance of pictorial literalism or "visual likeness" versus engagement with the "logic" of the interactions depicted.

Based on preliminary results, which made clear that, like *ELIZA*, the *DI tool* was being treated as realistic and informative, it seems to be an appropriate time to investigate this "virtual as real" process more systematically. This work is done with the intent to characterize *what* kinds of engagement are facilitated and *how* that gets enacted relative to the simulation tool itself. By way of contrast with the *DI tool*, several computer programs exist that purport to replace in-class practice for future teachers, streamlining the process of learning how to teach. The program *simSchool*, for example, is described as a "virtual apprenticeship" and a "flight simulator" (Zibit & Gibson, 2005). Both

descriptions assume that increasing the realism of the program will result in increased value or utility.

There is also a tacit sense in such environments that "realism" is best created through, and associated with, a kind of pictorial literalism. The classroom must look like a real classroom, and this level of visual realism should extend down to details like hair on simulated students' heads stirring in the breeze from a simulated open window. This pictorial way of construing what it means to be "realistic" structures expectations for how a simulation of classroom interactions should work. Part of the warrant for this thesis comes from the contrast between this expectation and the high level of user engagement observed with seemingly unrealistic environments like *ELIZA* and the *DI tool*.

It is important to evaluate possible technological tools for learning to teach from a critical perspective rather than simply assuming that existing paradigms for simulations (being more, or less, visually realistic) will apply to representations of teaching as well. This work is structured to draw out contrasting senses of what it might mean for simulations to be realistic in ways that instantiate and support user engagement with practices associated with teaching. The focus is on *what* responses emerge in relation to use of the simulation and then on *how* the engagement is related to the representation capabilities of the *DI tool*.

MY PERSONAL CRITICAL PERSPECTIVE AND ITS CONNECTION TO THIS WORK

Three examples may help to situate my own critical perspective – how I understand what is at stake – as it relates to my involvement in this work.

Imagine a clear cylinder, 2 feet long and 8 inches in diameter, mounted on a horizontal axle so you can tumble the contents. You are idly spinning the cylinder and

enjoying the shifting patterns made by the tiny colored beads inside. A nearby sign indicates that the beads have been installed to represent ratios: most of the beads are white, but one in every ten is yellow, one in every hundred is green, one in every thousand is red, etc. There are exactly one million beads in the cylinder, and one of them is black. How long do you stand there, rotating the cylinder and looking for the black bead? If you have not found the black bead after twenty minutes, do you ever buy a lottery ticket again? In other words, how might this experience inform your possible future actions in the real world?

Another example: you are watching lighted dots travel along a single line on a computer screen. The dots follow two rules: 1) if the preceding dot is at or within X distance, continue to travel at the base speed and 2) if the preceding dot is farther away than X distance, travel slightly faster than the base speed until the gap is closed. You spend half an hour or more using sliders to adjust the base speed, the number of dots and initial spacing. Somehow, the dots always form clusters. It is then revealed that the dots can represent cars and the ensemble of these dots can be taken to represent traffic patterns. Does your active, prolonged engagement (Humphrey & Gutwill, 2005) with this representation, and your experience that clusters were inevitable, make you less susceptible to road rage when you encounter a traffic jam?

And a third example: a computer screen displays a rectangular field of small tan squares (Meir, 1998). At a click, one-fifth of the squares turn green, and you group the green squares into irregular clumps divided by expanses of tan squares. Another click results in a scattering of blue squares appearing, overlaid on the underlying field of tan and green. You click a button labeled “go” and the blue squares start to wander around. The ones on tan squares wink out irregularly as you watch, and the ones on green spawn new blue squares, which in turn continue to wander. After a few minutes, you notice that

the number of blue squares is decreasing, and a few more minutes pass until they have all disappeared. You reset the field, changing the locations and sizes of the green clumps. Same result. How might your engagement with these colored squares inform expectations about whether the Fender's blue butterfly of Oregon is doomed? What if you tried every possible arrangement of tan and green squares that you could think of, and still saw the blue squares wink out, one by one?

From these and other influential experiences I have had with models, I developed a long-standing interest in complex behaviors that arise from local, simple systemic rules carried out using agent-based modeling (Arai, Deguchi, & Matsui, 2005; Bonabeau, 2002). I also am intrigued by the underlying problem of how people understand and use these seemingly unrealistic or nominally pictorial representations, especially as they inform future expectations about how to act. Dots of light do not resemble cars in ways that would be judged adequately pictorial, but the traffic jam model was instantly recognizable to me and the museum visitors around me. Beads are not events, but my internal sense of a "one-in-a-million" chance was permanently altered. Decisions about land conservation and butterfly populations were being made on the basis of models similar to the one I explored, though blue squares are not like butterflies. In all three examples, the representation has almost nothing in common with a picture-like depiction of the represented reality.

Despite the apparently weak literal correspondence, I used these models to make decisions and generate explanations about reality. The designer of each model intended me to do so, without seeing it as necessary to make the representation in any sense "picture perfect." The direct simplicity of these models focused my attention on the *process* rather than the image, on the *logic* that created the outcomes rather than the pictorial fidelity of the elements. And in each case, I was confronted with a surprising

outcome that I had not consciously detected, even though I had experienced the corresponding reality depicted by each of these models.

This backdrop of surprise and explanation through modeling was latent in me through years of working in museums and the beginnings of graduate school. I was still relatively new to education as a field of study when I took *Systemic Reform*. It is a core class, and intended to be a capstone course, "broadly synthesizing and situating" (Stroup, 2013) many aspects of research in science and math education. Instead of being an opportunity to synthesize, the course turned (for me) into an explosion of ideas and intellectual interactions.

To my surprise, I also encountered explicit references to and explanations of some of my personal experiences. The traffic jam model I experimented with at a museum turned out to be based on a set of programs written by high school students in the language StarLogo (Resnick, 1996). StarLogo was a precursor to the language NetLogo (Wilensky, 1999), that I used to create the *DI tool*. I myself must have participated in research about Logo, another precursor language, as an elementary school student near Boston, Massachusetts, in the early 1980s. I remember writing programs that would make an on-screen "turtle" draw given figures, and arguing with an adult (not my teacher) over whether one program could be "better" than another if they resulted in identical output. My position was that the shorter program was better, and I was surprised to be asked for reasons and explanations.

My experience in *Systemic Reform* renewed my interest in the behavior of systems and in the role of structure. Some outcomes are products of the structure of the system, rather than intentional acts on the part of elements. That interest had formerly been diffuse, but centered loosely around ecology and natural history as a result of my experience in those fields. I now saw that educational change at all levels was influenced

by structural factors. Even something as personal as the development of an identity as a teacher could be seen as a result of the structure (i.e., the order and nature) of influential experiences.

In my own history of influential experiences as a student and as a teacher, I had noticed a fundamental tension between prediction and surprise. My best moments in either role have come from what one might call constrained surprises, as I leaped just beyond what was expected or enjoyed seeing another do the same.

One particular manifestation of this tension that I began to find problematic was the relationship between past and future performance. Previous learning successes and failures affect students deeply, both in terms of accessibility of new content, and in terms of motivation and attitudes toward learning and one's identity as a student. I had seen this among my classmates and experienced it in my own learning. If learning success is simply the accumulation of successful moments, student performance will be consistent over time. This gives rise to the shorthand summarizing of the "A student" or the "C student." The reality is more nuanced, though; many students intermittently do much better (or worse) than their previous performance predicted, and this is lost in the on-going averaging of student outcomes associated with most formal schooling.

As a result of these experiences, I see a basic "problem" in teaching as this: balancing the effects of history with the potential of the future. I hope that all teachers might recognize the difficulty of this balancing act, but my own experiences suggest that teachers are typically left to struggle in relative isolation to articulate their thoughts, decision-making and motivations surrounding why students fail or succeed in school. In some significant sense, the work of developing and exploring the use of the *Direct Instruction tool* is framed by these experiences. It is also framed by my desire to probe the “balancing” I see as fundamental to the activity of teaching. The *DI tool* is of a kind

with the kinds of simulations that have been meaningful to me. The work of this thesis allows me to do more to articulate the *what* and *how* of these types of compelling tools relative to classroom teaching.

With this personal framing and situation of my own perspectives in place, features of the *DI tool* can now be discussed.

FEATURES OF THE *DIRECT INSTRUCTION TOOL*

Against the backdrop of my personal experiences, I designed a computer model for my final project in the *Systemic Reform* course. The *DI tool* is an agent-based model programmed in NetLogo (Wilensky, 1999), and consists of entities representing knowledge interacting with (i.e., being transmitted to) entities representing students. The knowledge agents are assigned an arbitrary numerical difficulty, sampled from a uniform distribution up to the user-defined maximum value. The student agents are similarly assigned a numerical value called "receptivity." The starting receptivity for all student agents in the *DI tool* is half the maximum knowledge difficulty. For example, if the maximum knowledge difficulty is set at 30, the starting receptivity value for any student agent is 15.

In each iteration of the program, knowledge difficulty is compared with receptivity on a per-student basis. If difficulty is lower than receptivity, then that student's receptivity increases by a fixed increment. If difficulty is higher than receptivity, then that student's receptivity decreases by the same fixed increment. The program rules repeat until stopped by the user. Importantly, two types of graph are displayed: one is a plot of the average student receptivity through time, showing how the average value has changed since the beginning of the model run. The second graph is a histogram of student receptivity values in the current iteration (a “snapshot” rather than a

view through time). The contrasts in what might initially seem like redundant depictions of the same data end up playing an important role in how the simulation is used and understood.

When the program is run, student receptivity values form a Gaussian distribution within a few iterations. As the model continues to repeat the rules described above, receptivity values diverge. The model eventually stabilizes with a proportion of the receptivities at the maximum possible (given the current setting of the knowledge difficulty slider) and the remainder at the minimum possible value.

Like *ELIZA*'s use of an easily-enacted form of therapy, direct instruction was chosen because of its relatively simple rules and not, in either case, as an endorsement of these rules as such. The modality assumes that the primary interactions of learning are between a teacher (or transmitter of information) and one or more individuals (students), whether solo or en masse, as would be found in a classroom. Student-to-student interactions are minimal in a lecture hall, where all are listening intently to the speaker. As a result, information is "broadcast" to the agents representing students, and there is no information exchange between students in the model. Similarly, I excluded other features that have an important role in real teaching, but were judged to be immaterial to the primary interaction described above. In the *DI tool*, there are no physical arrangements (e.g., the layout of a classroom) or structural decisions (e.g., types and timing of assessments), like one would expect in a real-world classroom setting.

The *DI tool* was originally intended to make visible, and then possibly open to challenge, the deterministic position that student success is controlled by past success. In developing the simulation, I adhered strongly to the idea that "teaching done right is a subversive act" (Petrosino, personal communication). My creation of this simulation was in itself a teaching act on my part, and was meant to be taken as plausibly subversive, by

opening up lines of critique relative to this deterministic position. Though the *DI tool* might initially sound like a reasonable approximation of reality, careful examination of the rule that success or failure in one learning event affects subsequent chances of success or failure leads to problematic results. Most testing events in most classrooms show, or are expected to show, a Gaussian rather than multimodal distribution of test scores.

To my surprise, given the limited logic implemented in the model, the very first users (students in my section of *Systemic Reform*) of the *DI tool* – like the users of *ELIZA* – not only seemed to accept the underlying rules as reasonable, they considered the non-Gaussian outcome to be realistic as well. Users spoke in terms of "students you just can't reach" (Comment 01) and "every class has some kids who won't learn" (Comment 02). These statements were in marked contrast to the ideal of "every student can learn" that those same users also espoused in ordinary classroom conversation. I was shocked and intrigued by the apparent conflict in these future teachers' internal landscapes. This work marks my attempt to link what users “see” in the model to how the model seems to function as a support to articulating insights that are then associated with what happens in real classrooms. The research questions, discussed next, are meant as a follow-up to my interest in the *what* and *how* provoked by my surprise at my peers’ engagement with the *DI tool*.

Research questions

Like *ELIZA*, the *DI tool* was not developed with the intent of copying reality, but clearly each program has some elements that users see as realistic or meaningful. Pilot data suggested that users interpreted the *DI tool* in some consistent ways, but also brought their own learning and teaching experiences to their interpretations. The program appears to structure experience both backward and forward (i.e., explain the past as well as predict future outcomes). The goal of my dissertation study is to follow up on the *what*'s and *how*'s associated with users of the *DI tool* by seeking answers to the following questions:

1) **What are the kinds of claims or insights that respondents generate in relation to using the *DI tool* to organize their experiences?**

What do they see as making sense? What do they see as surprising or warranting further investigation and reflection? This question includes the issue of whether there are themes or patterns in what respondents articulate as part of their understanding of teaching. In what ways do they extend, contrast, or elaborate on their expectations about how classroom teaching, in a broad sense, "works?"

2) **How do the functionalities of the *DI tool* fit with or support what respondents see as meaningful?**

The focus of this question is on how the respondents' interactions with aspects of the simulation fit with, or serve to elicit, users' expectations about teaching and learning. What features matter, are pointed to, or engage respondents as part of their sense-making?

In essence, these questions might be expected to illuminate broader issues of reality, and how fidelity in representation works, for simulations that are "unrealistic" (or at least "un-picture-like") but display internal logic that is apparently compelling to users.

To get at the (1) *what* and (2) *how* of use, the methodology focuses on users' reports of their emotional and intellectual responses, as well as their intended actions. A protocol of prompts was developed out of previous pilot work and faculty reports from using the software in classes (e.g., both *Knowing and Learning* and *Classroom Interactions* in the UTeach STEM teacher certification program). Although this protocol and analysis is structured by previous use, the methodology employed is meant to be close to those aspects of Grounded Theory that allow for patterns to emerge in careful coding and reflection on complex artifacts (in this case, written responses to prompts). This methodology retains the tentativeness associated with what is admittedly an early effort to make sense of the meaningfulness of what, by contrast with the presumed "state of the art" in computer simulations, are unrealistic, un-picture-like, representations.

