A Longitudinal Examination of Middle School Science Learners' Use of Scaffolding In and Around a Dynamic Modeling Tool

by

Eric Bruce Fretz

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Education and Psychology)
in The University of Michigan
2010

Doctoral Committee:

Professor Joseph S. Krajcik, Chair
Associate Professor Elizabeth A. Davis
Associate Professor Priti R. Shah
Assistant Professor Christopher L. Quintana

To my father,

Dr. Bruce Fretz,

and my mother,

Dr. Barbara Fretz,

for setting the example;

and to my wife,

Dr. Jennifer Fretz,

for making all things possible.

ACKNOWLEDGEMENTS

In the course of a 12 year journey, the list of those to whom I am indebted grows long indeed! We will start with my family – parents, wife, and children – who have borne the burden of having to share me with graduate school, and who have, in the case of my amazing PhD-toting parents, shown legendary restraint in NEVER ONCE harassing me about finishing. And my beautiful, brilliant wife Jennifer – without my "spousal fellowship" grants, I would never have been able to continue this journey, so thank you Peanut!

My time in graduate school has been an endless tour of superb scholars, helpers, and mentors. I thank my colleagues who helped me collect our KDI data, Dr. Hsin-Kai Wu, and Dr. Bao-Hui Zhang, as well as one of the PI's of the KDI project Dr. Elliot Soloway who makes cool educational software happen. My graduate program, the Combined Program in Education and Psychology has been unflagging in its support, generous to point of unreason, and I will always appreciate the kind but firm guidance of chairs Dr. Paul Pintrich and Dr. Kevin Miller, as well as the loyalty and sterling professionalism of CPEP's Janie Knieper.

I want to thank my advisor, Dr. Barry Fishman for tirelessly supporting his "oddest and lowest maintenance advisee" - I appreciated your counsel and friendship over the years.

I want to thank the science teachers with whom I worked and collected this data. Your school is amazing, but your classrooms are the best part of an amazing school. Thank you for showing me just how good great teaching can be.

I want to make a broad general thank you to the many fellow students and excellent professors who contributed in innumerable small ways to my personal and intellectual development and the School of Education staff who looked out for me along my journey. The University of Michigan is truly a source of excellence, and I am proud to be a part of it.

To Dr. Juliane Blazevski and Mr. Stephen Zekany, I owe great thanks for their tireless efforts in assisting me in polishing this dissertation to whatever luster it possesses. Your technical expertise and "outsider" perspective, as well as your professional perfectionism, were just what I needed in the final miles of this marathon. I look forward to helping you with your PhD someday, Steve, not that you'll need it!

I want to thank Dr. Phyllis Blumenfeld for teaching me and showing me what it means to think and write and act like an Educational Psychologist, and for bringing the concept of "academic press" to my existence. I have profound respect for your teaching and scholarship, and I hope to always measure up to "The Phyllis Standard" in my future work.

I want to thank my committee. To describe you as patient and long-suffering would be a gross under-characterization! I have, on more than one occasion, when asked to name my committee members, gotten a raised eyebrow with a smile and the statement, "that might be the best committee I have ever heard of". I concur. You have been unfailingly supportive through my various travails, and your many revisions were all so on the mark and just made this dissertation so much better – THANK YOU SO MUCH! Dr. Chris Quintana, once a fellow student (was it really back in the 1900's?!), thanks for creating interesting software, being a model teacher/researcher, and showing me it is possible to graduate! Dr. Priti Shah, thank you for being a friend and for teaching the most interesting classes and modeling successful work/life balance. Dr. Betsy Davis, thank you for collegially and patiently mentoring me through my first lead author experience (an honor I am still not sure I deserved), and for helping shape this study from its earliest stages.

And finally my chair, Dr. Joe Krajcik, who is the only reason I stayed in graduate school. After my second year, having switched research groups, behind in my progress, doubting my own qualifications, you turned to me as we prepared our group KDI AERA presentation and said, “You should present our stuff”. Forcing that shift from “doubting worker-bee” to “assertive colleague presenting the team’s work” was the hinge around which my whole grad school experience pivoted, and while it might have been just another minor mentoring moment to you, I hope you will take this as a reminder that there are no small mentoring moments! I know I have exasperated you at times with my slowness and tentativeness. While my life has contained its share of drama in the last decade, I look back and realize I could have gotten this done sooner. I apologize to you and to everyone for not working harder and faster, but in the end, you can only do your best in the moments you have. Your patience and thoughtful guidance, and your friendship, have meant much to me, and I hope we can collaborate in the future. I will never be a “rock star” of science ed like you, but you have shown and taught me a lot about being excellent in general. I hope to give back to academia over the coming years, and when I do so, it will be in your honor.

On My Honor, I Will Do My Best...

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

A Longitudinal Examination of Middle School Science Learners' Use of Scaffolding In and Around a Dynamic Modeling Tool

by

Eric Bruce Fretz

Chair: Joseph S. Krajcik

Scaffolding is a term rooted in multiple research communities over decades of development. Customized, contingent support can be provided to learners to enable performances beyond what they can do alone. This dissertation seeks to examine how effectively scaffolds designed to promote articulation (written expressions of learner understanding) actually work, and if this effectiveness and/or the quality of the resulting models changes over time. It longitudinally examines the use of scaffolds designed into a dynamic modeling tool, as it is used by middle school science learners to create, test, and revise models of complex science phenomena like stream ecosystems. This dissertation also reviews the origins of the scaffolding construct, and summarizes conceptions of scaffolding from various lines of research. Scaffolding can be provided by both human and non-human agents, such as computers, which require specialized interface design to ensure maximum effectiveness.

In the study, learners created models in four curriculum units over the seventh and eighth grade school years. Additionally, this dissertation examines the nature of the discussion learners have while using these scaffolds and the frequency and types

of interpersonal scaffolds employed during the creation of models. Model quality is also examined, using a rubric developed through review of prior research on assessing models and concept maps. Learner pairs' model creation sessions on a computer are captured with screen video and learner audio, and then distilled to transcripts for subsequent coding and analysis, supported by qualitative analysis software.

Articulation scaffolds were found to succeed in promoting articulations and the quality of those articulations improved over time. Learner dialog associated with these written articulations is of reasonable quality but did not improve over time. Quality of model artifacts did improve over time. The overall use of scaffolding by each learner pair was contrasted with that pairs model quality, but no relationship was found.

Software design and classroom implementation implications of these findings are discussed. The frequency of interpersonal scaffolding provided by teachers highlights the need to consider scaffolding holistically and synergistically, with design decisions for software tools made in light of careful analysis as to what human and non-human agents can and should each provide.

CHAPTER I

Introduction

> "Most of the psychological investigations concerned with school learning measured the level of mental development of the child by making him solve certain standardised problems. The problems he was able to solve by himself were supposed to indicate the level of his mental development at the particular time ... We tried a different approach. Having found that the mental age of two children was, let us say eight, we gave each of them harder problems than he could manage on his own and provided slight assistance ... We discovered that one child could, in cooperation, solve problems designed for twelve year olds, while the other could not go beyond problems intended for nine year olds. The discrepancy between a child's mental age [indicated by the static test] and the level he reaches in solving problems with assistance is the zone of his proximal development."
> (Vygotsky "Thought and Language" 1986 p. 186-7)

Scaffolding has been defined, among other things, as support provided in the Zone of Proximal Development (ZPD), the important window of educational opportunity between what a learner can understand or do without assistance, and what they can understand or do with assistance. The provision of scaffolding is a key aspect of any learning dyad, and is central to the power of one-on-one tutoring, such as Bloom's

"2 sigma" effect. Bloom (1984) noted that no other educational intervention thus devised showed the ability to move a learner two full standard deviations upward in the same way that focused one-on-one instruction can. With education in general, and science education in particular coming under considerable political and public scrutiny (AAAS,1993; NRC, 1996, 2000), efforts to improve science education are of particular importance. Specially designed software tools for science learning can play an important role in such efforts.

Drawing on constructivist principles, and integrating the tremendous potential of computers and software in education, a number of science educators have developed software tools integrated with curriculum and other supports to attempt to improve or reform science education. These efforts are informed by prior theoretical work of constructivists like Piaget and Vygotsky, but also seek to extend these theories and account for the unique context of combining learners, teachers and technology in Interactive Learning Environments (ILE's). In these efforts, a scaffold, or scaffolding, is considered both as a design principle and as a way to develop scientific knowledge and competence in the most efficient manner.

Scientific modeling is an important aspect of developing scientific knowledge and competence. Modeling is highlighted in science education reform efforts and is a benchmark for scientific literacy (AAAS, 1993; NRC, 2007). In order to provide the cognitive benefits of model creation to learners, a number of computer based modeling tools have been developed by researchers. Providing a tool that enables young and novice learners to make dynamic models of complex systems requires more attention to interface design and pedagogy than simply creating a modeling tool for scientists or experts to use.

This study examines middle school science learners who are using scaffolded technology tools in an ILE. The tool is dynamic modeling software that allows even young learners to create, test, and revise dynamic models of scientific phenomena (e.g., a stream ecosystem). I am interested in how the scaffolds designed into the tool are used, and how this and the scaffolding behaviors used by teachers and learners support the goal of creating substantive models that both represent, and help develop, robust scientific understanding.

There is currently debate about the nature of scaffolding, and what can be a scaffold. Additionally, the interaction of scaffolding designed into technology tools with the scaffolding provided by human agents is currently of interest. Design frameworks for scaffolds in technology tools have been proposed, but have not been fully validated. Methods for evaluating dynamic concept maps or models are evolving, but precise standards are not agreed upon. Within the ongoing debate about scaffolding, evidence linking the successful use of scaffolds, both in and around technology tools, to successful learner outcomes (e.g., final artifact quality) would be helpful.

This study is designed to investigate questions of scaffold use and model quality, and possible relation between the two. Two major questions and four sub-questions are more fully elucidated in chapter 2. By examining data from three classrooms across two academic years, generalizability is enhanced and changes over time can be examined.

1.1 Definition of Key Terms

A number of critical terms or constructs appear throughout this dissertation, and are thus defined here at the start. I consider "scaffolding", which is more fully reviewed in chapter 2, to be a form of contingent and transient support, offered intentionally to a learner, by a more knowledgeable other in the form of interpersonal activity or an intentionally designed tool, for the express purpose of enabling a level of performance or understanding not otherwise within reach of that learner.

"Technology tools" are any type of computer hardware/software combination provided to learners to enable or enhance a learning performance. Examples would include calculators, handhelds (iPhones), personal computers, etc.

An "Interactive Learning Environment" (ILE) is any learning environment where interactive computer technology is integrated into larger ongoing instruction. For example, a computer room where students go to use rote practice programs is *not* an ILE. The solitary, stand-alone, remedial, one-size-fits-all nature of this sort of use falls short. An English classroom where students use the Internet to research original Shakespeare manuscripts and create a new play using a word processor, and share their plays using a collaboration tool (e.g., wiki) *is* an ILE.