Chapter 2: Literature Review

COMPUTER TECHNOLOGIES AND LEARNING

This literature review follows the discussion of the development of the *DI tool* and the posing of the research questions because, unlike what might be expected from work being done within a particular realm of well-developed investigation, this review follows from the outcomes of using a particular computer environment. The literature review does not lead up to a set of questions, but rather clarifies issues after they initially emerge, in sometimes surprising ways, from users' responses. These issues do overlap with other, long-standing traditions in education and teaching research. The literature review, then, is positioned to augment issues that emerged from the data.

One place to start in helping to clarify and deepen the engagement with the research questions that follow from use of the *DI tool* is the "computer environments for learning" (CELs) literature. CELs is a broad term that encompasses models, simulations and animations. CELs can be used for a variety of purposes, including delivery of information, predictive modeling, adaptive testing and more. Perhaps because of the relatively recent use of computer modeling technology in classrooms, researchers have not come to a consensus about the effects of these types of tools – or computer technology in general – on learning (Means, Toyama, Murphy, Bakia, & Jones, 2009). Some benefits are clear: for example, computer environments can provide access to phenomena that are too time-consuming, remote, complex or unethical to experience or manipulate directly (Schwartz, 2007; Finkelstein et al., 2005). Computer games with educational content have been shown to motivate skills practice (Miller, Shell, Khandaker, & Soh, 2011) and to create a compelling cycle of learning-related user

behavior and positive game feedback (Garris, Ahlers, & Driskell, 2002); both conditions may increase engagement and subsequent learning.

It would seem logical to evaluate the utility of a CEL by making a direct comparison between the computer environment and live experience, when possible. Studies of this type have shown no clear benefit to technology use (Bell & Trundle, 2008; Winn et al., 2006). When students have the opportunity to construct models that reflect their own thinking, however, they show significant gains in content knowledge as well as understanding of modeling in general (Schwarz & White, 2005; Zhang & Linn, 2011). These results suggest that the specific type of computer environment has a substantial impact on its role in learning. Descriptions of CELs need to be linked to the context and the underlying functionality of the environment.

CONTRASTING CASES

For the purposes of this study, it is particularly important to distinguish the *DI tool* from other, superficially similar, CELs. The *DI tool* models an interaction between teacher, student(s) and "material to be learned," while making few claims about the nature or characteristics of any of those entities. I will set aside for the moment that the interaction is deliberately simplified and distorted, and that the entities are highly abstracted, and return to those points in later sections.

The *DI tool* is also explicitly a tool, to be used in a variety of contexts and perhaps for different purposes, depending on the context. It is not a replacement for any particular aspect of learning, and it does not represent a more efficient (or effective) "method" for achieving established outcomes. As such, it is distinct from CELs that are designed toward those ends, like the examples below. Each of the following cases

represents an effort to improve learning by making some aspect of the teaching or learning experience less laborious than it was in the past.

Teaching machines

Teaching machines originated as mechanical devices intended to make the process of learning more efficient and individualized (Skinner, 1961). Each student operated his or her own machine. Information and a "learning task" (question, fill-in-the-blank, *etc.*) were presented visually on slides mounted on a rotating frame in a set sequence. The student completed each task by entering a response in another window of the machine, and that response was immediately scored. In some versions of Dr. B. F. Skinner's teaching machine, the machine itself judged the response (e.g., a punch-card response to a multiple-choice question); in other versions, the correct response would be displayed for the student to compare to his or her own response. In both versions, correctly-completed tasks would be taken out of the rotation, so a student might cycle through the slides 2 or 3 times, completing a smaller subset of tasks in each iteration. It was the intent, however, that most students would complete the sequence successfully on the first attempt. Students would "quickly learn to be right" (Skinner, 1958) and the tasks were designed in a deliberately incremental fashion, such as having an early task offer information necessary to complete later tasks.

Machines were designed to teach arithmetic, spelling, and memorization of poetry. They were intended to have broad applicability in different disciplines and at different levels, though: one machine could "induce the student of high-school physics to talk intelligently, and to some extent technically, about the emission of light from an incandescent source" (Skinner, 1958, p.972).

Modern versions of the teaching machine are not strictly tied to the mechanical limitations of early versions. Any program that presents tasks or information in a flash-card format, requiring an answer from the student and re-presenting problems with incorrect solutions, can be considered a modern analog.

The teaching machine, in either mechanical or digital form, is a tool used *by the designer* on the learner. Communication is uni-directional in the sense that the designer (i.e., the team comprising content expert and computer programmer) learns nothing from interacting with the student via use of the machine. In fact, such interaction is made impossible by the time shift (the machine must be fully designed before the learner uses it) as well as by the limitations on student responses. Even if the designer were in the same room as the learner, the teaching machine was not intended to supplement or replace "the productive interchange between teacher and student in the small classroom or tutorial situation" (Skinner, 1958, p.969). Its primary purpose was to reproduce accurately what were seen as the best aspects of one-to-one interaction from teacher to student: frequent feedback, and a sequence of tasks likely to lead the student to success.

These features are in contrast to the *DI tool*, which is used *by the learner* to experience, reflect upon, or communicate about phenomena. There is neither an underlying "rightness" nor an explicit task. The *DI tool* presents a distorted version of the real-world interactions it represents. The lack of explicit tasks obviously leads to a lack of sequencing; there is no "moving through the material." These features mean that the learner is the one who creates meaning from the use of the tool, rather than receiving the meaning projected by the designer.

The teaching machine also has a very narrow range of utility. It is designed for a particular level of student, and is meaningless to learners outside of that range, whether more or less knowledgeable. As the tool for the user, rather than the designer (Halverson

& Smith, 2009), the *DI tool* is widely applicable to any participant in the represented interaction. It has meaning to students, to future teachers, and to current teachers at all levels of experience.

Expert systems

Digitizing teaching machines creates the opportunity for expert systems. These software programs consist of "teaching machine-like" structures of sequenced tasks, combined with explanations. The pattern of individual student responses influences both the information and the tasks that are subsequently presented by the program. These programs are sometimes designed to model the instruction that would be provided by an expert teacher (e.g., *Ms. Lindquist* (Heffernan, Koedinger, & Razzaq, 2008) was programmed to imitate a real algebra instructor of that name). The underlying structure of the interaction between program and learner remains similar to that of teaching machines. Current research on the expert system *Cognitive Tutor* (Anderson, Conrad, & Corbett, 1989) consists of investigating the relative effectiveness of offering incorrect examples alongside or in place of correct examples as part of feedback to the learner (Booth, Lange, Koedinger, Newton, & Newton, 2013).

Also like teaching machines, expert systems are uni-directional from the designer to the learner. Though the process of designing an expert system involves studying common student errors and effective system responses, the completed program is specifically and explicitly an expert (in contrast to the student's novice status). An expert system is intended to be a better judge than the student himself of what needs to be done or experienced in order to learn.

The contrasts with the *DI tool* described above apply to this case as well. An expert system is not a tool that the learner uses to interrogate his or her own thinking, or

to communicate with people. It is a tool the designer uses to communicate to the learner and to influence the learner's behavior. That commonality between expert systems and teaching machines can be obscured by the versatility of the former: expert systems have much more flexibility in responding to learners at various levels of knowledge. Nevertheless, the two types of CELs are fundamentally structured in a similar way.

Expert systems are currently clearly non-human, though the explanations and sequencing of tasks may be modeled after the actions of real instructors. Early users of *Cognitive Tutor* preferred short, "machine-like" hints and text over natural-language sentences (Anderson, Corbett, Koedinger, & Pelletier, 1995). Given current advances in natural-language processing, one can imagine a computer tutor that is indistinguishable from a human tutor interacting remotely with the student. As those technologies develop, learner preferences may shift toward human-like language.

Education-oriented gaming environments

There are many CELs that are primarily games, with some educational elements or intentions. These types of environments can be emotionally compelling, leading to vivid experiences that persist over a long period of time. For example, *Oregon Trail* (Rawitsch, 1978), a computer game initially designed to teach middle-school history, became widely available to students in the mid-1980's. Its iconic, pixelated graphic of a wagon and ox is still re-printed on t-shirts today, constituting a social marker of the wearer's (approximate) age and his or her positive experience with the game. Few educational experiences achieve such a following. Because the "educational" aspect of gaming environments may be minimal, I focus on one game that is explicitly intended to be an immersive educational experience.

Quest Atlantis (Barab, Thomas, Dodge, Carteaux, & Tuzun, 2005) combines social tools and science-related tasks. Learners navigate an extensive digital world and interact with elements of the world (i.e., "collecting" water samples by clicking on a pond). They can also digitally chat with programmed characters and with other learners who are currently in the program. Different parts of the digital world present different sets of challenges, either interdisciplinary or focused around a particular knowledge base (e.g., astronomy).

The intent of *Quest Atlantis* is to allow children "to engage in inquiry activities similar to those of experts" (Barab & Dede, 2007, p.3). Many of the scenarios require learners to consider multiple perspectives on a real-world problem or to make sense of contradictory data. As with other CELs, the computer allows learners to complete activities that would be prohibitively time-consuming in the real world, such as waiting twenty years to discover the long-term outcome of the learner's recommendation for water quality management.

Despite the intent stated above, much of the expert decision-making is made by the designer, which limits the inquiry available to learners. The program provides a structure for interaction around disciplinary knowledge, but it is not an investigational tool. It is a distorted copy of the real-world actions and interactions that scientists experience, programmed to guide learners through a particular sequence of tasks.

One important difference between the *DI tool* and *Quest Atlantis* is that, in the latter, the science concepts are the tools (Barab et al., 2009), rather than the computer being the tool. This focus on the science concepts assumes that the computer and the game are transparent to the user. As I discuss below, that assumption is problematic.

Another difference is that the *DI tool* models a phenomenon that can be considered as separate from the user, which creates the opportunity to question the reality

of the phenomenon. The object of consideration is the representation. In contrast, *Quest Atlantis* models the socio-technical context (Barab & Dede, 2007) or the community that exists around a particular suite of activities. That same context arises when students work together on real-world projects, but the digital version compresses space and time. Immersive, participatory simulations (Barab & Dede, 2007) may provide access across distance and time, but what they are providing access *to* is other people.

Networked participatory simulations

Gaming environments such as *Quest Atlantis* are described as participatory to indicate that the user is engaged in a holistic investigation with social, as well as scientific, ramifications. There is another category of participatory simulations that uses network capabilities to make visible to a group the actions of many individuals simultaneously, through avatars on a screen. Each student controls one avatar, and emergent behavior becomes visible as a product of individual action and decision-making. Simulations of this type can be completely abstract (algebra taking place on a coordinate plane) or based on a real-world system (traffic movement through a grid of streetlights); in both cases, individual students are literally participating in the behavior of the model (Wilensky & Stroup, 1999).

In theory, a teacher could create a coordinate plane by marking the axes on the classroom floor, then ask students to stand on the plane according to a particular rule (e.g., "where your y-value is equal to your x-value"). That is the essential task of the Function Activity, one example of a networked participatory simulation. The simulation *retains the individual agency* of each student in the group, *streamlines* data collection, and *automates* the creation of a graphical representation of the group's collective outcome. Multiple representations of ideas exist through concurrent digital and live

interaction: the placement of a student's avatar reveals his or her mathematical thinking, making that thinking immediately accessible to discussion.

The *DI tool* has no participatory element. It is the digital equivalent of a demonstration. As such, it should be nominally generative (Stroup, Ares, & Hurford, 2004), because it represents a single outcome with a single path to achieve that outcome. No user actions change the outcome of the model. One feature of the *DI tool*, however, is that the outcome is unexpected to most users. The *DI tool* gives the illusion that there might be multiple outcomes, depending on user choices such as class size. Perhaps for that reason, the tool has a disproportionately generative effect (as this study shows). The *DI tool* may surface thinking about teaching and learning, allowing users to explore the "quality of pathways to an endpoint" (Stroup, Ares, & Hurford, 2004).

AFFORDANCES AND CONSTRAINTS OF COMPUTER VISUALIZATIONS

The *DI tool* is not a sequence of tasks, a copy of a person, or a copy of a social context. It is a dynamic and interactive visualization of a specific phenomenon (Linn, Chang, Chiu, Zhang, & McElhaney, 2010; Zhang & Linn, 2011; Svihla & Linn, 2011). It is critical to explore how affordances and constraints of visualizations shape the way learners behave, because all of the technologies described above have the implicit goal of controlling the user's actions. That goal may, in fact, be explicit: the increase in "on-game" or "hallway" talk in the real world is a sign of success for the designers of *Quest Atlantis* (Hickey, Ingram-Goble, & Jameson, 2009). Designers intend to reach through the model to create new real-world phenomena or change real-world patterns. Though the *DI tool* could have an influence on the practice of teachers who spend time exploring it, the focus of this study is the localized influence of the program on claims and narratives that users generate.

The affordances and constraints common to visualizations of this type are well-described for other instantiations. In particular, the research findings from the Physics Education Technology (PhET) Simulations design and research program at University of Colorado at Boulder can be applied to the *DI tool*'s future development as well as understanding its functionality. The dozens of PhET simulations (i.e., interactive dynamic visualizations) that have been built cover a range of topics (primarily in math, chemistry and physics) and give users the opportunity to, for example, manipulate factors and compare multiple representations. Recent simulations are designed more quickly and are more effective, due to a well-developed and carefully-articulated design process involving multiple iterations and testing with end users at each stage (Podolefsky, Perkins, & Adams, 2010). The useful lessons to be drawn are not only about how the simulations operate to support learning, but how the designers turned an idea into a useful computer tool.

Simulations are not simply copies of the real world or convenient replacements for real-world experiences. With carefully-designed affordances and constraints, simulations can actually be more effective than their real-world counterparts. A direct comparison between a simulation about electrical circuits and an equivalent lab experience with real equipment showed that students using the simulation had a more thorough conceptual understanding of currents and circuits. In addition (and somewhat surprisingly), students who learned from the simulation used real-world equipment faster and more accurately when asked to build a challenging circuit (Finkelstein et al., 2005). Similar results were seen when comparing the performance of students who learned from a simulation to that of students who watched a lecture with an associated real-world demonstration.