"Human Computer Interaction" (HCI) refers to the study of human use of computer hardware and software interfaces and more broadly to the community devoted to such study. The HCI community is inclusive of, but not limited to: software design and computer science, educators who need and use technology tools, and cognitive and educational psychologists who develop theories and pedagogy related to implementing ILE's.

"Articulation" is the process whereby a learner makes their thinking visible in words. Similar to the ideas of explanation or elaboration, it forces the learner to take a persistent public position in terms of their knowledge, which can then be shared, assessed, and modified. For the purposes of this study, which also looks at verbal dialog, articulation is focused expressly on the use of written words to create an external representation of concepts or ideas as they are understood by the learner.

"Articulation Scaffold" is a scaffold designed to encourage or enable learner articulation, as defined above. Such a scaffold would at a minimum provide space for a learner or learners to specify the details of their understanding. For the purposes of this study, articulation scaffolds enable written articulation only.

Research Questions and Hypotheses

In exploring issues related to these constructs, this study asks two major research questions, each with sub-questions, in regards to the use of scaffolds in a dynamic modeling tool. First I ask – What types of student and teacher interactions occur with and around scaffolds in a dynamic modeling tool, and do they change over time? Specifically I ask:

- Do tool scaffolds for articulation lead to learners articulating their thinking in writing, does this change over time, and if so, how?
- While using tool scaffolds for articulation, what is the quality of discussions learners engage in, does this change over time, and if so, how?

The hypothesis for this question is that learners will successfully use scaffolds in the tool to articulate their thinking and they will have cognitively valuable discussion while doing so. Secondly I ask – What is the quality of learners' models and how does that quality change over time? Specifically I ask:

- What is the quality of learner pairs' final model artifacts?
- Does the quality of these models, in terms of content or complexity, change over time, and if so, how?
- What is the pattern between learner interactions with and around scaffolds and the quality of their final model?

The hypothesis for this question is that learners will create models of varying, but generally high, quality and that those pairs making better use of the scaffolds will create higher quality models, and that the quality of models created will increase over time.

1.2 Overview of Study

In this study, pairs of middle school science students are recorded (computer screen and voices) as they use a scaffolded dynamic modeling tool (Model-It) to create models during four curriculum modules over a one year time frame that covers parts of two grades. This recording provides a rich record of how the pair used the software and their interactions while engaged in the task of model creation. Additionally, their final model artifacts are analyzed. In general two learner pairs from two or more classrooms were designated as target students, and the classrooms were observed during four curriculum exposures. A total of 20 cases are analyzed. A case consists of two to five process video tapes (daily recording of student audio and computer screen video) and the final model artifact resulting from those days of work. Process video refers to a method for recording computer screen activities of students. The video from the computer screen is patched to a VCR while the audio track for that video is patched from microphones worn by the students. The resultant videotape shows all screen activity combined with all learner discussion. Process video tapes are digitized and transcribed, and the transcriptions are then enhanced by entering in descriptions of screen gestures (if any) that relate to transcript discussion. Transcripts are then entered into a qualitative analysis software (NVIVO) and coded for uses of scaffolds provided by the tool or teacher, as well as the nature of the dialog or screen activity during and immediately after the scaffold use. Model files are scored using a rubric based on prior research in evaluating concept maps and models, which addresses Accuracy, Completeness, and Function. Models for each case are described in a summary text document and coded for score in the database. The database can then be queried in almost infinite detail to develop warrants to answer the research questions.

1.3 Potential Contributions

As is clarified in chapter 2, this study can contribute in various ways to the study of scaffolding, the field of Human Computer Interaction (HCI), and ILE design. Scaffolds for articulation are mentioned in almost all scaffolding frameworks and taxonomies, and calls have been made (e.g., Quintana et al., 2004) for explicit testing of a scaffolding framework, of which articulation and reflection is a major part. Interpersonal scaffolding (the scaffolding that occurs around the tool) has received increasing attention in the design of ILE's (e.g., distributed scaffolding of Putembakar and Kolodner, 2005 and the synergistic scaffolds of Tabak, 2004). The interplay between scaffolds provided by tools vs. other aspects of the learning environment is a topic of importance in ongoing research, and this study's holistic examination of scaffolds *in and around* a software tool may contribute to our understanding of synergistic scaffolds. A detailed review of learner model artifacts can contribute to the understanding of how to evaluate dynamic models and similar concept maps, and how learners' modeling efforts change over time. Also, the discovery of any relationship between scaffold use and the quality of the final model artifact would provide valuable evidence for the importance of scaffolding as a design concept for ILEs.

1.4 Dissertation Structure

In chapter two, I review the literature on scaffolding, covering the evolution of the term and the current debates about its definition and application. Groundwork is laid for the coding schemes proposed in chapter 3. Additionally, I discuss the value of modeling in science education and the design of a scaffolded dynamic modeling tool, known as Model-It. Past research involving Model-It is reviewed, along with a recently proposed set of scaffolding design guidelines and the theoretical rationale for scaffolds designed into Model-It. In chapter three, I describe the context and goals of the study, as well as review the data sources and collection methods. A detailed explanation of data reduction and analysis methods is made, justified to the literature. Comprehensive coding schemes for scaffolds, articulation, dialog, and model evaluation are put forth. Chapter four presents my findings, organized around my research questions. Chapter five discusses the implications of my findings, makes clear the value of this study in relation to the previous work described in chapters two and three, and suggests possible future directions for further studies.

CHAPTER II

Literature Review

2.1 Introduction

In this chapter I discuss the origins of, and review current disagreements about, the term scaffolding. I review the use of scaffolding both as a term and a technique in the areas of teaching and tutoring, as well as the design of Interactive Learning Environments (ILE's). Sources of scaffolding include interpersonal scaffolding from human agents, and tool scaffolding from non-human agents or artifacts. A comparison of multiple conceptions of tool scaffolding reveals common themes. I discuss the value of student verbal dialog and written articulation for learners constructing shared understanding. I review literature on the educational value of model building and the design of, and past research on, a scaffolded dynamic modeling tool called Model-It. Next, I investigate the theoretical rationale for Model-It scaffolds, specifically for articulation, which is one of the common themes of tool scaffolding. Finally, I summarize and point out where this research fits into and can contribute to the existing literature.

This study examines three classes of middle school science learners as they use a scaffolded dynamic modeling tool in several project-based curriculum units. They create models in pairs, as a learning task to integrate and make visible their understanding of complex scientific phenomena like stream ecosystems. Videotapes of screen activity and learner discussion are used to create transcripts which are analyzed, along with the final model artifacts. Learner use of scaffolds, their discussions while using scaffolds, and the quality of their final models are investigated.

The potential contributions of this study are in the field of Human Computer Interaction (HCI), and ILE design. Within the broad HCI field, the Learner Centered Design (LCD) community would be the area most interested in investigations of various types of scaffolds in and around software tools. Articulation scaffolds are those which assist learners in using text to make their understanding explicit. Articulation is mentioned in the majority scaffolding frameworks and taxonomies, and Quintana and colleagues (2004) call for explicit testing of their scaffolding framework, of which articulation and reflection is a major part. A detailed, longitudinal examination of how often and how well learners use articulation scaffolds can inform future refinements to such frameworks. Further, the role of interpersonal scaffolding, which is well studied outside of technological contexts, receives increasing attention in the design of ILE's (e.g., the differentiated scaffolds of Tabak, 2004). The careful study of the interplay between scaffolds provided by tools vs. other aspects of the learning environment can enrich the emerging theory of distributed scaffolding. The detailed review of learner model artifacts will allow validation of an evaluation rubric that includes the best aspects of prior (static) Concept Map evaluations but also accounts for the dynamic functions of computer models and newer "Cyclical Concept Maps" (Derbentseva, et al., 2007). Finally, any relationship discovered between scaffold use and the quality of the final model artifact would provide valuable evidence for the importance of scaffolding as a design concept for ILEs.

2.2 Scaffolding

Scaffolding has its origins in the literature of psychology, as discussed below, but there are evolving definitions and design guidelines. Sources of scaffolding include teachers, and, depending on one's perspective, other sources such as: peers, tools, computer software, etc. Scaffolding provided by non-human agents, particularly computer tools, provides a rich area for discussion since they are designed by humans but cannot fully replicate what a teacher can provide. New theoretical constructs of distributed and synergistic scaffolding address the interaction between various sources of scaffolding. Four central themes related to scaffolding from human and non-human agents are discussed below.

2.2.1 Origins of the Scaffolding Term

While the term scaffolding is most strongly associated with Vygotsky, the term is generally acknowledged to have first been used by Wood and colleagues in their 1976 article on the role of tutoring in problem solving. They planted the seed from which all subsequent definitions have grown, saying tutoring "involves a kind of scaffolding' process that enables a child or novice to solve a problem, carry out a task, or achieve a goal which would be beyond his unassisted efforts" (p. 90). Years later, as the theories of Vygostky were more widely read in translation and social constructivist theories rose in popularity, his theories of the Zone of Proximal Development (ZPD) became closely associated with the idea of scaffolding. The ZPD is that area between what the learner can do on their own, and what they can do with the assistance of a more knowledgeable other. Scaffolding is now often simply referred to as "support provided within the ZPD".

Though it is a conceptually powerful term, scaffolding is subject to significant debate and misunderstanding. As Palincsar (1998) notes, scaffolding as a construct tends to be "most used yet least understood" precisely because of its great descriptive power (p. 390). As a construct it is not yet fully articulated, nor rigorously empirically verified, but that has not stopped it from becoming a prominent construct in several fields of research, including: learning disabilities, tutoring, and human computer interaction (HCI).

Each of these fields tends to have its own set of assumptions about what scaffolding means and requires. Palincsar (1998) and Stone (1998) debate the utility of the scaffolding construct for the field of disabilities research, and whether broad and atheoretical use renders the term inert. A key issue of contention is whether scaffolding is a purely interpersonal activity, or one that can involve other agents or artifacts. Yet the original definition (Wood, Bruner & Ross, 1976) still serves to bound the discussion, and provides a trail to follow when reviewing the literature.

2.2.2 Scaffolding as Interpersonal Activity

From the earliest conceptions of scaffolding, it has been an interpersonal activity. A long tradition in ecological psychology (Bronfenbrenner, 1979) has maintained the emphasis on interpersonal dialog and the important presumption (discussed again later) that the scaffolding dyad contains a more knowledgeable other who is intentionally providing guidance. As Wood originally noted, "The effective tutor must have at least two theoretical models to which he must attend. One is a theory of the task or problem and how it may be completed. The other is a theory of the performance characteristics of his tutee" (p. 97). Under these assumptions, the provision of scaffolding is no small or simple matter. It is presumed that this other has: knowledge/mastery of the task and sub-parts of the task, knowledge of the learner, knowledge of where the learner is in relation to the task and its sub-parts, knowledge of where learners in general have trouble with the task, and, finally, some idea of where this particular learner might have trouble with the task and/or any sub-part of the task.