One reason for the success of simulations is that productive constraints are explicitly designed. This makes simulations a novel tool, not simply a substitution (*sensu* Riley, as described in the Introduction of this work) for real-world activities (Podolefsky, Perkins, & Adams, 2010). Simulations elicit user engagement and exploration by providing the opportunity to manipulate factors that learners tend to think should be influential. The *DI tool*, for example, has a slider for controlling class size because many users consider class size to be an important factor in student learning, with smaller classes being "better" for students. Therefore, the *DI tool* offers the opportunity to increase or decrease the number of students even though that number has no effect on the outcome.

Productive constraints are provided partly by designers (i.e., inherent to the simulation) and partly by the context of use. For example, a high level of guidance or instruction in how to use the simulation reduces exploration and subsequent learning (Wieman, Adams, Loeblein, & Perkins, 2010). A student's approach to a simulation, whether dictated by an instructor or the student's own personality, influences effectiveness; in one study, students who were prompted to predict simulation outcomes before exploring learned more than students who explored the simulation without making predictions in advance (Finkelstein et al., 2006). Previous exposure to a particular simulation may push a student into "performance mode," consisting of superficial exploration, avoidance of novel challenges, and insistence on remembered knowledge instead of using evidence from the current simulation to explain phenomena. Careful structuring of the related tasks and assessments can counteract a student's tendency to switch from exploration to performance (Adams et al., 2008). Regardless of the context, most PhET simulations provide sufficient implicit and explicit scaffolding to allow students to use them independently (Finkelstein et al., 2006).

Early concerns about the lack of transparency of computer simulations (Roth, Woszczyna, & Smith, 1996) have proved to be unfounded in the long term. These older simulations had a long learning curve, with users focused on *how* to use the tool for up to several hours before being able to *use* the tool, as intended, to investigate phenomena (Suchman, 1987). Extensive design work with the PhET simulations shows that careful design is enough to make dynamic interactive visualizations stand alone to the inexperienced user, with the computer tool becoming "ready to hand" to the user in a very short time (one or two minutes). This compression of the time it takes to learn the tool represents a change in the interaction between humans and computers – a development of common language, one might say, as the technology improves and users become more accustomed to using computer tools as an extension of their own thinking and communication.

REPRESENTATIONS OF TEACHING

In addition to being a computer tool of a particular type, and having characteristics in common with other such tools, the *DI tool* is a representation of teaching. A representation serves as both an elicitor of ideas about teaching and a medium of communication among a community of sense-making individuals. The *DI tool* is an extreme abstraction along the same continuum set by the transition from video to animations (Chazan & Herbst, 2012) and has many of the same affordances and constraints of moderate abstractions of teaching.

Visual realism is problematic for both the *DI tool* and for other representations of teaching, because realism is neither necessary nor sufficient for meaningful sense-making. This study shows visual realism to be unnecessary in the case of the *DI tool*. In the case of video representations, one would assume that strict visual realism would

support sense-making. In fact, even authentic (i.e., unstaged) video can seem too "distant" and unrealistic for pre-service teachers unless coupled with live observation (Santagata & Guarino, 2011). The fact that an event actually occurred or that a video clip represents real people is not by itself sufficient to make the representation a useful tool for learners of teaching.

As with the *DI tool*, teachers project their own thoughts and experiences on the representation. A video representation is constrained because it is about particular people, and many teachers find it difficult to treat the particular case as a generality. In addition, video interferes with the discussion of alternative outcomes. The outcome that actually happened is the only one represented by the tangible; as a group of teachers discusses "what would happen if" a teacher or student responded differently, their speculations are unsupported by a shared artifact that eases communication. Animations of teaching events remove both constraints while retaining the affordance of using a representation of teaching to support conversations (Chazan & Herbst, 2012).

Animations of teaching can provide opportunities for practice and community-building (Chieu, Herbst, & Weiss, 2011). The practice in this case, however, is narrowly focused on the skill of noticing, rather than attempting to provide practice in teaching as a whole (as *simSchool* claims to do). "Noticing" as a narrowly-defined skill is demonstrably practicable through a video format paired with discussion, so using the animation constitutes an extension of a form of practice that is known to work for teaching. A group of learners (e.g., pre-service teachers) discuss the sense-making moves depicted in either the video or the animation, and through that process become increasingly alert and attentive to student thinking.

Another particularly valuable affordance of both video and animations is the representation of time in teaching, because the practice of teaching depends on the

timeliness of actions and the way teachers respond during critical moments. The same teacher behavior may have very different effects, depending on the context or time placement. Video representations offer the necessary sense of time, but are also "thick" with additional layers of detail. Animations, in contrast, retain the element of time, but also focus user attention on the core abstraction by removing some of the layers of obscuring detail. An animation that represents the teacher as a square and each student as a triangle, for example, removes the opportunity for a user to make sense of the teaching moment through the lens of race. A video of the same event necessarily includes the option (though not the need) to explain actions and outcomes in racial terms (Herbst, Chazan, Chen, Chieu, & Weiss, 2011).

A teaching animation, therefore, constitutes a move toward a teaching-specific Rorschach blot or an artifact that is more important for the stories it elicits than for the fidelity of its own representation (Chazan & Herbst, 2012). The *DI tool* is even more abstracted than a teaching animation, and creates opportunity for individual and communal sense-making by drawing particular attention to "structural" aspects of teaching (rather than emotional aspects, intellectual aspects, or the specifics of decision-making). As with other dynamic interactive visualizations, the model in some cases is potentially more effective than the real world (i.e., observation) or mere selection thereof (i.e., video).

INQUIRY AND DYNAMIC VISUALIZATIONS

The duality of dynamic interactive visualizations is that they both enable and dampen inquiry. Visualizations help users engage new ideas and clarify thinking, and, if sufficiently abstracted, provide enough flexibility that different users can superimpose

different ideas on the same visualization. Though the usability of a visualization can be explicitly designed, the purpose for and outcomes of its use are context-dependent.

One risk of visualizations is that, by making the invisible easily seen or tangible, they may promote a false sense of secure understanding on the part of the user (Linn et al., 2010). Designing appropriate constraints reduces this problem. Respondents to the *DI tool* were surprisingly committed to the reality that the tool presented, which suggests that some aspect of the tool's design or context was implying that the *DI tool* is "right" about teaching (or "true"). In a broader sense, CELs can be designed to problematize teaching in general and counteract the "apprenticeship of observation" that tends to reduce change in teaching practices (Lortie, 2002).

A visualization is an artifact that is both a source of evidence and a social organizer (Roth, Woszczyna, & Smith, 1996). For inquiry to be applied effectively to the learning of teaching, one needs artifacts (tools) that function in both roles within the teaching context. Simple tools, designed to be nearly transparent to the user, will be more useful and versatile than complex tools that attempt to encompass many aspects of teaching.

Pre-service teachers differ in their epistemologies, or ideas about what it means to know (White, 2000). Knowledge about teaching is particularly problematic. Many people, even current teachers, think that instruction is something that teachers do to learners rather than with them (Cohen, Raudenbush, & Ball, 2003). Part of learning to teach is developing a more nuanced view of the roles of teacher and learner, and what constitutes legitimate action for both parties. Furthermore, a teacher's epistemology determines how he or she responds to many other aspects of the daily routine of school learning. Different teachers may view an identical event as a learning failure on the part of the student or a learning opportunity for the teacher (Hammer, 1997). In short,

effective teachers have to be aware of their own epistemologies. Because CELs are about a digital world, not a human one, they disconnect the emotional and relational aspects of learning from the epistemological ones and allow users to articulate their thoughts (Chazan & Herbst, 2012).

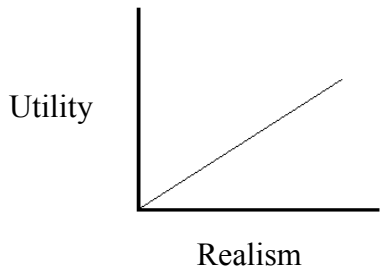
CELs provide a vehicle for specific case studies, especially the combination of "what happened?" with "why did it happen?" The integration of practice and theory helps teachers refine their skills and understanding of what teaching is (Shulman, 1986). Shulman distinguishes between propositions (teaching rules) and conceptual knowledge about teaching, and calls for case studies to help teachers build the theories that lead to flexible, generative teaching. Animations provide affordances specifically in this building of case study, because the decision point can proceed to either outcome with equivalent authenticity (Chieu & Herbst, 2011).

Because of their lengthy experiences as learners in classrooms, pre-service teachers often need to learn explicitly that teaching is complex. They need influential experiences that problematize teaching, rather than reinforcing the limited view they (in many cases) have previously held, that teaching is simple, procedural, and straightforward. Reflection is an effective tool for engaging teachers in the process of questioning their teaching knowledge and actions in a growth-oriented way (Sweeney, Bula, & Cornett, 2001). The technology that would allow a computer environment to effectively replace a classroom is still a long time in the future, so the utility of CELs for teaching as a replacement for live practice is extremely limited. As a tool for reflection, however, CELs that provide the appropriate level of abstraction, that serve as an extension of gesture, and that are designed to problematize rather than to obscure difficulties, will become an increasingly vital use of new technology.

REALITY AND DELUSION: THE PROBLEM OF "REALISTIC"

In retrospect, my initial attempt to fit the *DI tool* within organizing axes designed for a more limited group of models (Schwartz, 2007) was a mistake. One of the key aspects of Grounded Theory is that theory emerges from the data, and the researcher explicitly approaches the data without pre-existing explanations in mind. My pilot data suggested that some meaningful phenomenon was at work, so I entered the process of data collection and analysis for this study with some ideas already. I describe how I balance those tensions in the Methodology section.

Intuitively, one might predict that a computer representation with more features in common with reality is also more true, useful or valuable, as depicted here:

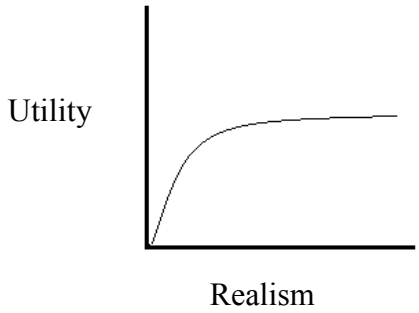


I contend that even if a computer representation that perfectly modeled reality in every way could be created, it would fail to provide insight. First, a perfect copy simply transfers the problems of the current reality into the new setting. If one writes a formula on a chalkboard, and a student fails to understand the meaning of the formula, displaying the formula on a computer screen is probably not going to help. Second, the theoretical endpoint of literalness is impossible to reach with current technologies. The designer of the representation is therefore making decisions on behalf of the user about which features to include. One of the assumptions underlying the relationship depicted above is that it is the accumulation or number of features, rather than the specific nature of the

features, that is meaningful. The designer then includes – and excludes – features with perhaps too little regard for the different meanings or values of those features in the eyes of the user.

A clarifying analogy is the prospect of going to the theater. A play that is exhaustively realistic would be merely boring, rather than insightful, as Aristotle points out in his Poetics (Aristotle, n.d., 2006). Good theater provides a focused, limited, perspective on the world.

Responses to the *DI tool* and *ELIZA* suggest instead the following relationship:



In this case, a small number of features are sufficient to make the computer representation useful. The level of correspondence with reality may be quite low, as long as the threshold level is reached (i.e., that point at which the above graph levels off). For the purposes of this work, I will refer to this type of model as a "threshold-type" representation. Though it is relatively easy to imagine that this type of representation exists, it is more difficult to distinguish it from the usual type of model, in which increased utility depends upon indefinitely increasing realism. I propose the following three criteria for recognizing a "threshold-type" representation:

First, it is used to organize forward in experience, i.e., respondents make claims or commitments about their future plans as a result of interacting with the representation.

This phenomenon indicates that the model is seen as useful and relevant enough to impact the user's real world experience.

Second, suggestions for change focus on how the model communicates with the user, not on the relationship between representation and reality. Given a sufficient level of utility, users want to ensure they are understanding all the information the model could be providing. They may even ask for output (e.g., graphs, labels or text) related to processes that are assumed to be happening in the model because a real-world correlate exists, even when these process are in fact *not represented*.

Third, because the number of meaningful features is limited, users' recommendations for additional features, if any, will converge on whichever features are important, but inadequately represented. In theory, it would be possible to compile a list of meaningful features. In practice, it is users' responses that create that list, and the existence of convergent suggestions for additional features indicates that one or several meaningful features are insufficiently represented.

Given that a "threshold-type" representation might exist, one then asks, how can such a patently unrealistic environment be seen as useful, or even realistic? At this point, some systematic framework for thinking about representations and reality is necessary.

The philosopher Jean Baudrillard (1929-2007) describes a progressive series of categories, in which the copy (which he calls the icon) moves from (mere) reference to representation to simulation to replacement of reality. An icon at the first three levels in the progression has an exchange rate, or a relationship with reality: "the sign and the real are equivalent" and separate (p. 484). A reference is essentially true, or a "reflection of basic reality" (p. 484). A representation distorts reality. In both these cases, the exchange is between the icon and a positive reality (Baudrillard, 1983).

In contrast, "simulation is… the generation by models of a real without origin or reality" (p. 482); simulations create something that is not really there. To give a concrete example, Baudrillard describes a patient simulating illness. The simulator produces symptoms rather than merely claiming to be ill.

At the extreme end of the progression, the icon can become reality. This endpoint reality may be only the extension of the previous simulation, or it may replace a reality that has some existence beyond the icon.

Scientific icons such as computer animations of ecological phenomena illustrate the gradual rather than categorical relationship between the steps in Baudrillard's progression. The steps are more idealized waypoints along a pathway than discrete categories. For example, a "computer model" of predator-prey relationships is an abstracted icon of what happens in the natural world. No matter how faithful it is to the rules that govern the interactions of actual lynxes and arctic hares, the icon distorts some aspects of the reality, placing it closer to a representation (step 2) than a reference (step 1). It is inaccurate, however, to say that it necessarily "masks and perverts" (p. 484) the reality. One can certainly imagine a similar computer model that is deliberately misleading, and so even more a representation: a construal with meaning, yet not bound to correspond with a presumed depiction of reality.

This gradual relationship extends to the progression from representation to simulation as well. An increasingly false (but increasingly complete and persuasive) representation becomes a simulation. If we consider an icon close to the simulation end of that transition, the icon makes no claim about the reality – it is a pure heuristic, a "what if?" that we explore without regard to the connections between the icon and an increasingly unnecessary other. This is in fact the natural next gradation from an icon that knowingly misleads, because the user's attention has already been diverted from a

constant comparison between the icon and reality. Once we stop asking, "what aspects of this are real?" the icon becomes a simulation, which "threatens the difference between 'true' and 'false,' between 'real' and 'imaginary'" (p. 483). Part of the role of a simulation is to obscure that difference, or sense of separateness, leading the user to conflate elements of the icon and reality.

It is when the comparison between icon and reality becomes impossible that the icon has reached step 4. Simulacra are what Baudrillard calls those icons that are their own reality. They are icons that, instead of "exchanging for what is real" (p. 484), exchange in themselves (author's phrase reworded, p. 484). This happens as an originally weak and imperfect simulation of the reality that might be there, so to speak, becomes perfect, or at least sufficient, and meaningful of its own accord.

There is one additional point to keep in mind about this progression. The author does not claim that all of reality is simulacra, only that some icons have the potential to replace reality as both a point of departure and as that which gets referred to. Not all icons or representational activities make the journey. However, because the steps are related to each other, even a representation has the opportunity to influence reality without becoming a simulacrum.

The *DI tool* is at least a malefice in Baudrillard's terms: a flawed reflection of a basic reality. In reality, students succeed or fail by a more complex process than is enacted in the *DI tool*. The *DI tool* proposes a vision of that process that is known to be untrue and over-simplified. Whether the program is also consistently a simulacrum remains to be seen. As a representation verging on simulation, pilot users commonly justified icon phenomena with "real-life" rules. For example, one respondent said "I think this does happen in classrooms because comparisons with smarter students does hurt the self-confidence of many students and, as a result, hurts their modivation [sic] and

performance in school" (Comment 03). This respondent postulated the existence of "real" processes based on a correspondence between the icon phenomena and the end result of some (unknown) actual processes that cause students to do poorly in school.

The apotheosis of that progression to simulacrum would be the reverse: a respondent interpreting phenomena in a live classroom in terms of icon rules. One pilot user said the following: "Yes, I've had a teacher who showed us the type of equations we would be solving by the end of the year and then slowly worked us up to that type of equations and I believe all the students kept up and understood" (Comment 04). This respondent's description of past experience is the exact correlate of a "solution" that exists in one version of the *DI tool*. In Baudrillard's framework, the *DI tool* acts as a simulacrum in this instance.

Though Baudrillard's progression from reference to replacement of reality provides some insight into the differences between the *DI tool* and other exemplars of CELs for teaching, it has a flaw: this progression fails to explain how pilot users of the *DI tool* experienced some shared realities. Baudrillard's simulacrum is left as idiosyncratic to each individual. I turn now to another author for an explanation of the existence of convergent experiences.

Charles Sanders Peirce (1839-1914) claims that we are immersed in signs and systems of signs. These signs are not "built up" from first principles, but rather exist as a system prior to any particular individual's notice. A given observer is always arriving mid-story. "Real" and "meaningful" only have meaning within the system, and one cannot step outside the system to get an independent perspective (Peirce, 1997).

This suggests a kind of phenomenology, in which different descriptions of experience are each taken as valid. Peirce argues, however, that shared beliefs, or, indeed, communities with beliefs and communities of beliefs serve to organize individual

experiences. In this context, “reality” is a set of representations and practices, plus the community that uses them. Reality may be small-scale or expansive, so one is always participating in nested and overlapping realities (De Waal, 2013). Because the collection of representations is reality, in Peirce's argument, the shared communality of representations (and practices associated with using them) give rise to common and intelligible experience.

The *DI tool* has organizing power shared across individuals, which is inexplicable in Baudrillard's framework. Peirce offers a complementary framework that accounts for the similarities in pilot user responses to the *DI tool*. Not all representations are equally truthful, even if they are part of the same system. One of the questions I propose to investigate (research question #1) is how respondents use the *DI tool* to organize their experiences despite its apparent mix of truthful (“realistic”) and untruthful elements.

Chapter 3: Methodology and Analysis

CONSIDERATIONS IN THE SELECTION OF METHODOLOGY

I analyzed written responses, submitted via a web form as described below, using grounded theory (GT). GT is a systematic approach to data analysis, usually associated with qualitative data. In contrast to most other data analysis methods, researchers using GT do not generally approach data with theory already in mind to be validated or contradicted. Instead, theory emerges through the iterative process of identifying concepts, making comparisons, generating tentative connections and evaluating the fit of the emerging theory with the data (Glaser & Strauss, 1967, p.21).

GT can be applied to all types of data, so using grounded theory in this study creates the opportunity for comparative qualitative studies to follow. This flexibility is particularly important given the possibility that pedagogical models and simulations may have predictable effects along several organizing axes; different situations may call for different data sources. GT also allows a transition to mixed methods, if appropriate; one can quantify the incidence of particular codes and use statistical patterns as indicators of underlying thoughts or interpretations (Creswell, 2003, p.212).

Due to my background in the natural sciences, I was initially drawn to the objectivity derived from the systematic, repeated comparisons and validations inherent in GT. The research questions are best answered by attending to the meanings users articulate or point to, and not by developing some metric quantity for comparisons based on standard measurement theory.

In the last 15 years, however, there has been a diversification in the details of grounded theory methods. The systematic objectivity that I value is present in Glaser's

approach (Glaser, 2001), who considers data from all sources and particularly emphasizes integrated concepts organized around a core category. For this work I, too, seek explanation – the essence of theory. For that reason, I follow Glaser in rejecting mere description, however detailed, objective and accurate (Glaser, 2004). But Glaser's insistence on an atheoretical beginning to data collection jarred my pragmatist sensibilities. Pragmatism acknowledges that we always come to observation and sense-making with prior ideas and theory. More importantly, such a commitment did not meet the needs of my study. This research exists at the intersection of several fields, and the details of that intersection are obscure enough without eschewing in advance what little light can be shed by previous research. Instead, my epistemological goals and needs were met by Strauss's approach, in which the researcher explicitly confronts and articulates the pre-existing theoretical context that accompanies data collection (Strauss & Corbin, 1998, p.44).

It should be apparent that I approached this study with some theory already formed, in contradiction to the data-first approach mandated by one of the standard interpretations of GT. The existing theory is the result of prior iterations and partial iterations of coding and hypothesis-generation, so it is not as complete a break with standard GT methodology as it might seem if the account began only with data collected for this thesis. My dissertation study therefore constitutes "theoretical sampling" (Strauss & Corbin, 1998, p.201). The data that I collected extended hypotheses that arose from my pilot data, and allowed me to make comparisons that elaborate existing relationships and categories (Strauss & Corbin, 1998, p.210).

With the amendments noted above, the nature of my research questions is generally well-suited to GT. The *DI tool* is neither fish nor fowl, neither reality nor a copy of reality, but it has some elements of each. In order to attempt to answer the

question of "what is it?" fully and accurately, my methodological framework needed to be open, not restrictive. I could not assume in advance the answer to this question. Regarding my research questions specifically, Question 1 is about a process of building connections that occurs internal to respondents, and calls for explanation of that process. Question 2 is about articulating relationships between the *DI tool* itself and the meaningful elements that users choose to respond to. My research questions go beyond the primarily descriptive focus of many qualitative data analysis methods. For my work, theory is implicated both coming into and emerging from analyses of the data.

Other methodologies were considered and rejected prior to the adoption of GT. To help clarify the commitments of this study, it is worth reviewing some of these alternatives. Ethnography is commonly used to make sense of qualitative data similar to that collected in this work, but ethnographic analyses are framed in terms of a notion of culture. The mere common use of a tool like a computer model, however, does not unite pre-service teachers into a common culture. I would consider both the *DI tool* and *simSchool* to be tools, not cultural artifacts arising from within a cohesive group. Phenomenological analysis is another framework that might have been well-suited to descriptions of the way study participants interacted with the computer program. It would contribute little, however, to my research questions, which are about connections and relationships. My research questions emerged from prior analyses of how the environment I developed was used, and I wanted a methodology that could predict, or project forward to, what might happen with the next new computer environment for learning to teach. This requires going beyond mere accounts of the phenomena (i.e., the phenomena of simulations of classroom interactions).

Before considering these two methods, I thought to use case study methodology. That could have been suitable if there existed any theory of computer environments for

learning to teach. What I found was that existing environments were inadequately theorized and were unconnected to other teaching tools, or even to theories of how people learn to teach. The most well-developed example was explicitly analogous to a flight simulator, but I was dissatisfied with the organizing power of that comparison. A teaching simulator is not like a flight simulator in several important ways. Unlike flight simulators, in which pilots can sit and handle the steering yoke and other tools as in a plane, teaching simulators do not reproduce the physical aspects of the activity that is simulated. One does not physically walk around the virtual room, handle the virtual books or tools, or get tired after long hours on virtual feet. Another key difference is the "translation" of visual information. In the case of a flight simulator, a digital image or copy of an instrument gauge provides all the same information as the original. Digital avatars, in contrast, fail to communicate as effectively as live students, because digital "body language" is much less nuanced than that experienced in live interactions. In the face of these concerns, I returned to GT as a way to develop a theory or principled account that might plausibly be used to make sense of both existing and future computer environments; although some of the literalism of flight simulators has been taken up by other environments simulating teaching, such a literalistic approach is only one of several possible stances. Theory about teaching simulations needs to extend across the range of possible environments.

LESSONS FROM THE PILOT STUDY

The *DI tool* was shared with University of Texas undergraduates enrolled in teacher preparation courses. In this context, user talk was limited and researcher-driven. The pre-service teachers seemed reluctant to speak freely, and kept their comments short, even in the face of results that were visibly provocative and surprising. I piloted a semi-

anonymous written response format, with questions that 1) checked the user's understanding of the *DI tool* results and 2) provided open-ended opportunities for the user to comment on reality, utility, explanatory power and flaws of the program. These written comments turned out to be long, detailed and intensely reflective.

A key goal of the original environment design was to demonstrate that average student scores on an assessment may not represent a feature of the class as a whole, and it did so effectively. As the *DI tool* runs through its iterative program, fewer and fewer students actually exist at the average receptivity value. Though the graph showing average receptivity may stabilize in the middle of the possible range, the histogram that shows the distribution of receptivity values displays a bi-modal distribution with peaks near the maximum and minimum.

Most pilot respondents understood this contrast: "the first graph…kind of gives a sen[s]e that the whole class is moving forward" (Comment 05), and "the average student…does not show up in reality" (Comment 06). Pilot respondents who offered a solution to this epistemological problem uniformly favored differentiated instruction: "switch back and forth between high difficulty lessons and low difficulty lessons" (Comment 07) or provide "extra support" (Comment 08) to low-receptivity students.

All pilot respondents were able to accurately describe the model's behavior under specified starting conditions, and were able to manipulate conditions. This confirmed that the environment is useable even when respondents explore without researcher guidance.

The concept of "receptivity" was a rich one in the pilot study. Respondents explicitly conflated receptivity with a variety of real-world concepts such as grades, understanding and motivation. I expected users to argue that the condition of identical starting receptivities is problematic. The pilot respondents did not disappoint, with about

one-third of their critiques mentioning that student receptivity values should begin at different levels. This starting condition has minimal impact on the behavior of the *DI tool*. The "realistic" starting condition of receptivity values drawn from a range actually hastens the inevitable divergence into a bimodal distribution, which many pilot respondents rejected as an "unrealistic" outcome.

Another suggested modification was that class size should control knowledge difficulty, with smaller class settings lowering the maximum of the distribution from which difficulty would be drawn. Finally, one pilot respondent suggested that opportunities to learn, represented by the comparison of receptivity and difficulty, are not equally distributed across all students in a class. Teachers can and do "pay more attention" (Comment 09) to some students than others.

These preliminary results shaped the subsequent research questions by focusing my attention on two patterns in particular. One, though many aspects of the *DI tool* are unrealistic, pilot respondents discriminated between features on the basis of the organizing power (meaningfulness) of a feature rather than purely on the basis of its realism. Two, the features that drew a particular respondent's attention seemed to be related to – and elicited narratives of – his or her individual experiences as a teacher or a learner. The pilot study resulted in a refinement of methodology that enabled me to more effectively investigate the research questions articulated above.

PARTICIPANTS, CONTEXT AND DATA COLLECTION IN THE DISSERTATION STUDY

Twenty respondents participated in this study. Seven of them were University of Texas undergraduates enrolled in teacher-preparation courses: EDC 365D (*Classroom Interactions*) and EDC 365C (*Knowing and Learning*). Pre-service teachers enrolled in these courses plan to become certified in middle-school or high-school science,

mathematics, or engineering, and all are pursuing their teaching certificate along with a bachelor's degree in a scientific subject area. The remaining respondents were in-service K-16 teachers (n=3) and graduate students in STEM Education (n=10). This diversity of respondents is appropriate to GT as a methodology, because their differing perspectives add depth and detail to the codes that are identified as meaningful among the responses of pre-service teachers. With GT there is no *a priori* idea of a "minimum sample size" such as might be associated with the use of inferential statistics.

Data were collected between 27 February and 17 March, 2013. Each participant explored the *DI tool* independently, using an instructional handout and the response questions as a guide. Because participants accessed both the *DI tool* and the response questions remotely, through an internet connection, the physical location of data collection is not known. Participants were instructed to run the program under specific conditions as a starting point, but also had the opportunity to explore freely, with no time limit (observed time ranges were 11 to 96 minutes).

Written responses to survey questions (see Appendix A) were submitted via a form in an online survey tool. Each question on the form was associated with a text box, in which participants typed their answers. After the initial verification page, on which participants indicated their status (pre-service teacher, graduate student or other), all questions were presented on a single page. Participants could skip any question, though this happened only once.

Pilot work conducted in a previous semester with graduate students exploring the *DI tool* indicated that interview comments and critiques could be extensive and sufficient to warrant systematic attention. Attempts to use an interview protocol with undergraduate pre-service teachers were, however, not successful. Pre-service teacher participants were hesitant or unwilling to elaborate, and rarely shared their interpretations

of the real-world equivalent of model behavior. A pilot study that offered parallel opportunities to write responses as well as speak them revealed that the written comments were long and detailed and contained extensive interpretation. Written responses often included narratives of personal experience as a teacher or learner; such narratives had been essentially absent from spoken responses. On the basis of this observation during the pilot phase of this research, I use the rich data source offered by written responses.