It is from research on tutoring and disabilities that a number of empirically based definitions and taxonomies of scaffolding have been proposed. The proposed taxonomies or behavior categories regarding scaffolding behaviors (Hogan & Pressley, 1997; Lepper, et al., 1997; Roehler & Cantlon, 1997; Wood et al., 1976) can be aggregated around five concepts. These concepts are: Changing the Task, Modeling Behavior, Enhancing Motivation, Sharing Knowledge, and Sharing Metacognition.

Changing the task relates to any effort by the teacher to make the path of the learner easier, or to help the learner maintain progress. By selection of and/or modification to the task or its sub-elements the teacher changes the task in some temporary and contingent way to allow the learner to see it through to successful conclusion. This includes both the selection and setup of the task (Hogan, 2000) and the ongoing supervision of the task in progress (e.g., Wood and colleagues' "direction maintenance"). Additionally it includes the P from Lepper and colleagues' (1997) INSPIRE framework, Progressive routines to gradually introduce material and identify difficulties.

The *modeling behavior* theme relates to teachers providing exemplars of desired performance for learners to emulate. They can do this by their own behavior, or by inviting and supporting learner behaviors. This can include overt demonstrations (Wood, et al., 1976) of a procedure, or more subtle modeling of behaviors, as well as explicit invitations to practice a skill or task while the teacher observes (Roehler & Cantlon, 1997).

The *enhancing motivation* theme relates to a wide range of activities required to manage learners and their ego needs. By recruiting interest (Wood, et al., 1976) and pre-engaging the learners in a shared goal before the task (Hogan, 2000), the teacher can then carefully attend to learner efficacy and affect during the task (e.g., Lepper and colleagues' Nurturant, Intelligent, and Encouraging which embody praise or constructive criticism). This motivational concept also captures how the teacher can maintain interest and reduce possible frustration and risk (Hogan, 2000; Wood, et al., 1976), ensuring the task is completed successfully.

The *sharing knowledge* theme relates to broader verbal interactions designed to engage learner thinking. By having Intelligence (content mastery) and using Socratic methods (Lepper, et al., 1997) and careful matching of content and metaphor to observed learner needs the teacher can offer explanations (Roehler & Cantlon, 1997), tailored assistance (Hogan, et al., 2000) and clarify to bring the learner to a shared understanding.

The *sharing metacognition* theme relates to a more specific sharing of knowledge about thinking as opposed to the specific content focus above. By explicitly attending to metacognition (e.g., helping learners Reflect (Lepper, et al., 1997), diagnosing their needs (Hogan, et al., 2000), or making explanations/generalizations (Hogan, et al., 2000), the teacher can enhance learning and performance, now and in the future. This is distinct from dealing with a specific concept or misconception as would occur in the "Sharing Knowledge" theme.

These five themes were found in the previous research on scaffolding behaviors of teachers or tutors. The environments for such research did not involve technology, and often involved only a teacher and learner dyad. However, it is reasonable to expect to see these behaviors when we examine teacher and peer behaviors in the context of an ILE.

2.2.3 Scaffolding in Contemporary Learning Theories

A number of contemporary theories address scaffolding either explicitly or implicitly. Though Stone (1998) mentions the Cognitive Apprenticeship of Rogoff (1990) and Collins and colleagues (1989), and the Legitimate Peripheral Participation of Lave and Wenger (1991) as alternatives or reactions to an over-emphasis on the idea of scaffolding, these theories still contain similar concepts. Teachers, through their classroom behavior, can make a learner into a sort of "cultural apprentice" (Rogoff, 1990). In the model for cognitive apprenticeship of Collins and colleagues (1989) scaffolding is explicitly listed as one of six desirable methods to use in ideal learning environments. Discussions of apprenticeship and tutoring focus on the master/apprentice or teacher/learner dyad, and inevitably on the way the master provides tailored assistance in the areas the apprentice is able to grow, then gradually removing the assistance and moving on to the next, more challenging step. Within all of these situations or definitions, there is a constant tension between ensuring the learner is pushed, but not beyond their abilities. The concept of "distributed cognition", popularized but not originated by Salomon (1996), shows up in numerous conceptions of learners interacting with agents and artifacts (e.g., Computer Supported Collaborative Learning, CSCL). Scaffolding is also a theme in the Project Based Science curricula of Krajcik and colleagues (1998), which invokes multiple sources of scaffolding (to include technology, discussed later) that can support student efforts at complex inquiry tasks.

2.2.4 Unique Concerns for Scaffolding from Peers

An important consideration when discussing applications of scaffolding is what counts as a knowledgeable other? Obviously an adult with greater knowledge or experience can fill that role, but the answer is less clear for peers. Piaget (1959) felt that peers were a better source for learning discussions, particularly the sort of discussion that would lead to cognitive restructuring, because the unequal power of the adult or mentor will tend to cause the child or learner to simply accept the answer without question. In Piagetian constructivist theories, the goal is to cause disequilibrium so as to motivate the learner to modify their schema and restore equilibrium. As evidence, for example, non-conserving children were shown to attain conservation 80% of the time when paired with a conserving peer for discussion, as opposed to attaining conservation only 50% of the time when paired with an adult. Recent work in peer tutoring showed that peers benefit from tutoring activity because it clarifies their own understanding (Roscoe & Chi, 2007). The issue of this desirable "cognitive conflict " will be discussed again later. Vygotsky (1976) similarly stated a “more capable peer” could be effective helper. The Vygotskyian view emphasized the interaction between peers as an important part of development, as opposed to the more individualized development conception of Piaget. However both theorists acknowledged the value of peers in learning, even if they did not discuss scaffolding behaviors directly.

2.2.5 Scaffolding One-to-One Versus One-to-Many

A final key concern for discussions of interpersonal scaffolding is the fact that one-on-one tutoring, while historically valid, desirable, and powerful (Bloom, 1984), is rare. Far more common is the standard classroom situation where one teacher supervises the education of dozens of learners. Can the original conception of scaffolding in the learner/other dyad accommodate multiple learners? Contemporary educational theorists (i.e., grounded in classrooms or apprenticeship situations) seem to think so. Palincsar (1998) noted “it is helpful to recall that ZPD’s include not only people but artifacts, and that ZPD’s are embedded in activities and contexts”(p. 371). Brown and colleagues (1993) first described “overlapping zones of proximal development” (p. 194) in the classrooms they were working in. Then in their book chapter on communities of learners Brown and Campione (1994) described classrooms with multiple ZPD’s and emphasized that ZPD’s could contain non-human agents. Hogan and Pressley (1997) offer a thorough review of instructional practices for scaffolding student learning in classrooms, and explicitly discuss strategies for scaffolding while working with whole class discussion. Thus scaffolding is not limited to a one-on-one dyad, but can, at least for some theorists, can be provided by one to many. While the strictest possible interpretation of scaffolding, or focus on the original tutoring context from whence it came, can seem to exclude one-to-many classroom settings, such a focus may be unnecessarily tight.

2.3 Scaffolding from Non-human Agents

Many claims have been made over the last two decades as to what non-human agents or artifacts can provide and whether they can be considered a form of scaffolding. In early research that presaged current work on prompts, Scardamalia and Bereiter (1985) used "procedural facilitators" which were notecards with sentence prompts to enhance online argumentation. Similarly, written guides and checklists were used to scaffold learners as they mastered the techniques of reciprocal teaching (Palincsar & Brown, 1984). Collins, Brown and Newman (1989) postulated that scaffolding could take the form of reminders or help, and then also included the concept of physical supports (e.g., cue cards for a debate team or ski-tethers for skiers learning to snowplow). Rogoff (1990) also provides a very inclusive definition of scaffolding as supportive situations created by adults to help children stretch present knowledge and skills to higher level of competence. Several researchers have addressed curriculum as scaffolding. In the case of Kolodner (2003) a "launcher unit" was analyzed as a scaffold and McNeill (2006) addressed scaffolds intentionally designed into a curriculum unit. Azevedo (2005) distinguishes between "embedded static scaffolds" (as part of software or a tool) and "adaptive human scaffolds" from peers and teachers (p. 201). So, if scaffolding is not restricted to a learner/teacher dyad, and if various cultural tools and physical artifacts can be scaffolds, then perhaps ever more powerful computer software can also play a role in scaffolding learners.

2.3.1 Scaffolding from Software Agents

If scaffolding can be provided "one-to-many" then computer software (which is by its nature supporting many learners) can exhibit scaffolding. But is it scaffolding if one half of the learning dyad is a machine, a computer? A computer that, the strongest critics would allege, can have no personal understanding of a given learner, their past history, their classroom tendencies or unique needs. This concern can be addressed by the possibilities of artificial intelligence (AI), the "voice of the programmer", and distributed scaffolding, each of which will be discussed below.

2.3.1.1 The voice of the programmer.

One solution is to acknowledge the limitations of present software design and processor capacity, but to extrapolate forward and argue for what could be possible, someday. State of the art research in Artificial Intelligence will not be reviewed here, but suffice to say the solution may be close but it is not here now. Multifaceted reforms such as Problem Based Learning (Krajcik & Blumenfeld, 2002) call for use of scaffolding from learning technologies as one of five key concepts. Even a decade or more ago, Intelligent Tutoring Systems (in geometry, for example) offered up to 8 subtly graded levels of help to learners that was performance contingent (Anderson, 1995). Even in domains that are not as strictly bounded or organized as geometry, highly organized systems of prompts can (as part of a larger scaffolded environment) enable success in complex inquiry projects (Davis, 2000, 2003). With increases in storage and processor capacity, we may not be far away from a "brute force" approximation of a "guide at the side" in subjects like science or even history. However, current technological limitations prevent software tools from fully emulating scaffolds from human agents (Azevedo, 2005).

Leaving aside this speculative AI option, we can take up the idea that for now, intentionality will come from how the software is designed and implemented. In a classroom in Russia, in a very early study of a rudimentary software program (somewhat similar to the modern LOGO program) a group of researchers (Griffin, et al., 1993) noticed a peculiar phenomenon. As they field tested their software and studied its use in classrooms, it became clear that many decisions the programmers were making were having effects on the learning context. They noted that the computer (and thus the software) was not a static object, and had in fact joined the teacher and learner's dialog to create what they called a "polilogue". They coined the term "voice of the programmer" (Griffin, et al., 1993), noting "programmers had to be treated as "hidden" members of the communicative interactions, with distant but powerful "voices" (p. 126). In a similar vein, Puntambekar and Kolodner (2005) apply Woods, et al.'s, (1976) previously discussed requirement for theory of task and theory of tutee to scaffold design. They argue that in employing software as a source of scaffolding it is critical to know the common aspects of intended learners so situation and software can be tailored for the multiple ZPDs that will be encountered.

Finally, in a related theory, Learner Centered Design (LCD) makes the equivalent of "voice of the programmer" assumptions in its theory of design (Quintana et al., 2001). LCD argues that learners (in software design terms) are fundamentally different from users, and require unique support. That support is based on an analysis of the needs of that learner population in relation to the learning task, which again invokes Wood and colleague's (1976) theory of task and theory of tutee, if in a somewhat broader sense. The "voice of the programmer" helps to address concerns about scaffolding in technology tools, and is, in the end, quite focused (perhaps unrealistically) on the capacities of the software design to embody all the required burden of the "more knowledgeable other".