DATA ANALYSIS

GT does not proceed in a direct, linear fashion. It is necessarily iterative. As an approach, however, it can be divided into a series of specific descriptive, interpretive and analytical actions. The following steps overlapped as I shifted between generation and validation of concepts. For the purposes of clarity, I selected the development of four important and related final codes to feature in this description of the data analysis process. Similar analysis was done with other concepts.

Microanalysis, open coding and labeling: At this stage, I read closely to identify meaningful phenomena that exist in the data. These phenomena, or "concepts," may be events, ideas or properties of other phenomena (Strauss & Corbin, 1998, p.103). Concepts are abstracted and labeled, either by the researcher or by the study participant. The goals of "open coding" and labeling are to mine the data as fully as possible for concepts, to generate the word or brief phrase that summarizes each concept (if needed), and to begin describing the properties of concepts (Strauss & Corbin, 1998, p.121). This aspect of GT was particularly suited to my study because I was investigating the elements that users considered as meaningful, as well as the connections they drew between their experience with the model and their previous teaching and learning experiences. Open

coding involves explicitly searching for the variety of analogies and identities imposed on the *DI tool* by users.

Codes at this stage are preliminary, because they depend on a tentative grouping of multiple exemplars. This inductive reasoning is characteristic of GT (Strauss & Corbin, 1998, p.136). The abstraction process is influenced by the researcher's experiences and ideas, but the next stage provides an opportunity to check codes against the data for internal consistency, validation and explanatory power (Strauss & Corbin, 1998, p.156).

I began data analysis by reading through the written data with my research questions in mind. In particular, I was looking *for patterns in the claims and interpretations* that respondents made of the *DI tool*. I was also alert to indications that affordances and constraints of the model itself shaped the participant experience. Deep and iterative engagement with the data is the core of microanalysis, so I read all responses multiple times. After the initial reading, I organized participants by category (undergraduate, graduate, in-service teacher) and read each group of responses as a connected whole several times. I then read responses grouped by survey question, again, multiple times.

During each reading of the complete data set, I wrote memos (Strauss & Corbin, 1998, p.217) that contained notes on themes that were emerging as well as my reflections and questions. Through this iterative process of reading, reflecting and writing, I developed a preliminary list of codes that encompassed the concepts expressed by study participants. Though GT is a methodology in which theory emerges from the data, this study also includes some aspects of theoretical sampling. In the context of this work, theoretical sampling includes an effort to validate, extend and deepen some elements of

prior existing theory about the experience of using the *DI tool*. Specifically, I made use of some codes generated during pilot work.

Examples of preliminary codes that were developed from my data are:

"D student": this concept was labeled directly by pilot respondents (an "in vivo" code) (Glaser & Strauss, 1967). It suggests that students have an inherent property that influences the grades they achieve. This property might apply across different subjects or classes. This study deepens understanding of that code by indicating that respondents imposed the fixed property of "D student" only to other students. Their own grade-influencing property was perceived to be context-dependent rather than fixed.

"paying attention": this concept describes one way that teachers and students interact with each other and the subject matter. Students can pay attention to the teacher or to the subject, either of which may lead to learning. Teachers can pay attention to students in both positive and negative ways. Whether teachers pay attention to the subject matter was an open question going into this study. Responses suggest that teachers "pay attention" to the subject matter enough to make instructional decisions about whether to teach in a way that is "easy" or "hard."

"unrealistic": to my surprise, pilot participants used this concept when talking about the "real world" (another in vivo code) as well as when talking about models in general or specific features of the *DI tool*. Certain teaching actions were described as possible for teachers to do, but "unrealistic" because of time or attention constraints.

The above examples remained unchanged through the subsequent steps of GT, and appear as final codes. Other preliminary codes went through substantial refinement. The concept of "making it hard" is one example. This code arose from responses stating that teachers might present material too quickly or give lessons that are too difficult for a

given group of students. In addition, participants referred to being a hard or easy teacher: one who, presumably, makes the subject hard or easy.

Questioning assumptions: I cannot assume that my interpretation and labeling of participants' words reflects what was intended. At this stage, I revisited codes by reviewing all exemplars of a particular code together (Strauss & Corbin, 1998, p.95). I generated alternative labels and searched the data for evidence to support each articulated alternative (Strauss & Corbin, 1998, p.93). By constantly asking myself "what else could this mean?" I remained open to the concepts and relationships that were emerging. Codes themselves changed, if the data indicated that my initial interpretations were unsupported (Strauss & Corbin, 1998, p.159). This fits with the iterative commitments of GT.

After iterative development and modification of the preliminary list of codes, I further refined the list by using the data to 1) quantify the frequency of each code, providing one way to represent concept relevance and 2) validate the consistency of each code across the whole data set. Because several preliminary codes had been broken down into more specific component concepts, I again grouped exemplars of participant text that were consistent with each code, then critiqued my own assumptions by generating alternative meanings for each exemplar and comparing them across exemplars. The final list of codes is included in Appendix B.

This element of the process was particularly critical to shaping my analysis. For example, questioning assumptions while analyzing my pilot data generated the hypothesis that respondents distinguished between two types of "unrealistic," specifically when talking about computer models. One type was that the *DI tool* was operating by rules that were in conflict with the "real world" (my initial assumption). The other type

was that the *DI tool* simply failed to include some important factors; the rules were realistic but the outcome was not. Recognizing and articulating this difference allowed me to differentiate between respondents who were using similar language to describe quite different experiences.

In the case of the preliminary code "making it hard," as I reviewed grouped exemplars, it became apparent that different respondents were focusing on different aspects of the experience of learning difficulty. In particular, respondents distinguished explicitly between *identity* and *action* as teachers (or prospective teachers). The outcome of "making it hard" can arise from "being a hard teacher," which is a claim about one's inherent identity that in turn influences a range of teacher actions or behaviors. "Making it hard" can also arise from a teacher "choosing" an easy or difficult method, activity or strategy for introducing the material. At this stage, I split the original preliminary code into two working codes: "being a hard teacher" and "imposed difficulty." The latter code was itself distinct from the concept that became the code "inherent difficulty," which describes the incomplete learning associated with specific subject areas that respondents perceived to be cognitively more demanding.

Member checking (Lincoln & Guba, 1985) provides an opportunity to increase accuracy, credibility and validity of qualitative data. Participants review analyzed data for authenticity, alert to coding or proposed connections that carry meanings not intended by the participants themselves. In this study, member checking was performed with participants after the full development of initial codes. I shared the list of codes with three participants selected both in terms of their availability and because their responses collectively spanned the full range of responses that occurred in this study. One participant was a current teacher, one was an undergraduate, and one was a graduate student. I verbally described the initially-emerging patterns that connected some of the

codes. These participants expressed verbal agreement that their experiences were encompassed by the preliminary code list and by the patterns that I described. In particular, the codes were consistent with aspects of their own responses to working with the *DI tool*, and did not misrepresent or distort their responses in idiocentric ways.

Question 2 on the response form asks participants to describe the outcome of the *DI tool* under specified starting conditions. I used responses to this question to ensure that participants were able to manipulate the *DI tool* and understand the output graphs at a basic level. Then, having confirmed that there were no participants whose data needed to be excluded due to inability to use the software, I grouped respondents so as to move toward generating "typical" response profiles (see Appendix C for examples of characteristic responses).

Meaningful response profiles were based on conceptualization of the *DI tool*, not the category of respondent (e.g., not in terms of undergraduate/graduate status or major). This was done as part of the implementation of GT because the responses of pre-service teachers were not consistently similar to each other and different from the responses of graduate students or in-service teachers. Instead, the profiles emerged from my analyses of the responses in light of the research questions. Profiles were centered around participants' views of the bi-modal receptivity distribution of the *DI tool* as either an *inevitable and acceptable* outcome or a *problem or flaw* that must be explained. That is, respondents were unequally divided between those that found that aspect of the model fundamentally consistent with their lived reality of teaching and learning, overlooking minor inconsistencies (n=12, pre-service teachers, 5 graduate students and 2 in-service teachers), and those that found the inconsistencies too overwhelming (n=2, both graduate students). Six respondents (2 pre-service teachers, 3 graduate students and 1 in-service

teacher) did not commit either way, articulating aspects of the model that they found realistic and other aspects that they considered unrealistic.

Axial coding: At this stage, I categorized and grouped codes according to their properties (Strauss & Corbin, 1998, p.126). Some codes became subcategories of other codes. Some were recognized as describing processes, relationships or conditions. As those groupings emerged, I developed "hypotheses," or conditional explanations that linked two or more concepts (Strauss & Corbin, 1998, p.129). I then searched the data for examples and counter-examples, which served to elaborate the conditions of each hypothesis – or provide evidence that the hypothesis was not a good fit with the data after all (Strauss & Corbin, 1998, p.136). Continued "axial coding" also involves identifying the relationships between categories, when relevant (Strauss & Corbin, 1998, p.141).

Using the final code list, I did axial coding to find the connections among codes that exist in the participant responses. In this context, axial coding primarily centered around identifying conditions or circumstances under which specific codes apply. For example, "being difficult" was a preliminary code that turned out to apply to subjects and teachers, but not to students. That preliminary code was separated into "being a hard teacher" and "inherent difficulty." I also grouped some codes to reflect central themes that were becoming apparent. For example, the theme of the "experience of being a student" encompassed the codes "glazing over," "emotions related to learning," "matching expectations to work level," "membership in a particular receptivity group," "narrative of own experience," "student boredom," and "tension between attitude and ability." This grouping revealed that the concept of outcome was not represented, leading to the development of the code "meaning of not learning." Consistent with the

use of GT, these connections derived from the axial coding comprise a substantial portion of the Results section of this thesis.

Another important example of concepts being revealed through axial coding occurred during analysis of the pilot data. An underlying pattern or commonality appeared to exist in the second type of "unrealistic" described above. Several respondents independently described the *DI tool* as unrealistic because it lacked a penalty for agents with maximized receptivities. Not only did those respondents claim that a highly-successful learner (student agent with high receptivity) should still have the opportunity to fail (which is actually a feature of the program), but they argued that such an entity is more likely to fail than an agent with a mid-range receptivity. The suggested mechanism is that "smart" or "good" students get bored and stop paying attention. Surprisingly, this "boredom penalty" was one of the most common early criticisms of the *DI tool*.

When I grouped the codes, "being a hard teacher" and "imposed difficulty" (codes that themselves evolved from close analysis of "making it hard"), the overall role of the teacher was called into question. Returning to the data to investigate this potentially-new concept, I found that respondents agreed that the role of the teacher is to make choices. Responses differed regarding what types of choices (i.e., whether the choices were about students, about the subject, about responding to unexpected events, about instructional methods in a broad sense, etc). "Role of the teacher" became another code. With "role of the teacher" in mind, I searched the data for additional concepts not already encompassed by the existing codes. Two respondents described ways in which their perception of the role of the teacher depends on the particulars of the students involved. The code "teacher responsibility toward students who are behind vs. students who are ahead" captures this context-dependency.

Integration: The final step of GT is integration, or organizing connected codes around a central concept that has broad explanatory power. Integration begins with choosing a central category that relates to many codes and provides a connecting structure for the data as a whole (Strauss & Corbin, 1998, p.147). Additional coding is done to fill out poorly-developed categories, to validate against new data, to check for internal logic and consistency, and to build in variation (Strauss & Corbin, 1998, p.158). This study is exploratory and stops short of developing a fully-integrated theory. In fact, there is no *a priori* reason to believe that participants' experiences of the *DI tool* in this study constitute a single phenomenon that can be integrated *sensu* Strauss and Corbin (1998). Concepts in this data set grouped around clear and provocative themes, but the existence of the distinct participant profiles suggests that this data set may not be fully representational. Additional data are needed.

In addition to the profiles and examples of responses (see Appendices C and D), I felt readers would benefit from hearing participants describe some of their experiences in their own words. The comments included in other sections of this thesis are not single codes, but rather segments of responses selected to illustrate the concepts and connections (Bogdan & Biklen, 2003) that emerged through the process described above. The comments are numbered for ease of reference. It is important to note that these comments do not correspond directly with analytic units. For the purposes of analysis, codes reflect concepts, which are words, phrases or sentences that express a phenomenon, description or connection (Strauss and Corbin, 1998). This definition is consistent with "idea units" (Gee, 1986); expressions need not be restricted to a standard size or number of words. The comments included vary in size and in some cases include surrounding context or several codes. The point of including participant comments in the text is not to

augment the methodology, deepen the descriptions of findings from within the modified GT methodology used in this work.

Chapter 4: Results

The pilot study I conducted produced insight into pre-service teachers' beliefs. How these teachers conceptualize various aspects of teaching, learning and classrooms became apparent through the eliciting effect of the *DI tool*. My goal with this dissertation study, however, was not to explore the beliefs that teachers hold, but rather to explore the mechanisms by which the computer model functions to elicit, engage, or shape beliefs. In other words, it is the affordances and constraints of the model itself that are my focus: the function of drawing the curtain aside using a digital representation, rather than the views revealed once the curtain is drawn.

In this analysis, I attend to the features and characteristics of the *DI tool* that allow it to do organizing and eliciting work (research question #1). Then I describe the theory of mechanism that arose from participant data (research question #2). Finally, I revisit the dimensions of reality first articulated in the literature review, and situate the *DI tool* within a theoretical relationship between representation and reality. Throughout, I use the following quantifiers to indicate the prevalence of specific concepts: "a few" indicates 1-3 respondents, "many" indicates 4-12, "most" indicates 13-16", and "nearly all" indicates 17-19 respondents.

WHAT CLAIMS AND INSIGHTS DO RESPONDENTS GENERATE AND USE TO ORGANIZE THEIR EXPERIENCE?

Realism

Unsurprisingly, participant responses focused heavily what might be called the "realism" of the computer model. Most respondents considered the *DI tool* to be quite realistic, in the sense that multiple elements of their experience using it aligned with aspects of their real-world learning and teaching experiences. There was no need for

realism to be consistent across all aspects of the model, however. Respondents independently assessed the realism of the represented entities (student and teacher avatars, "knowledge"), the rules by which those entities interact, and the outcome of running the model (the stable, bimodal distribution of student receptivity values). In this study, respondents described each element as aligned with personal experience (or not) without regard for the internal logic of the model itself. Specifically, most respondents accepted the outcome and the rules, but questioned whether the representation of students was an adequate fit to the real world. Internal consistency would suggest that an inadequate student representation cannot produce an adequate copy of a group of students. In the case of the *DI tool*, however, the classroom and the process of learning were considered to be a good fit to the real world, even though the students might plausibly have been considered mere caricatures, thereby suspending ongoing engagement by the user.

For example, one respondent suggested adding in "outside factors" (Comment 10) such as a "personal life" (Comment 11) to the representation of students. The same respondent, however, felt that the *DI tool* "ties in well" (Comment 12) with his or her classroom experience, and stated that "student receptivity will depend on the way in which the material is present to them" (Comment 13).

Intuitively, it is easy to see that these types of personal narratives reveal what future teachers see as normal or natural to the classroom habitat. That is an important and non-trivial use of a computer model like the *DI tool*. Responses also reflect the aspects of experience that are perhaps most susceptible to elicitation. The themes that arose in this study are a function of both participant concerns and model affordances. The two themes I describe below are intimately connected with the way the *DI tool* represents processes. This result suggests that it would be possible to design a series of

simple, eliciting models that target narrow aspects of experience. Properly sequenced, and supported by experience in real classrooms, such a set of models could be a powerful tool for teacher learning.

Deterministic vs. probabilistic views

The *DI tool* is probabilistic, in that its numerical values for knowledge difficulty are randomly generated from flat distributions. While a low receptivity makes it more likely for a student avatar to fail, causing its receptivity to decrease, there is no external force keeping receptivity low – only probabilities. In this study, participants particularly struggled with the tension between "unlikely" and "impossible." In the words of one respondent, "Is there a possibility of increasing student's receptivity after they drop below 10%? or are they just doomed?" (Comment 14).

There is a possibility of increasing receptivity from a low value. Unlikely sequences of knowledge difficulty sometimes occur, allowing a student to "cross the divide" (Comment 15) (i.e., move from the low-receptivity group to the high-receptivity group). No one respondent has explored the *DI tool* long enough to see this occur; I have only seen it happen two or three times in the course of programming and testing several versions of the model. There is no threshold beyond which a student will never learn, in the model world. Such a threshold may exist in the realities of study participants, however.

One respondent reported, "I was always worried more about losing students that may be behind, than challenging students who are ahead… because if I lost someone in the beginning, they would eventually be lost forever" (Comment 16). Another described an important time in a class being "…when the kiddos were still trying even though the material was really difficult" (Comment 17). In both examples, there exists an implied

before and *after*; before students were permanently lost or before they gave up, and after these events occurred.

The existence of this threshold suggests that important classroom events are not evenly distributed through time. The *DI tool* is deliberately ambiguous regarding what one iteration of the model's rules actually represents. Respondents agreed that, "initial success is critical" (Comment 18) in both the model world and the real world. They differed widely, however, in their definitions of what knowledge opportunity (i.e., model iteration) might mean. Definitions ranged from an entire run of many iterations representing one question asked by a teacher, to representing a whole school year.

In addition to persistent low receptivity values, other forms of stability exist in the *DI tool*. The average receptivity stabilizes within about 200 iterations of the model. The receptivity of each student stabilizes as well. Though a few respondents argued that most students end up somewhere "in the middle of the pack" (Comment 19), rather than at high or low receptivity values, most respondents explicitly agreed that each student has a relatively stable receptivity. That value is context-dependent and a function of both the teacher and the subject matter; as one respondent explained, "I'd automatically be at a certain receptivity level when going to that class" (Comment 20). Respondents used that context-dependency to place responsibility for receptivity at the feet of the teacher, rather than considering it a fixed property of the student.

"Difficulty" is both positive and negative

The concept of "difficulty" – subjects, teachers, and teaching methods – was another key theme. Respondents uniformly agreed that the teacher controls the perceived difficulty of a learning moment, rather than difficulty being a characteristic of the subject or of the student (level of preparation, for example). A typical response indicating this

was, “I should not make the topic more difficult than the students can handle” (Comment 21). While that seems to be a reasonable principle to follow as a teacher, one wonders: what if the topic is genuinely difficult? Is there such a thing as a topic that is “too difficult” for a particular group of students, and how would a teacher decide?

One artifact of the way the *DI tool* operates is that the maximum student receptivity is capped at the maximum knowledge difficulty, which is controlled by the user on a scale of 1 to 100. If knowledge difficulty is kept low, all the student receptivity values remain low. This phenomenon led one respondent to comment, “I don’t think you need to be a hard teacher to get students to learn” (Comment 22). Many respondents agreed, considering “hard teaching” (Comment 23) to be a source of negativity due to the large number of un-receptive students it created (in the model world). They stated, for example, “I should not make my class difficult all the time” (Comment 24).

A few respondents took the opposing view that difficulty has a positive effect on learning. As one expressed it, “the more you challenge students, the more receptive the receptive ones will be” (Comment 25). This view underlies one consistent criticism of the model: respondents argue that high-receptivity students should be penalized because, in the real world, they would quickly become bored and disruptive. Given the level of academic success achieved by all respondents, they are likely to have been in that high-receptivity group at least some of the time. One wonders whether the “boredom penalty” reflects their real-world experiences or only their real-world fantasies.

HOW DO THE FUNCTIONALITIES OF THE *DI TOOL* SUPPORT WHAT RESPONDENTS SEE AS MEANINGFUL?

Commonalities of emotional experience imply a shared reality

The *DI tool* creates a shared reality among respondents by "smoothing over" idiosyncratic differences in personal experience. It de-emphasizes some aspects of experience and highlights others. In particular, respondents identified strongly with their past emotions and motivations as students in either a high-receptivity or low-receptivity cohort. Most respondents explicitly described their educational experience in terms of their personal receptivity.

Given the high level of academic success achieved by the participants in this study, it is surprising that only one respondent placed him-or-herself firmly in a high-receptivity group. The respondent stated, "I was always a top tier student so I did not think about the 'bottom' kids" (Comment 26).

In contrast to the certainty offered by that respondent, many other respondents described a mix of high-receptivity and low-receptivity experiences, such as:

> I've always been the kind of student who set the curve and aimed to meet the high expectations. I think that there's a point at which that type of behavior breaks down, though. Depending on my interest, I would sometimes (although rarely) zone out completely and not get what was going on in class. (Comment 27)

This respondent admits that his or her identity as a high-receptivity student was context-dependent, not a fixed aspect of personality. Many respondents, though they shared that context-dependency, focused their comments on times when they were in the low-receptivity group. Apparently those negative experiences were the ones vividly evoked by the bimodal distribution that appears in the *DI tool*. These responses provide a poignant look at academic failure and despair:

"If I am doing badly in a class, I don't have much desire to improve or get back on track." (Comment 28)

"When things are too difficult for me in class, it is difficult to stay engaged. Eventually, I tune out and let the kids that get it participate in discussion." (Comment 29)

"It remind[s] me of the glazing over, when you realize you are lost in a class." (Comment 30)

"I do remember that if I got too frustrated, or felt that the material was too hard/given too quickly then my interest in putting effort into the course waned, and therefore my receptivity waned also." (Comment 31)

"I remember taking a very difficult class once ([name of course excerpted], ugh). Early in the semester, I had trouble following the lecture, and after a while, I stopped trying to follow the lecture – since it seemed hopeless, and eventually I stopped going to class all together (making my receptivity definitely = 0)." (Comment 32)

These highly successful students, whose commitment to learning and teaching is taking them into education as a career, independently described very similar negative experiences. The experiences themselves must have made up a tiny proportion of each respondent's academic career – students for whom "zoning out" is a common occurrence do not usually attend college. The *DI tool* effectively created a shared reality (*sensu* Peirce), which we can infer by the eliciting of these common fragments of experience.

Simulacrum per Baudrillard: model replaces reality

Within the experience of each individual, newly-discovered realities have the power to replace the reality that existed at the time of an event. Consider the case of a significant other, startled by your sudden appearance, who hides something from you as you approach. In the moment, your thoughts turn to betrayal, angry feelings rush up, and you pick a fight about where to go for your six-month-anniversary dinner. When you

later discover that the "something" was a surprise gift for you, the actions and words that seemed so justified in the moment are revealed to be foolish, unnecessary, responses to a reality that existed only in your mind. The later, shared reality replaces the experience you had at the time. You never describe the incident in terms of your initial reaction; it becomes "that time I was so mad at first, until I found out what was actually going on." Once you know the full story, all parts of it must be told for the narrative to make sense.

The *DI tool* organizes backward, which means it has that quality of replacing reality. We take for granted that real-world experiences organize backward, but it is not a common feature among computer models. One might imagine that a physics model explains why a rocket follows a certain trajectory, but we do not expect the observer's description of that trajectory to invoke forces that operate only in the model world. We expect the description to invoke real-world forces that the model has drawn attention to.

Like the example of a trajectory description that draws on non-real forces, respondents exploring the *DI tool* interpreted past experiences in model terms. For example, "there usually were 2 groups of students, and sometimes, either every [one] knew the answer, or no one did" (Comment 33). This exactly describes the behavior of the model (if one assumes that receptivity is the same as "knowing the answer") but is rare in real-world classrooms.

Another respondent stated, "If material was too easy or hard, I would initially be receptive but then gradually (or quickly) lose interest" (Comment 34). Again, this description exactly copies the outcome of the model but loses the complex ebb and flow of real-world student life.

A few respondents went one step further toward conflating the real world and the model world: they explained model outcomes by generating new real-world forces. Three respondents invoked a "student receptivity stabilizing force," operating at the

individual or the classroom level, by suggesting, "there needs to be an option for a student's receptivity to not go up or down" (Comment 35), and "there are plenty…of the students who stick around the average receptivity" (Comment 36). These same respondents accepted the realism of the rules by which receptivity increases or decreases in the model, but could not reconcile the logical outcome of those rules with their own classroom experiences.

REPRESENTATION AND REALITY: REACHING THE THRESHOLD OF "SUFFICIENTLY REALISTIC"

In the literature review, I proposed that some digital representations require only a small number of meaningful features to become effective influencers of users' thoughts. Instead of becoming more realistic with the addition of many of interchangeable features, representations of this type reach a threshold level of realism with the inclusion of specific, contextually-important features. I proposed theoretical criteria for recognizing a "threshold-type" representation: 1) it is used to organize forward, 2) suggestions for change focus on how the model communicates with the user, not the relationship between model and reality 3) recommended additional features are patterned or consistent among different respondents. Based on participants' responses, it seems clear that the *DI tool* meets these criteria.

Organizing forward

Most respondents embraced the model as a source of guidelines for future teaching. Their claims center around the implicit challenge that the model proposes:

"It is hard to reach all students." (Comment 37)

"It is difficult to make sure that all students are receptive to a lesson." (Comment 38)

These statements reflect an underlying frustration with the existence (in the model) of low-receptivity students. In this version of the *DI tool*, it is not "hard" to increase the receptivity of all students – it is impossible. Interestingly, respondents interpreted this fact as a challenge to their (future) teaching abilities, rather than as a narrative of futility. What is impossible in the model world is merely difficult in the real world. Only one respondent balked at the consistent student failure by stating, "I couldn't find a setting which would enable most of the class to pass… This is not realistic" (Comment 39).

Other statements proposed a connection between difficulty (of the content or teacher) and success of the students, such as:

"the harder the teacher, the more students learn." (Comment 40)

"If you teach above students ability, then you lose about half [of the students]." (Comment 41)

"The more difficult or challenging the subject or material, [the more likely] you will get students who are absolutely receptive [and] those who do not care at all." (Comment 42)

Again, these statements express likely relationships, rather than the pre-determined outcomes that the *DI tool* actually presents.

Many of these respondents also made personal claims about their own teaching plans. Examples include:

"When I do direct teach, I am probably increasing the receptivity for some students, but decreasing receptivity for others." (Comment 43)

"This model could help me predict how engaged students will be in my own classroom depending on the difficulty of the material." (Comment 44)

Both of these statements demonstrate a willingness to apply the model's rules *and outcomes* to real-world teaching situations. In general, participants in this study experienced the *DI tool* as providing legitimate, relevant, and potentially predictive insight into the nature of teaching and learning.

A few respondents rejected the model, claiming that it did not apply to their anticipated or actual teaching circumstances. Interestingly, even among these respondents, some aspects of the model remained meaningful. For example, one respondent stated that the model did not relate to his or her teaching, but followed up with, "I teach very small classes of strong students… I am able to craft my lessons to fit very exactly to the receptivity of my students" (Comment 45). This respondent accepted the underlying entities of the *DI tool*, as well as the process by which receptivity increases or decreases on an individual level, but rejected the classroom-level effect of bimodal receptivity distribution that appears when the model is run.

Focus on how the representation communicates

Many respondents suggested changes, not in the relationship between the model and reality, but the way that the model communicated with the user. These types of criticisms appeared in both Q5 (how might the model be improved?) and Q6 (what features would you add or remove to make the model more realistic?). The existence of these criticisms demonstrates that the connection between model and reality is not

necessarily the most important aspect of the user's experience. These respondents took for granted that the *DI tool* was realistic enough to be relevant; what they requested was more detailed information about intermediate stages of the model as it was running. The following response is representative:

> I would like to see more graphs/measures of receptivity over time. Although I can watch progression of the graphs over time, I'd love to have this info summarized somehow, maybe as range of receptivity over time + average. Even being able to move around the average receptivity graph or have a separate sliding bar that can recreate the model results over time would be helpful. Or a graph for receptivity range per day. (Comment 46)

Other respondents went into less detail, but made similar requests for additional summarizing graphs, more detail on the existing graphs, and statistics that would capture intermediate states of the model.

Patterned additional features

Any representation can be made more realistic by the addition of features that copy the real world. For a "threshold-type" representation, however, most of the multitude of possible additional features are relatively meaningless. One would expect that user requests for additional features would converge on a few meaningful elements, rather than cover a broad range of possibilities. This is indeed what happened with responses to the *DI tool*.