2.3.1.2 Distributed scaffolding.

If computers will be part of the educational context for the foreseeable future, and if they cannot provide a single unified source for the scaffolding that learners require, perhaps a broader conception of scaffolding is warranted. Theorizing on his early research on the impact of computers on classroom interactions, Crook (1999) noted "The management and evaluation of computer experiences would benefit from focusing on the broader context of classroom discourse in which such experiences may be situated"(p. 60). A decade earlier, computer based instruction (CBI) tutoring pioneer Anderson (1989), after researching tools that were designed from the start to stand alone, acknowledged the value of teachers for providing assistance to learners when they fell outside of the scope of pre-designed help, as well as the value of classroom peers for sharing knowledge of how to use the tutor software, and motivate each other by comparing progress. When it comes to employing technological tools in educational settings, context matters, and the contemporary concept of distributed scaffolding continues this line of thinking.

The concept of distributed scaffolding was first advanced by Puntambekar and Kolodner (1998), in their discussion of their research on middle school science students learning science by design, as a new way of thinking about scaffolds in ILE's in context. Acknowledging the theme of the importance of classroom context found in other lines of research, Puntambekar and Kolodner (2005) found that no one source can provide all needed scaffolding. Instead they emphasized the importance of various social and material supports that are brought together to create an organized system of tools and agents. This is similar to the call from Salomon (1996) to recognize the importance of "cognitive tools" and interaction between various classroom elements (e.g., teacher, students, software, etc.). This more holistic and social constructivist view of scaffolding productively avoids specific arguments about what software can, can't and can't yet do, by considering the aggregate context of human and material supports arranged to support learners in complex disciplines, such as science.

Distributed scaffolding was reviewed and extended by the work of Tabak (2004), in her proposal for the concept of Synergistic scaffolding. Bringing forward the ideas of Differentiated scaffolding (multiple unique situations of one scaffold to one need) and Redundant scaffolding (several scaffolds to the same need, not necessarily concomitant) from previous research, including that of Kolodner (2003), she added a third type called Synergistic scaffolding (multiple scaffolds interacting to support one or more needs). As the concept of synergistic scaffolding gains empirical support, tool designers will not have to attempt to have the tool do it all, but rather will have to account explicitly for the larger learning context and apportion the scaffolds required for the task across those contextual resources that can best provide it.

2.3.2 Common Themes in Conceptualizations of Tool Scaffolding

From the earliest studies of the use of computers in classrooms, attention has been paid to not just their ability to provide rote drills, but also to encourage and support higher order processes. Salomon and colleagues (1989), in research that predates some of the more recent debates on scaffolding, found, using what we would now call metacognitive prompts as scaffolds for reading comprehension, that “computers can serve as tools that provide guidance in a child’s ZPD and can thus facilitate competence development” and that such use can “leave a desired cognitive residue“ (p. 626). It is generally acknowledged that simple rote learning (or “drill and kill”) software is one of the least useful ways to employ computers in education, and while it may have its place in specific situations, over-reliance on this sort of software can actually degrade learner abilities (Weglinsky, 1998). Reviews of educational software research also refer generally to “technologies used to scaffold thinking” (Bransford, Brown, & Cocking, 1999) and by this they mean the sort of higher order cognitive processes that would be going on in a tutoring dyad.

As the Human Computer Interaction (HCI) field discusses scaffolding in software, definitions vary a bit in how Wood and colleague’s (1976) original conception is applied. However there is clear agreement that scaffolding involves supporting a learner in accomplishing a goal that would not otherwise be possible (Brush & Saye, 2001; Fretz et al., 2002; Guzdial, 1994; Jackson, et al., 1999; Krajcik et al., 1998; Quintana, 2001). The primary consideration in HCI discussions of software and environment design is the scaffolds in tools, however in some cases a tool itself could be a scaffold. In cases where the tool is small and focused, it might be a scaffold in the manner of the artifacts discussed earlier. Quintana (2004) argues that well designed tools "provide scaffolding by transforming tasks in ways that lead to greater success and opportunities to learn" (p. 341). In a more complex example, the Scaffolded Integrated Tool Suites (SITE) of Quintana (2001) contain multiple tools (each with scaffolds), as well as additional scaffolds to integrate the use of the various tools in some larger sense (e.g., conducting a scientific investigation).

Although current conceptualizations of scaffolding in the HCI field can be traced back to early definitions in other fields, certain aspects differ. Technological tools can be seen generically as an "intellectual partner" (Salomon, Perkins, & Globerson, 1991) which assumes some of the cognitive/information processing load. When definitions become more specific, they tend to reveal the engineering and design aspects of the HCI community. Mentions of motivation and feedback on performance quality are rare, because these things are, to this point, difficult for technological tools to do at all, much less do well. Also, fading, a critical concept that will be discussed later, is often overlooked.

Researchers in the HCI community have conceptualized scaffolding in a variety of ways. Table 2.1 below lays out the specifics of six published conceptualizations of scaffolding for technological tools. I have grouped the items found in each conceptualization of scaffolding around five themes. There are two themes (articulation/reflection and dynamic diagnosis/coaching) that appear in nearly all the articles, and will be discussed in detail in the paragraphs that follow.

Under the "Articulation & Reflection" top row in Table 2.1, Guzdial (1995) focused on the need to elicit articulation from learners to identify programming misconceptions and share ideas. Metcalf (1999) similarly described "reflective" scaffolds as one of three types, whose purpose was to encourage learners to articulate their thinking and make predictions while engaging in science learning. Linn and colleagues (2004) noted providing social supports for peer discussion and encouraging reflection

	Guzdial, 1995	Metcalf, 1999	Linn, et al., 2004	Brush, 2001	Quintana, et al., 2004	Puntambeker & Kolodner, 2005
ARTICULATION & REFLECTION	X	X	X		X	X
DYNAMIC DIAGNOSIS & COACHING	X	X		X	X	X
MODELING & STRUCTURING THE TASK	X	X	X	X	X	
METACOGNITIVE ASSISTANCE				X	X	X
FADING		X				X

Table 2.1: Summary of Scaffolding Concepts from various ILE research.

were aspects of the learning environment worth scaffolding. In a set of five guidelines for tool scaffolds supporting science inquiry, Quintana and colleagues (2004) emphasized articulation and reflection as a specific learner behavior that software could support. Specifically mentioned were simple reminders and other support for learners to articulate their ideas and engage in sensemaking. Puntambekar and Kolodner (2005) noted the value of dialog and interaction between teachers and peers while using technology as one of five areas tool scaffolds could support. They proposed the concept of distributed scaffolding, a system of tools and agents, to specifically emphasize how no one agent should or could best provide all scaffolding. Thus we see that nearly all these conceptualizations of scaffolding note the importance of supporting articulation and reflection.

The second row of Table 2.1 deals with the slightly more complex idea of dynamic diagnosis and coaching, where scaffolding can be provided to learners as they attempt to master new concepts or tasks. Again five of the six conceptualizations address this issue, such as Guzdial's (1995) early emphasis on providing hints, tips and encouragement to learners, or Metcalf's (1999) Supportive scaffolds that explains subtasks and provides coaching. Similarly, Puntambekar & Kolodner's (2005) idea of Dynamic

Adaptive Support addresses how to gather information about learner weaknesses or misconception and then provide calibrated help. Brush (2001) and Quintana and colleagues (2004) discuss making strategies explicit and using ongoing diagnosis to provide calibrated help.

Modeling and Structuring the task is also addressed by five of the six researchers. Guzdial (1995) speaks of communicating the process via demonstrations, and Metcalf's (1999) Intrinsic scaffolds change the nature of the task. Linn and colleagues (2004) discuss making science accessible by modeling behaviors and making thinking visible by showing how experts think and act to solve problems. Brush (2001) discuss Procedural scaffolds that help with using tools, and Conceptual scaffolds that make problem solving and task steps clear. Quintana and colleagues (2004) similarly address Process Management to provide structure and expert guidance to make tool use easier.

Further down the chart, each row captures a theme with less broad support. Metacognitive Assistance is addressed by Brush, (2001) as learning about processes and metacognition, by Quintana and colleagues (2004) as sensemaking and the support of inquiry, and by Puntambekar & Kolodner (2005) as building shared understanding of a common goal. For the final row that deals with the idea of fading, or gradually removing support, only Puntambekar directly accounts for "fading and transfer of responsibility" in her conceptualization (while the other authors all directly or indirectly mention the idea of fading, only this one specifically accounts for it).

As shown in Table 2.1 certain themes occur across nearly all lists or taxonomies from research on scaffolding in ILE's. In particular, the abilities of software to encourage articulation and reflection are emphasized in nearly every system. Additionally, the themes of modeling/structuring the task and of dynamic diagnosis/coaching are also broadly applied. These are areas that do not require the software part of the ILE to exhibit anything like artificial intelligence. Encouraging articulation and reflection can be accomplished with aspects of design and by using prompts. In the areas of metacognitive assistance, helping the learners think about their thinking, there is less emphasis because it requires both artificial intelligence and a unique understanding of what the learner knows. Similarly, the issue of fading is only dealt with explicitly in one system. These are aspects of software design and interface that pose the most difficulty for programmers and designers, so it is not unusual that they have yet to be fully explored and articulated. Of central concern for this argument is that articulation and reflection are almost universally acknowledged as something that software can and should scaffold.

2.3.2.1 Fading.

One of the key point of contention when discussing scaffolding in technological tools is the idea of fading, the gradual reduction of support that has been part of almost every conception of interpersonal scaffolding. In Wood and colleague's original study (1976, p. 96), they refer to the tutorial function "withering away" as skill and age increase. In the cognitive apprenticeship work of Collins and colleagues, a more knowledgeable or capable other reduces participation in the task so the learner can move toward mastering the whole skill or process. It is not just interpersonal scaffolding that can fade; the work of Lee (2003) and McNeill (2006) showed that scaffolds designed into curriculum could be productively faded. In the area of ILE's, Guzdial's (1995) early work on software realized scaffolding provided multiple options for fading multiple scaffolds, and Jackson (1999) developed an entire system of Guided Learner Adaptable Scaffolding (GLAS). However, Lee and Songer (2004) found that fading of content specific written scaffolds too quickly reduced the quality of student explanations. Fading is difficult to manage even in interpersonal situations and even more so for software agents. Tabak (2004) continues to address the issue by making fading a key, at least implicitly, to her definition of scaffolds as mediational means that are not expected to persist into the future.