As described above, many of the respondents requested features that communicated the model processes or results differently, rather than features that changed the relationship between model and reality. A few respondents declined to name any features that would make the model more realistic, writing, for example, "it seems too simplistic for receptivity to only depend on difficulty, but I'm not sure what other features should be captured" (Comment 47). A few respondents did focus on visual

features, suggesting that changing the shape and layout of the colored avatars in the model would increase the realism: "make the model be spread out like a grid (just like desks in a classroom) instead of a circle" (Comment 48). Note that this particular response also reveals the normality of a grid of desks rather than a circle, in the respondent's reality.

Setting aside those three categories, nearly all of the remaining respondents requested the same additional feature: a more complex representation of students. Student in the *DI tool* have only one characteristic (receptivity), represented by a single number. Respondents suggested adding other numerical values to represent academic variability ("prior knowledge level" Comment 49), demographic information ("SES" Comment 50, "income, language barriers" Comment 51), internal classroom events ("things that could distract students, i.e., cellphones" Comment 52) and random variation ("to account for students having 'good' days and 'bad' days" Comment 53).

The fact that responses converged upon this particular additional feature suggest that it is more meaningful than, for example, adding teacher characteristics (one respondent) or modifying the rules that increase or decrease student receptivity (one respondent). Differential meaningfulness is the hallmark of "threshold-type" representations.

Overall, a nuanced set of patterns in interpretation and use did emerge. The complex view of "realism," the tension between deterministic and probabilistic experiences, and the focus on "difficulty" illustrate the *what* of what emerged (research question #1). The claims and insights that respondents generated were a mix of elicited personal narratives and connections between specific elements of the simulation and the real world. This analysis indicates that the *DI tool* was seen as relevant and useful despite a lack of internal consistency. In other words, portions of the model could be

considered as quite *unrealistic* without negating the aspects that were seen as realistic and meaningful.

Respondents struggled with the conflict between the known probabilistic nature of the simulation and their experiences of stabilizing average and individual receptivity values while using it. Their responses suggested that in the real world, as in the model world, past failure has a potentially overwhelming influence on probability of future success. What was different between the two realities was that respondents suggested that there might be a point of no return in the real world, or failure after which success is impossible. The existence of such a point implies that the timing of classroom events is key and, indeed, respondents interpreted the meaning of one model iteration in a variety of ways, to suit the other claims they were making.

The *DI tool* seems to promote a shared perspective on certain past classroom experiences, which speaks to *how* the software supported what respondents saw as meaningful (research question 2). Users readily co-opted the concept of student "receptivity" to interpret their memories of success and failure. Surprisingly, the existence of a large group of low-receptivity students in the model seemed particularly provocative to respondents. Despite the high level of academic achievement among study participants, many of them focused on narratives of their own "low-receptivity" experiences.

Finally, the *DI tool* meets the proposed theoretical criteria for a "threshold-type" representation. Having demonstrated that this type of representation exists, in contrast to the assumed benefit of indefinitely increasing realism, I can now begin to situate these findings relative to broader issues of representation and reality in the use of simulations, especially those simulations related to teaching.

Chapter 5: Reflections and Future Research

ARTICULATING AFFORDANCES AND CONSTRAINTS OF THE *DI TOOL*

The results of research in the intersecting areas of design and use of dynamic interactive visualizations points to the need for more theory of how these tools interact with users. The rapid development of these tools has outstripped the systematic understanding of them, and it has especially outstripped our understanding of the design process involved in creating them.

Research on other dynamic interactive visualizations shows that even minor design elements of a particular visualization have a substantial role to play in shaping the behavior and interpretations of users. One extension of this study is to investigate in detail the specific elements of the *DI tool* that appeared to influence users. For example, respondents in this study were surprisingly committed to the teacher as "in control" of learning, and particularly as the arbiter of difficulty. That stance is visually supported by the central placement of the teacher avatar and the description of knowledge as being "broadcast" from the teacher. Other versions of the *DI tool* could be developed that position the teacher among the students, or lack a teacher avatar altogether. By parallel with research on the design of PhET simulations, there is reason to believe that the relatively small visual difference could have a disproportionately large effect on user interpretations of the teacher's role.

In addition to direct comparisons among variations of the simulation, other learning-related phenomena could be modeled with simple programs similar to the *DI tool*. For example, some aspect of student-to-student interactions might be represented by each agent's "affinity value" influencing nearby affinities through a contagion process. Groups of students would form or disperse based on a defined relationship

between their affinities. Such a model would retain the ambiguous naming conventions that, in the case of the *DI tool*, allow users to impose their own experiences on the actions and interactions of the agents. It would also focus attention on a single aspect of the representation, potentially generating a commonality of experience in the way that the "low-receptivity" cohort in the *DI tool* seemed to generate a common reflection of experiences of failure in this study.

Video and animated representations of teaching afford the possibility of relating events through time, while still re-experiencing them outside of the chronology (by pausing or rewinding). The *DI tool* clearly represents time passing, but is deliberately ambiguous about the real-world interpretation of one iteration of the program. A way to retain that ambiguity and simultaneously increase the focus on time is to invite users to set the length of the "days" function (using a slider). Is one day represented by one iteration of the program, or by tens of iterations? Though I did not ask respondents in this study to specify the real-world equivalent of the *DI tool*'s "knowledge opportunity," related responses suggest that there is wide variance in how much time is perceived to have passed before the stable outcome is reached. In a follow-up study, I would analyze the justifications that users offer for setting a "day" as a particular number of program iterations.

Though ambiguity of elements is a key feature of the *DI tool*, allowing users to project their own interpretations on the events and entities of the program, another direction for future research is to modify the level of ambiguity. Users are likely to respond differently to the program if, for example, "receptivity" is called "ability" or "motivation" instead. Defining elements in more concrete terms may elicit additional ideas or narratives about teaching and learning that are otherwise hidden.

PLACE OR PROCESS

In particular, one of the key abstractions of the *DI tool* is its sense of place, of the classroom as a physical space. The *DI tool* is designed to be almost purely "process," and most respondents willingly ignored "place," or the physical and locational aspects of teaching. One wonders how a visualization of place without process would look: is it possible to recognize a classroom simply from the relative positions and movements of the entities? This would be an interesting question to explore.

If such a program did prove to be recognizable, it would support the value of varied representations in concretizing abstract ideas. Wilensky (1991) argues that learning consists in part of experiencing many representations and uses of a concept. It is through the relationship that grows between the learner and the concept that the initial abstraction becomes first intimately familiar, then a part of the learner's concrete, "robust and familiar" knowledge. The *DI tool* and its hypothetical place-representing counterpart might be disproportionately effective when used together as a pair of linked representations.

VALUE OF ABSTRACTION IS MEDIATED BY THE DESIRED END

In a more general sense, the *DI tool* is a further abstraction along the continuum established by live observation to video to animation. It can be considered a generalization about teaching. The fact that the *DI tool* is deliberately distorted ought to place it as a peripheral example rather than a core prototype of teaching (Rosch, 1973). Instead, respondents in this study treated the outcomes of the *DI tool* as typical of classroom teaching, or as exemplifying the bulk of their experiences as teachers and learners. If this tool can represent teaching effectively despite its simplification and

distorting intent, what features would be necessary to include (or exclude) to create a representation that is consistently perceived to be "wrong" rather than "right?"

Finally, uncoupling the assumed relationship between utility and realism provides a new lens through which to analyze computer representations in many disciplines, not just representations of teaching. Responses to the *DI tool* strongly suggest that a highly abstracted model that includes a few meaningful features can be just as effective (and much simpler to design and program) as a model that indiscriminately incorporates as many features as the technology will allow. More visual realism, more literal realism, is not always better. The partial realism provided by a highly-constrained computer environment may, in fact, lead to more and better learning than the sometimes extensive instructions needed to constrain the real world to meet learners' needs (McElhaney & Linn, 2011).

"Direct measurement videos" of physics demonstrations illustrate both the value and the limitations of turning the decision-making about constraints over to designers. These videos provide a semi-structured, time-controllable way for individual students to interact with the demonstrations beyond mere observation (Bohacek, Vonk, Iverson, & Kirk, 2013). The videos are overlaid with tools, such as a frame counter, ruler, protractor, etc. Each student can independently collect data about the event using the tools and re-winding the video as needed. In this way, students need not rely on the instructor to provide the data, nor do they need to be skilled with tools such as distance rangers or force sensors – indeed, they do not even need access to such tools. The teacher, in turn, avoids the difficulties associated with ensuring that a full set of classroom equipment is operational and synchronizing appropriately with computers. With measurements made from the "in-video" tools, students can calculate experimental values or work related physics problems.

Clearly, live data collection by students of physics places many constraints on the learning situation. Only a few types of events can be reliably enacted and measured in the typical high school or college laboratory setting, restricting the examples that learners consider. The creation of tables and graphs, however, becomes almost automatic with the use of electronic probes and sensing equipment synchronized to a computer.

Video cases give students access to a much broader range of physics events, and the in-video tools are essentially transparent even to new users. Learners have to take on the task of converting the event to a graphical or tabular representation, though. In addition, the exact problems that can be solved are highly constrained. Any student holding a distance ranger has the potential to create experiments that the instructor did not envision; students using a direct measurement video have little to no such opportunity.

The underlying tension revealed by this new tool for learning physics is: are students learning to do physics in the sense of making measurements and working calculations, or are they learning to become physics thinkers by creating experiments that reveal or test their understanding?

Designing constraints that help students focus on one of those two activities is a labor-intensive process for teachers. A well-designed visualization stands mostly alone as a learning event, without much instruction needed to facilitate the interaction of learners with the visualization. Whether this is a helpful reduction of the teachers' "burdensome chores" (Skinner, 1958) or a negative mechanization of the art of teaching remains an open question. The substantial influence of context suggests that, at least with current technologies, plenty remains for teachers to do in their role as designers of learning experiences. I view a clear and detailed understanding of the *DI tool* and other

representations, whatever the subject matter, as a benefit. With a better tool ready to hand, what are teachers and learners now free to do?

Appendices

APPENDIX A: SURVEY QUESTIONS

Page 1:

Q1: From which instructor/course did you hear about this study?

Page 2:

Q2: Using the two graphs, what happens if you set number-of-students to 45 and maximum-difficulty to 100?

Q3: If a group or classroom of students was like this in reality, what would it mean?

Q4: What does this model tell you about your own teaching (if anything)?

Q5: How might the model be improved?

Q6: What features would you add or remove to make this model more realistic?

Q7: Does this model explain or connect to any of your past experiences as a student?

Q8: Does the model's behavior fit with what you expected?

APPENDIX B: FINAL CODES LIST

"Glazing over" or "tuning out"

Being a hard teacher vs. being an easy teacher

Class size

Classroom randomness: students, events, teacher, subject

Consistent teaching vs. variable teaching

DI tool predicts what will happen in real world

Different instructional methods

Effective sequencing of instructional methods or activities

Emotions related to learning

Fixed receptivity (students who “stick” around the average)

Future is influenced by past experiences

Increasing difficulty causes increased receptivity

Inherent difficulty

Initial success is critical

Matching expectations to work level (student effort level)

Meaning of “not learning”

Meaning of low receptivity (ability, self-esteem, willingness to work, attention to subject, discouraged, low motivation, “too difficult, fail, angry, stressed and upset, therefore learn less”

Membership in a particular receptivity group

Narrative of own experience

Paying attention

Proportion of class that has low receptivity

Real world

Receptivity is context-dependent (too easy, too hard, or “just right”)

Role of teacher is to reach all students

Student attributes: SES, accommodations, distractions, attitude, lack of sleep, level of prior knowledge

Student boredom/boredom penalty

Student identity fixed (“D student” vs. “A student”)

Students are unique/cannot be generalized as a group

Students have a learning “fate”

Teacher responsibility toward students who are behind vs. students who are ahead

Teacher-created difficulty

Tension between student willingness/attitude and student ability

Unrealistic

APPENDIX C: TYPICAL RESPONSES BY PROFILE

Model consistent with reality

It [the model] is pretty accurate. Because, during direct instruction in classrooms, there is always a certain number of students that "tune out" and those would be the students at 0 receptivity. I work harder at first and then when I know what the expectations are I only work as hard as I need to work. Over time, I as a student, desired different instructional methods to make the content more engaging. It is difficult to make sure all students are receptive to a lesson. I did not have a set expectation when I began running the model but with the given parameters it does fit with my experience that if students are constantly feeling challenged, they do not always feel like they are making progress and so may give up. I like the "day feature, so…maybe having a single day's results isolated and shown separately so we can understand how quickly we might lose students on any day given the material.

Model not consistent with reality

Direct instruction is "at best" 50% effective according to the model; however, I don't necessarily see this myself. In my own teaching experience, I've found that direct instruction is very effective. I've taught all levels of math students, and never had a true issue with this. I expected some students' receptivity to go down and others to go up, but

[not as much as this]. I'd allow different variables to be changed for different students. I don't think that a univariate analysis provides sufficient information.

Mixed response

[The model connects to past experiences] pretty much… big classes, it is very difficult to get involve[d]. But I think that you don't need to be a very hard teacher to get your students to learn. [The model's behavior fits] a little bit. However, in reality, because we have many integrated factors associated with difficulty of learning, we can build much fitter [models] in the future. I suppose I would have been happy if half got it.

APPENDIX D: SELECTED RESPONSES

Pre-service teacher:

Q1: From which instructor/course did you hear about this study?

Dr. Stroup (Classroom Interactions)

Q2: Using the two graphs to explain, what happens when you set number-of-students to 45 and maximum-difficulty to 100?

What happens if that at the beginning of the model, receptivity is somewhat evenly distributed in the center. However, as the model progresses it is evident that receptivity goes to the polar opposite sides. There is no longer a middle ground, but rather a distinct black and white distinction. Which side it favors simply depends upon the randomness of the model.

Q3: If a classroom or group of students was like this in reality, what would it mean?

This means that eventually, your class will be solidly divided between a very high receptivity, and a very low receptivity. However, it also means that the more you challenge the students, the more receptive, the receptive ones will be.