The fading of scaffolds can be problematic in ILE's. It is partially bound up in the argument of intentionality (discussed below). In work studying the fading of scaffolds in ILE's Guzdial (1995) and Jackson (1999) found benefits for both letting learners control fading (adaptable scaffolds) and letting the software control fading based on learner actions (adaptive scaffolds). The idea of how best to manage fading of scaffolds in ILE's remains elusive, and the issue is not consistently addressed in research. Pea (2004) makes a distinction between "scaffolds for performance" (which persist in the environment, like a miter box on a saw) and "scaffolds with fading" (p. 438), and then re-labels scaffolds for performance as being part of his concept of distributed intelligence, essentially denying they are scaffolds at all. If we agree that true scaffolds must fade or be capable of fading, Luchini (2003) adds one final wrinkle, noting that learners can, in effect, fade certain scaffolds themselves, by simply ignoring or not using them after they have learned to do the supported task without the help. Fading remains a key concern in studying scaffolding, and it is tied up in the argument of what, exactly, a software program can do in terms of fully replicating tutoring functions.

2.4 My Definition

For the purposes of this dissertation, and in light of the foregoing discussion, I offer the following definition of scaffolding:

> Scaffolding is a form of contingent and transient support, offered intentionally to a learner, by a *more knowledgeable other* in the form of interpersonal activity or an intentionally designed tool, for the express purpose of enabling a level of performance or understanding not otherwise within

reach of that learner.

Scaffolding within software tools or ILE's (hereafter referred to as "tool scaffolds"), are thus design aspects or sub-components of such software that meets the above definition. For this definition, it is not required that the tool scaffold actually fade, only that it could be faded, so as to distinguish it from supports that would help any and all users engaging in similar activities. Scaffolding from human agents such as a teacher, mentor, etc. (hereafter referred to as "interpersonal scaffolding") would therefore be any combination of communicative efforts (speech, writing, drawing, gestures) undertaken in an interaction with one or more learners. This definition does not explicitly exclude peers, but as will be discussed later, true scaffolding among peers, in light of the above definition, is rather rare.

Now that scaffolding has been reviewed and a working definition laid out, I will next briefly review the literature on the value of modeling in science education to set the stage for a discussion of Model-It, the software used in this study.

2.5 Modeling

Making and using models in science education is a priority for policymakers, researchers, and teachers alike. Modeling is one of the benchmarks for scientific literacy (AAAS, 1993), and is specifically discussed in national science education and research documents (AAAS, 1990; NRC 1996; NRC 2000). Science education reform efforts highlight the importance of model creation and use (Clement, 2000; Gilbert, Boulter & Rutherford, 1998; Gobert & Buckley, 2000). The ever increasing prevalence of computers in classrooms and learner facility with using them makes visualizations like models a vital, if incompletely understood, aspect of science classrooms (Gobert, 2007), and makes them an "integral part of scientific literacy" (p 9).

2.5.1 Modeling in Classrooms

Modeling provides a number of cognitive benefits to the learner engaged in modeling practices (Stratford, 1996; Zhang, 2003; Sins, et al., 2005). A model is a representation that abstracts and simplifies a system by focusing on key features to explain and predict scientific phenomena (Schwarz, et al., 2009). Modeling practices have been described as representational practices or cognitive strategies, which are engaged in during the process of model creation. In prior work focusing on Model-It (Fretz et al., 2001; Zhang, 2003), modeling practices were identified to include: Planning, Analyzing, Synthesizing, Evaluating, Reflecting & Monitoring, Publicizing & Communicating. Zhang (2003) found that Model-It scaffolds allowed middle school students the opportunity to demonstrate a variety of desirable modeling practices, and that student modeling practices became more integrated, meaningful, and purposeful over time.

A model can serve as an important sensemaking tool as students construct scientific knowledge (Magnani et al., 1999, Krajcik & Blumenfeld, 2002). For example, models can be used in science classrooms to highlight concepts and structures of a complex system (Gobert & Discenna, 1997; Gobert, 2007). Models serve as a representation (often simplified) of a system, and they help focus attention on system components like objects, events, or processes (Gilbert, Boulter, & Rutherford, 1998; Ingham & Gilbert, 1991). Once created, they can be modified and manipulated to investigate the effects of changing system components. Model artifacts allow teachers to assess for higher level cognitive outcomes (Krajcik & Blumenfeld, 2002). Models also help students actively make links between the real world of objects, events, and relationships, and the world of theory and model, which facilitates their understanding (Tiberghien, 2000; Gobert, 2003; Schwarz, et al., 2009). As Wilensky and Reiman (2006) pithily note, “if you can’t build it, you don’t understand it” (p. 202).

2.5.2 Challenges in Model Creation

Model creation is the central task under observation in this study. The models in question serve as a summative assessment task in a project based science curriculum (more fully discussed in chapter 3). Modeling by students tends to involve a series of desirable activities referred to as modeling practices, such as: analysis, relational reasoning, synthesizing, and testing/debugging. These modeling practices are discussed and investigated by a number of researchers, including Stratford (1996) and Zhang (2003) and will be discussed in a subsequent review of Model-It research. Though the benefits of modeling are clear, the goal of making modeling available to all science learners remains challenging. Penner (2001) noted that "in contrast to mainstream practices of school science, accounts of the work of professional scientists paint a picture dominated by the building and testing of models" (p. 2).

One of the central values of modeling is the collaboration and struggle to create a shared understanding that it makes possible. As Penner (2001) stated, "the process of physical instantiation moves the model from the mind of the individual into a public forum where it can be discussed" (p. 2). This idea of collaboration and shared construction is very much in keeping with contemporary social constructivist theories, some of which were discussed previously. The key markers of such collaboration are what the learners discuss and what they articulate (or type) in their descriptions or justifications for their work. The concept of cognitive conflict and its value to the learning process will be fully discussed below, in the rationale for articulation scaffolds.

2.5.3 Computer Tools to Support Model Creation

As the use of computers in classrooms increases, so does the use of computer based technological tools to assist in the creation and use of models in science classrooms. While there is considerable research on computer simulations (which are essentially fixed pre-made models that can be manipulated), I will focus on the research on modeling software specifically. Making dynamic modeling accessible to secondary school learners has proven a challenging task. Gobert (2007), in discussing visual representations such as models, explicitly calls for scaffolding as vital to supporting students' knowledge acquisition while constructing models. In a review of modeling programs Penner (2001) found only three dynamic modeling programs used in the secondary school environment: STELLA (Mandinach, 1989; Reimann, et al., 2007), Model-It (Jackson, 1999), and Star Logo (Resnick, 1997) (considered along with its cousins Logo and Boxer). Of the three, STELLA and Model-It share the metaphor of connecting building blocks with relationship information, in a sort of highly advanced version of the concept map.

NetLogo (and updated version of StarLogo) (Wilensky, 1999; Wilensky & Reiman, 2006) takes a different approach by allowing a number of independent entities (or "turtles") to be programmed and then released into a bounded environment where their interactions can be observed. This is useful for modeling, say, ideas about how termites create large nests without centralized management or intelligence. NetLogo and its precursors are rooted in the "constructionist" philosophies of Papert (Harel & Papert, 1991), and emphasize the value of learning by making. Like the constructivist theories of Piaget and Vygotsky already discussed, constructionism agrees that knowledge is actively constructed by each learner through interaction with the world. Constructionism then extends this idea by emphasizing the need to project internal ideas into a shareable form which can be sharpened or improved. Of central concern is the change in learner beliefs and the dynamics of such changes. STELLA and Model-It share this desirable feature of creating a shared public representation of learner understanding that can be the subject of much discussion during construction and review. Between STELLA and Model-It, STELLA is by far the more complex tool. Used only in high school environments, research reports that learners would often take several class periods or a practice unit just to master the software function (Mandinach, 1989).

Model-It (Jackson, 1999) is the most heavily scaffolded of all these tools, takes a more learner centered approach and was designed to be accessible to even very young learners with minimal instruction. Model-It has been used successfully in several middle school and high school science contexts (Stratford, 1996; Jackson, 1999; Novak & Krajcik, 2006). It is this tool that is used in this study.

2.6 Model-It, a Scaffolded Dynamic Modeling Tool

Model-It is a learner centered, scaffolded, dynamic modeling tool, designed to make the benefits of modeling accessible to even very young (e.g., 6th grade) learners. Using the ideas of LCD, it provides a simple interface, a way to model qualitatively and still observe quantitative functions, and numerous scaffolds to support the learner. A full description of Model-It's design history can be found in the work of Metcalf (1999), but in summary this tool underwent intensive development through the 1990's to version 3.0 which is the version in this study. A detailed description of Model-It's interface and modes can be found in the work of Fretz and colleagues (2002), but a summary review follows.

Model-It uses three functional modes, Plan / Build / Test, to decompose the modeling task. Screen shots of these modes can be found in Figures 2.1, 2.2 and 2.3. While an expert can approach the complex modeling task without support, the interface of Model-It constrains the task and requires or encourages learners to proceed through steps in order or iteratively. They first create the 'building blocks' of their model (e.g., the Plan mode), and then they link those 'blocks' together with very specific details about how A affects B. (e.g., the Build mode) Once enough links are made, the function of the entire model can be tested (e.g., the Test mode).

In Plan mode (Figure 2.1), a learner can create Objects (such as a "factories" or “cars”) and then create Factors (which would be measureable aspects of those factors, such as “emissions” for factories or “exhaust” for cars). The Objects serve primarily as a way to organize and think about the Factors. In this study, Objects are referred to by their title, such as the Object “Factories”, and the subordinate Factors are referred to in combination with their parent Object, such as the Factor “Factories: The Amount of Emission”. In a model of water quality, for example, learners might want to address their stream primarily, and focus on some tests they did such as temperature, dissolved oxygen, and pH. Thus they would create an Object called “Stream” and a series of Factors like: “Stream:temperature” and “Stream: pH”. Within each dialog box for creating Objects and Factors there is an area for providing a definition/description of the Object or Factor. This is where learners should be articulating their definitions of the concepts they are working with.

Once a set of Factors has been created, they can be linked together with Relationships (Figure 2.2). This is accomplished in the program by dragging and arrow between two Factors and then specifying the details of the Relationship in a dialog box. The Relationship dialog box requires the learners to specify the direction and degree of the Relationship. So A might be related to B such that as A increases then B decreases, and that decrease is linear. There is also an articulation box with a partially filled in sentence that asks the learners to say why the Relationship is this way. So for a Relationship between the temperature of the stream and the dissolved oxygen level of the stream the Relationship would be: “As Stream:Temperature decreases Stream:Dissolved Oxygen decreases by about the same because colder water can carry less dissolved oxygen.” Both factors and relationships can be customized in various ways to be more quantitative and to reect directionality and rate of change. Using this basic premise, learners build up models that represent their understanding of complex systems (such as weather) or phenomena (such as decomposition).

Once built up, the models can be run and tested (independent factors can be

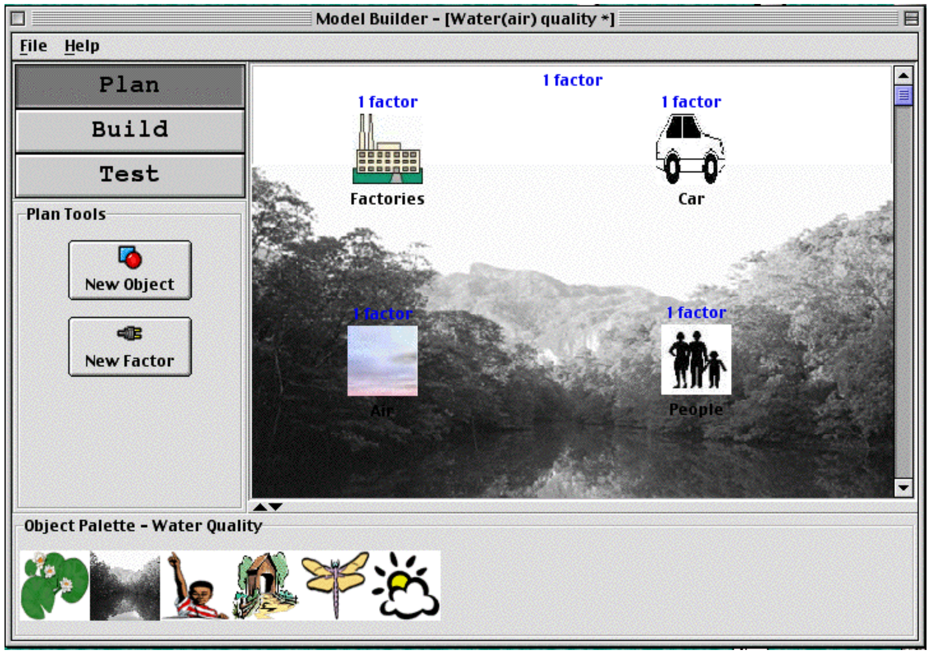

Figure 2.1: Plan Mode (custom background graphic, icons for objects e.g., Cars or Factories)

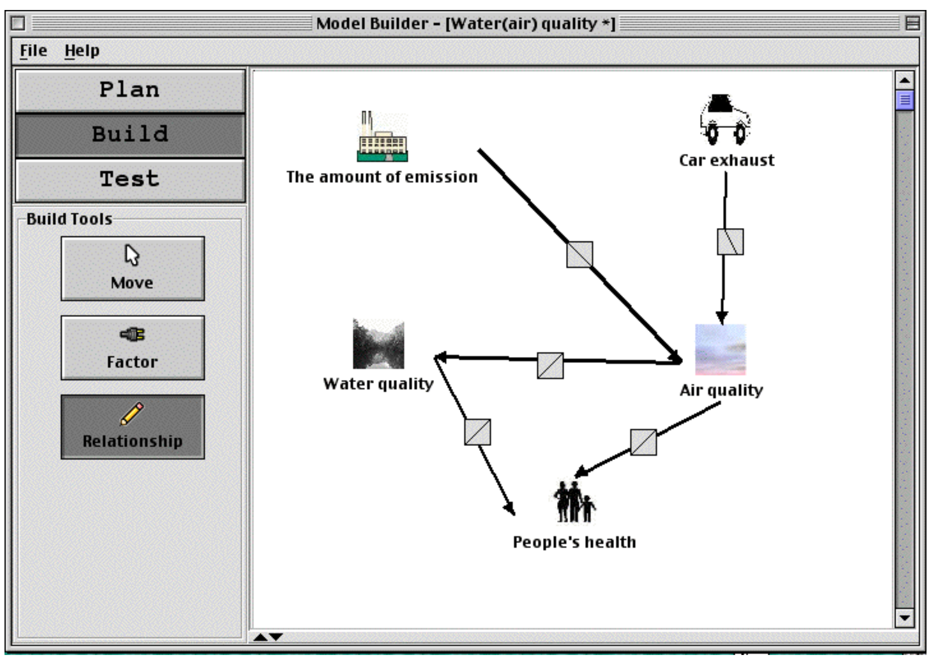

Figure 2.2: Build Mode (icons for factors e.g., Cars:Number of Cars, and arrows for relationships e.g., As Cars:Number of Cars increase then Pollution:Amount increases by about the same)

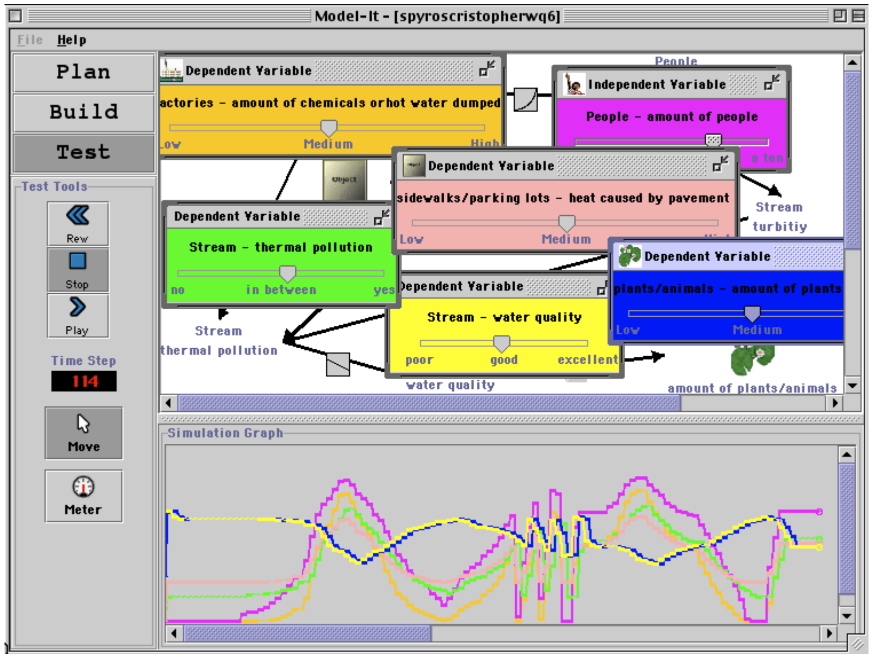

Figure 2.3: Test Mode, with meters up for all factors, and values over time plotted on graph)

manipulated) to observe model function (Figure 2.3). Changes in factor values over time can be graphed for examination. A bad model would be one that returns an unexpected result, such as setting input factors like pollution and predators to a high level, but still having an increasing number of fish. This debugging can lead to further discussion of model structure and science content, and then to revision of the model.

2.6.1 Prior Research on Model-It

Model-It has been the subject of a variety of research over the years, as summarized in Table 2.2 below. In the earliest studies, the focus was on the usability of the tool and determining what practices modelers engaged in. Model-It's initial programmer, Jackson (1996) and developer Stratford (1999) both found that Model-It was successful in allowing young learners to construct models of substantial quality. Stratford analyzed relationships within, and function of, learner models and found most models to have coherent structure, complex relationships, and good fidelity to the scientific concept being modeled. Focusing on the scaffolds themselves, Jackson found that most but not all scaffolds were used by the learners, though some were generally ignored (e.g., reflective scaffolding provided for them to explain what they were doing). This early version allowed students to remove or "fade" some scaffolds and learners generally did so as they gained expertise. Stratford assessed the modeling strategies used by learners and found that learners engaged in a variety of strategies similar to those used by experts, and additionally that they created models of substantial quality.

In follow-on research in different settings Shrader (2000) and Singer (2000) both examined the classroom environment around the use of Model-It. Shrader distilled a set of seven instructional strategies to assist teachers in using Model-It, which was an early acknowledgement of the importance of interpersonal scaffolding around scaffolded tools. Singer found that classroom supports enabled the best use of the Model-It tool, and that while learners created valid models they did contain one or more errors.

Research by Fretz and colleagues (2002) began to assess the quality of the artifacts made by modelers and explore the use of scaffolds in and around the tool. They found that teachers provided important scaffolding support in the use of Model-It but that peer scaffolding did not play a major role in the classroom. Additionally, using a simple model scoring rubric, they found that model quality tended to increase with multiple exposures to Model-It but that the increase was not orderly. Zhang (2003) examined the change in modeling practices over time was assessed, along with the ability of certain scaffolds to support modeling practices. Zhang (2003) noted the importance of peer discussion and teacher guidance in using the Model-It tool to learn modeling practices.

In summary, research has generally shown that learners can use Model-It successfully, that is they create models that demonstrate understanding of the components

Researcher / Year	Question(s)	Findings	Note
Jackson, S 1999	How well did scaffolding in Theory Builder work? Does it support the development of expertise by fading?	Most scaffolds worked, all were faded somewhat by almost all Ss. All Ss were able to model immediately. Reflective scaffolding could not be evaluated as it was used too infrequently.	GLAS also allowed ADDING features and complexity
Stratford, S 1996	What are the modeling strategies engaged in by middle school science learners and what is the quality of the model they produce?	Students engaged in a variety of modeling strategies, and those who used these strategies produced Models of substantial quality, reasonable scientific validity, avg 15 factors and relationships	
Shrader, G et al., 2000	What difficulties to students face when they engage in modeling tasks? What instructional responses seem to scaffold students through those difficulties.	Developed 7 items in a diagnostic toolkit to help teachers deliver instruction with Model-It, and deal with cognitive difficulties that students face.	Teacher assistance with use of tool was central point of study.
Singer, J 2000	What classroom supports enable technology to be introduced seamlessly, so students could construct appropriate air quality models?	Classroom performance supports did work (most effective when all four phases were brought together), but most models contained at least one error.	Teacher support of tool use was critically important.
Fretz, E et al., 2002 (AERA)	Do software scaffolds in tool seem to work when examined in detail? Are scaffolds from peers and teachers vital to the use of Model-It? Do we observe changes in scaffold use over time?	Teachers provided important scaffolding, more so for 7th graders. Using the INSPIRE categories, most scaffolding from teachers was in Socratic or Reflective mode. Peers could not be evaluated for scaffolding, not enough evidence that met coding standards Peers were more of a factor in 8th than 7th grade. Models scored were better form 8th graders than 7th graders, but did not show an orderly increase in scores by unit.	
Fretz, E et al., 2002 (RISE)	How often do modeling practices occur with scaffolds? Which modeling practices are most effectively supported by certain specific scaffolds? What does student use of scaffolding look like as they engage in modeling practices?	Some but not all modeling practices were supported by tool scaffolds. Both the articulation and dynamic testing scaffold supported the most modeling practices.	Calls for research on role of teachers and peers in association with tool scaffolds. Comparison of articulation/ explanation with model quality.
Zhang, B 2003	What modeling practices do we see learners engage in while using Model-It, to they change over time, and what features of the learning environment facilitate modeling practices?	Modeling practices elaborated, they became more integrated, meaningful and purposeful over time. Effective collaboration between peers and the guidance of the teacher were both important contributors to learner success.	No scaffolds evaluated directly.