Q4: What does this model tell you about your own teaching, if anything?

That the more difficult or challenging the subject or material, you will get students who are absolutely receptive but those who do not care at all.

Q5: How might the model be improved?

Make the graphs more reader friendly.

Q6: What features would you add or remove to made this model more realistic?

Possibly add in outside factors. These factors could include income, language barriers, personal life, etc.

Q7: How does this model connect to your past experiences as a student?

I think it ties in well. After a certain degree of just a teacher speaking, you are either going to be actively involved or you are not going to care.

Q8: Does the model's behavior fit with what you expected?

Yes it does. Student receptivity will depend on the way in which the material is presented to them.

Current teacher:

Q1: From which instructor/course did you hear about this study?

Other (please specify) current teacher

PAGE 2

Q2: Using the two graphs to explain, what happens when you set number-of-students to 45 and maximum-difficulty to 100?

More than half of the "students" failed. This happened very quickly. Average Receptivity leveled out at about 54.8, and the Receptivity Distribution was relatively evenly split between

Q3: If a classroom or group of students was like this in reality, what would it mean?

Meaning the extreme run of 45 students/difficulty 100? That the teacher was teaching material that was too difficult for the students. Slightly more than half of students would fail (and be justifiably angry), and all of them would probably be stressed and upset, and therefore learn less. If you mean the model as a whole, I'm not really sure. I think there are a lot of human and environmental factors that this doesn't/can't take into account, such as attitude, lack of sleep, etc. which might affect a student's day-to-day receptivity.

Q4: What does this model tell you about your own teaching, if anything?

Not much, really. I teach very small classes of strong students. I am able to give individual attention to struggling students, and therefore only rarely have anyone fail an assignment or test. I am able to craft my lessons to fit very exactly to the receptivity of my students.

Q5: How might the model be improved?

I couldn't find a setting which would enable most of the class to pass. It seemed like regardless of the number of students or level of difficulty, consistently slightly more than half of the students failed. This is not realistic.

Q6: What features would you add or remove to made this model more realistic?

I'm not sure. It would be nice if it were *possible* to create a scenario in which the whole class maintained a higher level of receptivity, for example, some model representing a multi-pronged approach to delivering material, so that students with higher

receptivity received more difficult material and students with lower receptivity received easier (or more slowly presented) material.

Q7: How does this model connect to your past experiences as a student?

I've always been pretty self-driven, so as a student, any deficiencies in my classes I made up by reading/studying on my own. I do remember that if I got too frustrated, or felt that the material was too hard/given too quickly then my interest in putting effort into the course waned, and therefore my receptivity waned also.

Q8: Does the model's behavior fit with what you expected?

more or less. I thought changing class size and difficulty of material would affect the results more, but the shapes of the graphs remained more or less the same.

References

Adams, W. K., Reid, S., LeMaster, R., McKagan, S. B., Perkins, K. K., Dubson, M., & Wieman, C. E. (2008). A Study of Educational Simulations Part 1 – Engagement and Learning. *Journal of Interactive Learning Research, 19(3),* 397-419.

Anderson, J. R., Conrad, F. G., & Corbett, A. T. (1989). Skill acquisition and the LISP Tutor. *Cognitive Science, 13,* 467-506.

Anderson, J. R., Corbett, A. T., Koedinger, K., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences, 4(2),* 167-207.

Arai, K. K., Deguchi, H., & Matsui, H. H. (2005). *Agent-based Modeling Meets Gaming Simulation*. Tokyo: Springer.

Aristotle. (2006). *Poetics*. Newburyport, MA: Focus Publishing.

Barab, S., & Dede, C. (2007). Games and immersive participatory simulations for science education: An emerging type of curricula. *Journal of Science Education and Technology, 16,* 1-3.

Barab, S. A., Scott, B., Siyahhan, S., Goldstone, R., Ingram-Goble, A., Zuiker, S. J., & Warren, S. (2009). Tranformational play as a curricular scaffold: Using videogames to support science education. *Journal of Science Education and Technology, 18,* 305-320.

Barab, S., Thomas, M., Dodge, T., Carteaux, R., & Tuzun, H. (2005). Making learning fun: Quest Atlantis, a game without guns. *Education Technology Research and Development, 53(1),* 86-107.

Baudrillard, J. (1983). Simulacra and simulations: Disneyland. In M. Poster (Ed.), *Semiotext(e)*. Brooklyn, NY.

Bell, R. L., & Trundle, K. C. (2008). The use of a computer simulation to promote scientific conceptions of moon phases. *Journal of Research in Science Teaching, 45(3),* 346–372.

Bogdan, R., & Biklen, S. (2003). *Qualitative research in education* (4th ed.). Boston: Allyn & Bacon.

Bohacek, P., Vonk, M., Iverson, E., & Kirk, K. (2013, July 22). Using Direct Measurement Videos to Teach Physics. *Pedagogy in Action: the SERC Portal for Educators.* Retrieved August 1, 2013, from http://serc.carleton.edu/sp/index.html.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences 99 (suppl. 3)*, 7280-7287.

Booth, J. L., Lange, K. E., Koedinger, K. R., Newton, & Newton, K. J. (2013). Using example problems to improve student learning in algebra: Differentiating between correct and incorrect examples. *Learning and Instruction, 25,* 24-34.

Chazan, D., & Herbst, P. (2012). Animations of classroom interaction: Expanding the boundaries of video records of practice. *Teachers College Record, 114, 030307*.

Chieu, V. M., Herbst, P., & Weiss, M. (2011). Effect of an animated classroom story embedded in online discussion on helping mathematics teachers learn to notice. *Journal of the Learning Sciences, 20(4),* 589-624.

Chieu, V. M., & Herbst, P. (2011). Designing an intelligent teaching simulator for learning to teach by practicing. *ZDM – International Journal of Mathematics Education, 43,* 105-117.

Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, Instruction, and Research. *Educational Evaluation and Policy Analysis 25(2),* 119–142.

Creswell, J. (2003). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (Second Edition).* Thousand Oaks, CA.: Sage Publications, Inc.

De Waal, C. (2013). *Peirce: A guide for the perplexed.* New York, NY: Bloomsbury Academic.

Finkelstein, N. D., Adams, W. K, Keller, C. J., Kohl, P. B., Perkins, K. K., Podolefsky, N. S., Reid, S., & LeMaster, R. (2005). When learning about the real world is better done virtually: A study of substituting computer simulations for laboratory equipment. *Physical Review Special Topics - Physics Education Research, 1, 010103*.

Finkelstein, N. D., Adams, W. K., Keller, C., Perkins, K., Wieman, C., & the Physics Education Technology Project Team. (2006). High-tech tools for teaching physics: the Physics Education Technology Project. *Journal of Online Teaching and Learning, September*.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, Motivation, and Learning: A Research and Practice Model. *Simulation & Gaming, 33(4),* 441–467.

Gee, J. P. (1986). Units in the production of narrative discourse. *Discourse Processes, 9,* 391-422.

Glaser, B. G. (2001). *The grounded theory perspective: Conceptualization contrasted with description.* Mill Valley, CA: Sociology Press.

Glaser, B. G. (2004). Remodeling Grounded Theory. *Forum: Qualitative Social Research, 5(2),* Art.4. Retrieved from http://www.qualitative-research.net/index.php/fqs/article/view/607/1315#g33

Glaser, B. G., & Strauss, A. L. (1967). *The Discovery of Grounded Theory.* Chicago: Aldine.

Halverson, R., & Smith, A. (2009). How new technologies have (and have not) changed teaching and learning in schools. *Journal of Computing in Teacher Education, 26(2),* 49–54.

Hammer, D. (1997). Discovery learning and discovery teaching. *Cognition and Instruction, 15(4),* 485–529.

Heffernan, N. T., Koedinger, K. R, & Razzaq, L. (2008). Expanding the model-tracing architecture: A 3rd generation intelligent tutor for algebra symbolization. *International Journal of Artificial Intelligence in Education, 18(2),* 153-178.

Herbst, P., Chazan, D., Chen, C.-L., Chieu, V.-M., & Weiss, M. (2011). Using comics-based representations of teaching, and technology, to bring practice to teacher education courses. *ZDM – International Journal of Mathematics Education, 43,* 91-103.

Hickey, D. T., Ingram-Goble, A. A., & Jameson, E. M. (2009). Designing assessments and assessing designs in virtual educational environments. *Journal of Science Education and Technology, 18,* 187-208.

Humphrey, T., & Gutwill, J. P. (2005). *Fostering Active Prolonged Engagement: The Art of Creating* APE *Exhibits*. San Francisco: Exploratorium.

Lincoln, Y., & Guba, E., (1985). *Naturalistic Inquiry*. Newbury Park C.A.: Sage Publications.

Linn, M. C., Chang, H.-Y., Chiu, J. L., Zhang, Z. H., & McElhaney, K. (2010). Can desirable difficulties overcome deceptive clarity in scientific visualizations? In A. Benjamin (Ed.), *Successful remembering and successful forgetting: a Festschrift in honor of Robert A. Bjork* (pp. 239-262). New York: Routledge.

Lortie, D. C. (2002). *Schoolteacher: A Sociological Study.* The University of Chicago Press.

McElhaney, K. W., & Linn, M. C. (2011). Investigations of a complex, realistic task: Intentional, unsystematic, and exhaustive experimenters. *Journal of Research in Science Teaching, 48(7),* 745-770.

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2009). *Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies.* Washington, D.C.

Meir, E. (1998). EcoBeaker. Boston: SimBio.

Miller, L. D., Shell, D., Khandaker, N., & Soh, L.-K. (2011). Teaching Using Computer Games. *Journal of Educational Technology Systems, 39(3),* 321–343.

Peirce, C. (1997). *Pragmatism as a principle and method of right thinking.* (P. Turrisi, Ed.). Albany, NY: State University of New York Press.

Podolefsky, N. S., Perkins, K. K., & Adams, W. K. (2010). *Physics Review Special Topics – Physics Education Research, 6, 020117.*

Rawitsch, D. (1978). Oregon Trail. *Creative Computing, 4(3),* 132-139.

Resnick, M. (1996). Beyond the centralized mindset. *The Journal of the Learning Sciences, 5(1),* 1–22.

Riley, D. (2007). Educational technology and practice: Types and timescales of change. *Journal of Educational Technology and Society, 10(1),* 85–93.

Rogers, C. (1951). *Client-centered Therapy: Its Current Practice, Implications and Theory.* Boston: Houghton Mifflin.

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology, 4(3),* 328-350.

Roth, W.-M., Woszczyna, C., & Smith, G. (1996). Affordances and constraints of computers in science education. *Journal of Research in Science Teaching, 33(9),* 995-1017.

Santagata, R., & Guarino, J. (2011). Using video to teach future teachers to learn from teaching. *ZDM – International Journal of Mathematics Education, 43,* 133-145.

Schwartz, J. L. (2007). Models, simulations, and exploratory environments: A tentative taxonomy. In R. A. Lesh, E. Hamilton, & J. J. Kaput (Eds.), *Foundations for the Future in Mathematics Education.* Lawrence Erlbaum.

Schwarz, C. V, & White, B. Y. (2005). Metamodeling knowledge: Developing students' understanding of scientific modeling. *Cognition and Instruction, 23(2),* 165–205.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15(2),* 4–14.

Skinner, B. F. (1958). Teaching machines. *Science, 128(3330),* 969-977.

Skinner, B. F. (1961). Why we need teaching machines. *Harvard Educational Review, 31,* 377-398.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research (2nd ed.).* Thousand Oaks, CA: Sage Publications.

Stroup, W. M. (2013). *Teaching*. University of Texas. Retrieved from http://www.edb.utexas.edu/faculty/wstroup/teaching.htm.

Stroup, W. M., Ares, N. M., & Hurford, A. (2004). A taxonomy of generative activity design supported by next-generation classroom networks. *Psychology of Mathematics Education – North America.* Ontario, Canada, pp. 837-846.

Suchman, L. (1987). Plans and Situated Actions: The Problem of Human-Machine Communication. Cambridge: Cambridge University Press.

Svihla, V., & Linn, M. C. (2011). A design-based approach to fostering understanding of global climate change. *International Journal of Science Education,* 1-26.

Sweeney, A. E., Bula, O. A., & Cornett, J. W. (2001). The role of personal practice theories in the professional development of a beginning high school chemistry teacher. *Journal of Research in Science Teaching, 38(4),* 408–441.

Weizenbaum, J. (1966). ELIZA – A Computer Program For the Study of Natural Language Communication Between Man and Machine. *Communications of the ACM 9 (1),* 36-45.

Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation.* San Francisco: W.H. Freeman and Company.

White, B. C. (2000). Pre-service teachers' epistemology viewed through perspectives on problematic classroom situations. *Journal of Education for Teaching, 26(3),* 279–305.

Wieman, C. E., Adams, W. K., Loeblein, P., & Perkins, K. K. (2010). Teaching physics using PhET simulations. *The Physics Teacher, 48,* 225-227.

Wilensky, U. (1991). Abstract meditations on the concrete and concrete implications for mathematics education. In I. Harel & S. Papert (Eds.), *Constructionism*. Norwood N. J.: Ablex Publishing Corp.

Wilensky, U. (1999). NetLogo. Evanston, IL: Center for Connected Learning and Computer-Based Modeling, Northwestern University. Retrieved from http://ccl.northwestern.edu/netlogo/

Wilensky, U., & Stroup, W. (1999). Participatory simulations: Network-based design for systems learning in classrooms. *Proceedings of the Conference on Computer-Supported Collaborative Learning*, Stanford University.

Winn, W., Stahr, F., Sarason, C., Fruland, R., Oppenheimer, P., & Lee, Y.-L. (2006). Learning oceanography from a computer simulation compared with direct experience at sea. *Journal of Research in Science Teaching, 43(1),* 25–42.

Zibit, M., & Gibson, D. (2005). simSchool: The game of teaching. *Innovate, 1(6).* Retrieved from http://www.innovateonline.info/index.php?view=article&id=173

Zhang, Z. H., & Linn, M. C. (2011). Can generating representations enhance learning with dynamic visualizations? *Journal of Research in Science Teaching, 48(10),* 1177-1198.