Table 2.2: Summary of previous research on Model-It.

and function of complex ecosystems. The modeling practices that learners engage in with Model-It have been elucidated and studied (Stratford, 1996; Zhang, 2003). The overall design of scaffolds to make the software usable has been validated (Jackson, 1999), and initial research on specific tool and interpersonal scaffolds has shown they are being used with some success (Fretz, 2002). However, initial investigation of scaffolds was somewhat simplistic, and involved mainly enumeration of scaffold instances by type, or assessments of how often certain modeling practices occurred with certain scaffolds. This research did show, however, that the tool scaffolds of articulation (text boxes) and dynamic representation (test mode) are potentially rich sources of evidence on how Model-it supports learning about science content. A more detailed examination of how Model-It is used over time, and how model quality varies over time could extend this line of research.

2.6.2 Articulation in Model-It

As discussed earlier, articulation is a vital and desirable facet of learning that shows up in almost every description of scaffolds in ILE's. Articulation is also explicitly or implicitly addressed in descriptions of modeling practices. The construction of models and the need for collaboration and discussion are both cited as core principles for science learning (e.g., Jacobson & Wilensky, 2006). Having learners articulate their descriptions of objects or factors, and explain WHY their relationships function the way they do is instrumental in making their thinking visible and creating a shared understanding with their peer(s). The articulation text boxes in Model-It are designed to encourage learners to articulate their reasoning when building each piece of their model (seen below in Figure 2.4). As they first create Objects, they can select a representative icon, and then enter a description in a blank text box.

In the next stage, as they create Factor(s) for each Object, the Factor window similarly allows customization (for example specifying a 14 pt scale for a pH factor) and again the learner can enter a description in a blank text box (Figure 2.5).

Finally, as the learners connect one Factor to another with Relationships, the rela-

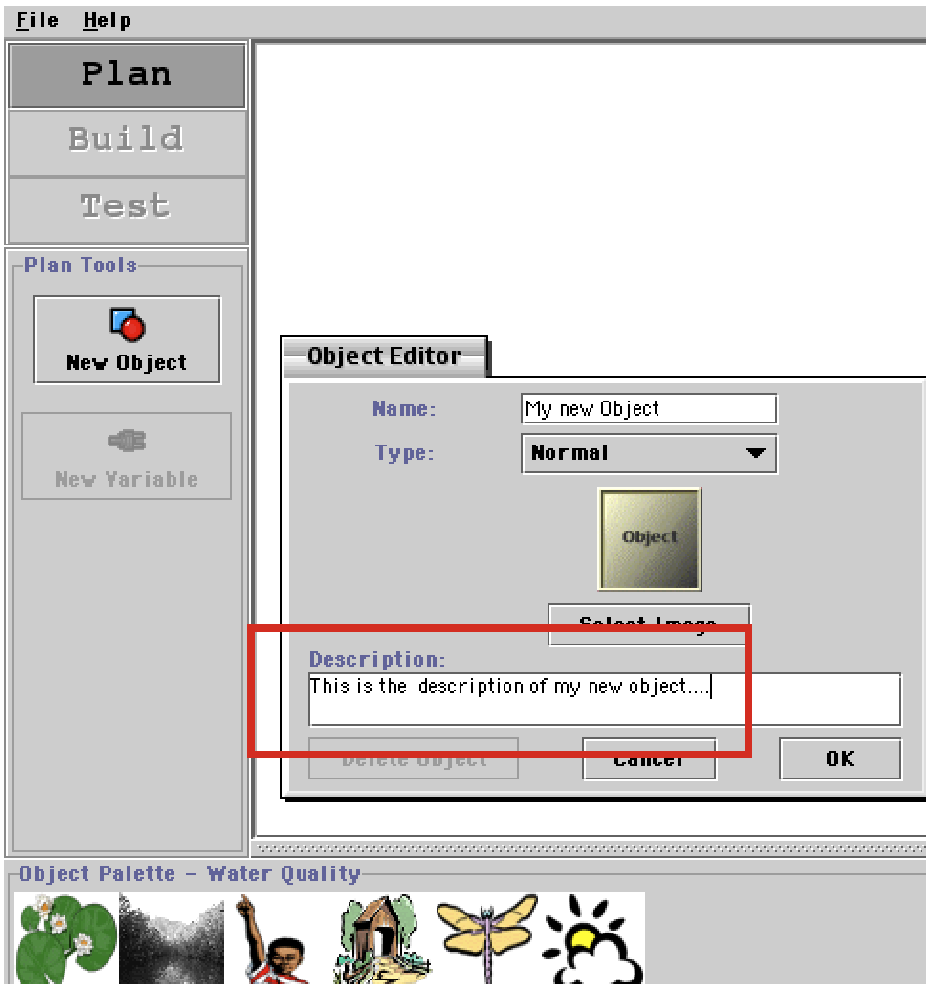

Figure 2.4: Plan Mode (articulation box in red)

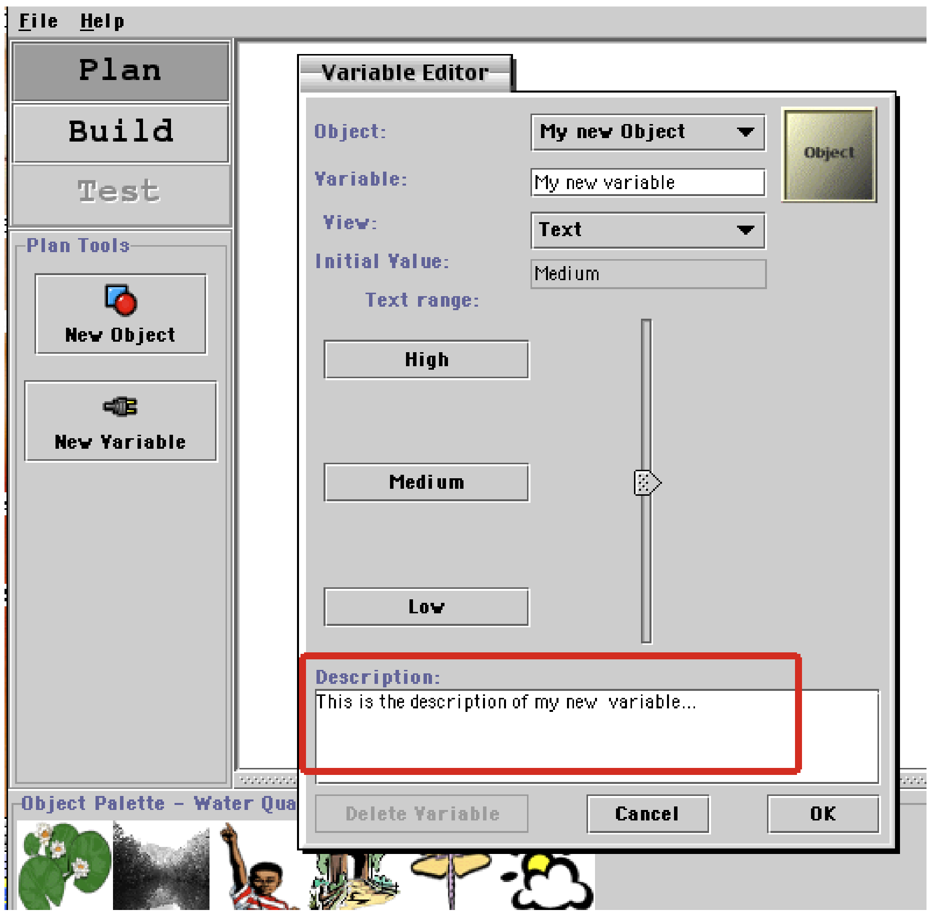

Figure 2.5: Build Mode, Factors (articulation box in red)

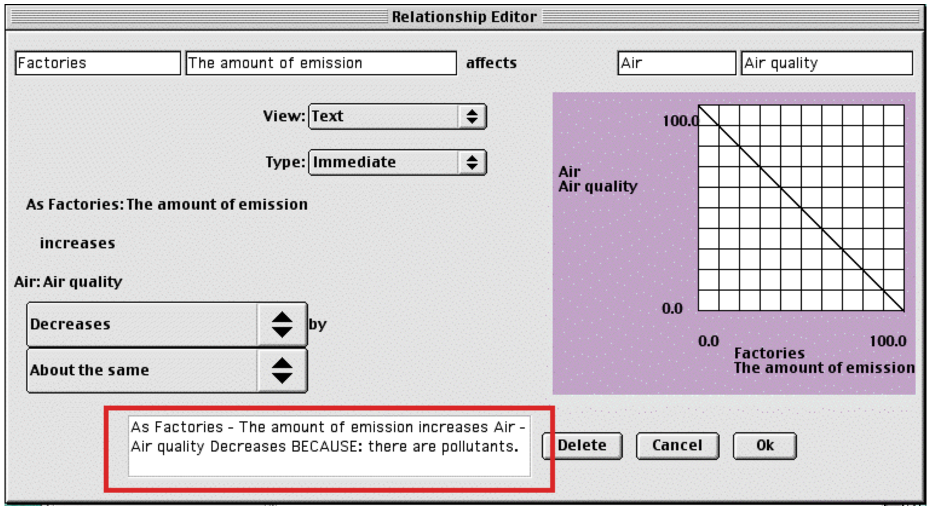

Figure 2.6: Build Mode, Relationships (articulation box in red)

tionship editor also has a text box (Figure 2.6). This text box is slightly different, in that in contains a partly filled out in the form of "as X increases Y increases/decreases, because" (the X and Y are filled in by the program based on what factors were selected). The Object and Factor boxes are blank, and the Relationship box has only a generic sentence completion prompt, and from the earliest research on Model-It (Jackson, 1996) it has been noted that students will often leave these boxes blank without teachers setting expectations and offering reminders. Research by Davis (2003) has also shown that while learners make more productive use of generic prompts, they do so when the overall interface provides highly scaffolded prompts for activity completion.

Although learners do not create self-explanations easily or consistently, research by Chi (2000) has shown that eliciting self-explanations from learners can help them develop better understanding. While these text boxes are a more simplistic form, research on similar "activity prompts" (Davis & Linn, 2000; Linn & Songer, 1991) showed that they assisted students in completing a task like articulating their reasoning behind a scientific design or argument point. Luchini (2003), in work with PiCoMap (a collaborative concept mapping tool for handheld devices), used a similar blank "add notes here" box to encourage learners to describe concepts and relationships created, because the "explanatory process can help clarify the current understanding" (p. 381). Guzdial (1994) found in his work on scaffolding learners mastering programming skills, one of three key areas where software could most help learners was eliciting articulation. Similarly, research on Linn's scaffolded knowledge integration framework emphasized the importance of articulation and reflection for learners to improve their understanding of science process and content (Davis & Linn, 2000; Linn, Davis, et al., 2004).

Scientific investigations "require the complementary processes of reflection and articulation" (Resier, 2004, p. 277). In the initial design of Model-It, these text boxes were referred to as "reflective scaffolding" (Metcalf, et al., 1999), with the goal of eliciting articulation about learners' conceptions of objects, factors, relationships, and the science behind them. Requiring an explicit shared representation helps learners create a shared understanding (Roschelle, 1992).

2.6.3 Dialog Around Model-It Scaffold Use

An underlying assumption about the articulation scaffolds in Model-It is their potential to create cognitive conflict, as noted in the earlier discussion of the value of modeling in science learning. By requiring learner pairs to specifically describe their factors, and articulate the reason(s) for their relationships, the articulation scaffolds can surface disagreements and misconceptions. For example, when required to describe the "why" of their relationship between pH and fish population, a learner pair may be forced to confront misconceptions about the pH scale and have a debate about what type of relationship to use, finally agreeing to use the "bell curve" type of relationship. First coined by Festinger (1957) as "cognitive dissonance", this concept supposes that disagreement can create an aversive state and that learners will be motivated to resolve the dissonance. This is similar to the Piagetian concept of disequilibrium, something that sounds bad, but is actually desirable as it drives learners to modify their existing knowledge structures. Since the goal, referred to in the earlier discussion of models in science education, is to have learners develop robust and coherent understandings of science content and concepts, it would seem that the more cognitive conflict the better. In a study of learners using a scaffolded computer tool, Hmelo and colleagues (2000) found that moderate amounts of socio-cognitive conflict can also drive learning if it leads to an attempt to produce coherent understanding.

Model-It's articulation text boxes have been shown to assist with a number of modeling practices, particularly analyzing, synthesizing, and explaining (Fretz, 2002b). While articulation and learner discussion can lead to more explicit and thoughtful work (Reiser 2004), and classroom argumentation between learners can increase content understanding (Bell & Linn, 2000), it can be difficult to get learners to recognize they need to articulate their reasoning (Scardamalia & Bereiter, 1991) and even when given opportunities to do so, they may comply in a non-thoughtful "fill in the blank" fashion (Davis & Linn, 2000). In the previously discussed scaffolding framework of Quintana and colleagues, Guideline 7c was "providing guidance including simple reminders for articulating ideas to promote sensemaking" (Quintana et al., 2004, p. 371). This scaffold is instantiated in Model-It in the articulation text boxes, and an investigation of how these scaffolds are used by learners in relation to their discussions could provide productive empirical validation.

2.7 Potential Contributions

In this chapter I have reviewed the literature on scaffolding and modeling. I have traced the evolution of scaffolding as a term and construct, and showed the various ways it has been studied and defined in various lines of research. I have highlighted key areas of disagreement or concern with definitions of scaffolding. I have argued for the value of learner articulation as a scaffolding goal, and the value of examining learner discussion. The value of models and model building for science education has been discussed, and past research on the dynamic modeling tool Model-It has been reviewed.

This study examines the following questions: What types of student and teacher interactions occur with and around scaffolds in a dynamic modeling tool, and do they change over time? Specifically, do tool scaffolds for articulation lead to learners articulating their thinking in writing and does this change over time and, while using tool scaffolds for articulation, what is the quality of discussions learners engage in and does this change over time, and if so, how? Then, what is the quality of learners' models and does that quality change over time, and if so, how? Specifically, what is the quality of learner pairs' final model artifacts, and does the quality of these models, in terms of content or complexity, change over time, and if so, how? What is the pattern between learner interactions with and around scaffolds and the quality of their final model?

There are a number of areas in the literature where this study can make contributions. In relation to research question one, which seeks to characterize scaffolding in and around Model-It, the study could:

- Provide empirical validation of articulation scaffolds that are commonly found in scaffolding frameworks but have not yet been explicitly tested.

- Extend current work in the area of interpersonal scaffolding behaviors around the use of Model-It and similar scaffolded tools (e.g., empirical validation of previous teacher/tutor taxonomies that were conceived and applied in a technology-free educational context), to consider teachers' use of strategies in the context of ILE's as well as the interplay between tool and interpersonal scaffolding.

In relation to research question two, which seeks to compare scaffolding use with the quality of the model artifact, the study could:

- Provide validation of a comprehensive model evaluation rubric that combines the best of many prior evaluation schemes.

- Examine the role of dialog during written articulation.

- Extend prior work assessing the quality of models learners can create with a scaffolded modeling tool, by tracking results over time through multiple curriculum units.

- Provide support for the argument that the successful use of scaffold use in and around technology tools leads to concrete benefits or improved learner artifacts, as shown by any relationship between the nature of scaffold use and final model quality.

In the next chapter, I detail the methods used to gather and analyze data in support of the research questions in support of the above contributions.

CHAPTER III

Methods

3.1 Introduction

This chapter reviews the context and data sources for this study and lays out the rationale for, and guidelines used in, the analysis of the data. Pairs of middle school science students' use of a computer based modeling tool are examined by coding transcripts and analyzing them with a qualitative analysis tool, and by scoring the artifact created by the learners (the model) using a rubric. First I describe the context of the study, to include participants, environment and technology, as well as how the source data were collected and the rationale for data digitization and reduction. I then put forth the study design and describe and justify coding, and analysis techniques organized around each research question.

3.1.1 Research Questions

1. What types of student and teacher interactions occur with and around scaffolds in a dynamic modeling tool, and do they change over time? Specifically related to question one, I ask:

    (a) Do tool scaffolds for articulation lead to learners articulating their thinking in writing, does this change over time, and if so, how?

   (b) While using tool scaffolds for articulation, what is the quality of discussions learners engage in, does this change over time, and if so, how?

2. What is the quality of learners' models and how does that quality change over time? Specifically related to question two, I ask:

   (a) What is the quality of learner pairs' final model artifacts?

   (b) Does the quality of these models, in terms of content or complexity, change over time, and if so, how?

   (c) What is the pattern between learner interactions with and around scaffolds and the quality of their final model?

3.2 Context

The subjects of this study were pairs of 7th and 8th grade science learners in an independent 6-12 school in a mid-sized midwestern university city. The software (Model-It) was integrated into a project-based curriculum that has been in use for several years. Model-It was used several times throughout a cycle that included all the 7th grade school year and the beginning of the 8th grade school year. One of these cycles was observed over an 11 month period. The timeline, class, and pair structure are detailed in Figures 3.1, 3.2, and 3.3.

3.2.1 Learners

The learner population for this study was quite homogenous. The learners were almost entirely white and upper middle class, and nearly all had access to a computer at home. In classes of around 18, the learners were paired, generally in mixed gender, for the duration of the curriculum unit (which included the modeling activity) with one other student, with whom they collaborated in and out of school, and with whom they completed their model and any other projects for that unit. In one case, three students worked together. The total number of learners involved across all units was 46, with 26 males and 20 females. In each class, teachers selected two pairs of target students to be video-taped, with the criteria of having learners who were neither particularly high or low performers, and who would be likely to work well together and, most importantly, verbalize their thinking while working on the computer. Learner pairs were typically given 2-4 days to create a model after they had been exposed to related content for several weeks (e.g., trips to examine local streams and conducting water quality tests).

3.2.2 Teachers

Three experienced science teachers taught the classes in this study. Tara, with 11 years experience, taught 1 curriculum unit. Ursula, with 12 years experience, taught 8 curriculum units. Sandra, with 27 years experience, taught 3 curriculum units. These teachers had all taught at this school for many years, and were well versed in project-based learning and use of technology in their classrooms. They had previously integrated various technologies into their teaching, such as hand-held computers with testing probes, and digital cameras. The teachers taught in a pair of classrooms joined together by a common prep and storage room. These teachers were the entirety of the science department for the middle school. They maintained a highly collegial atmosphere with each other, and employed an eclectic, highly engaged, cooperative teaching style (e.g., there was a mix of lecture, demonstration, activities, and technology, and the students were required to work in pairs and large groups as well as do solo work). The teachers had experience using an earlier version of Model-It in their classrooms prior to the time of this study. Model-It was added to an established science curriculum (Novak & Gleason, 2001; Novak & Krajcik, 2006) during project based units (Blumenfeld, et al., 1997, Krajcik & Blumenfeld, 2002) dealing with water quality and decomposition (7th grade) and weather (8th grade). Prior to their first use, the students were introduced to Model-It through a one class period demonstration by the teacher. The teacher reviewed the modes and functions of Model-It while demonstrating how to create a basic model. Model-It was then used in four curriculum units taught by these teachers, interspersed through the school year(s) with other science topics and units.

3.2.3 Technology

The classrooms each had 8 networked iMac computers, linked to a teacher station that functioned as a server. Students used the computers for a variety of other classroom tasks in addition to Model-It (such as report generation and web research). Students used the computers in pairs, and very occasionally in groups of three. The version of Model-It was quite reliable, with only occasional bugs. Model-It was one of many technology tools used in these classrooms during year-long use of project-based science curricula. Two of the iMacs were connected to a Process Video (Krajcik et al., 1988) setup, and the target student pairs sat at these machines and wore microphones while using the computer to create their model. Students saved their models daily on their local computer, and at the completion of each day the researchers would copy the model files to a designated folder on the teacher's server.

3.2.4 Curriculum Units

The specific curriculum units were similar in their project based focus, and use of Model-It, but did have some differences in scope and content. The first two units were both on Water Quality. In the first Water Quality unit, the students were introduced to relevant terms and the concepts of a water cycle and a stream ecosystem. They learned about water quality tests and used handheld probe technology to perform tests on the water in the stream behind their school. After this unit, another science topic, such as electrical circuits, was taught. Later in that term, Water Quality was re-visited in the Water Quality 2 unit. Here the students learned additional concepts such as the effects of pollution and more complicated water quality concepts like eutrophication.

Later in the year, after Water Quality 2 and another unit, the Decomposition unit introduced concepts of how trash decomposes (or fails to decompose) based on the nature of the trash, and the factors present (e.g., moisture, heat, worms, bacteria, etc). The student pairs constructed two "decomposition towers" in which they place various materials (e.g., banana peel, Styrofoam peanut, coffee grounds, dirt, etc) and then observe the state of decomposition over the next month. The pairs can also place their towers in warm or cool, light or dark environments so they can observe the differences caused by these factors. They record their observations in a notebook, as they do with each unit. Finally, in the first unit of the 8th grade year, the students participate in the Weather unit. In this unit they make weather observations and learn the various measurements taken to predict weather, as well as learning about various atmospheric effects and the nature of storms. In all of these units, one of the summative assessment activities is to have each pair of students make a model using Model-It that reflects the information and concepts they have learned in the preceding weeks. Once all students have completed their models, they take turns presenting them to the entire class, where they receive feedback on the content and structure of their model.

3.2.5 Classroom Norms and Model-It Introduction

The science classes in question were taught by experienced teachers using a smaller number of units to cover topics in depth using a variety of authentic experience-based activities. As detailed previously, the creation of computer models was included in several but not all of the units. Typically, Model-It was used as a summative evaluation activity where students were expected to show what they had learned by demonstrating their model to their classmates. Due to teacher reminders, student pairs were acutely aware of the need to develop their model for peer-review. These reviews provide the benefit of peer and teacher feedback, but also encourage more rigorous engagement of the material (Linn, Bell & Davis 2004), similar to the "pin up sessions" of Kolodner (2003). Generally the students would have several weeks of classroom discussions and activities before using Model-It (see Novak & Gleason, 2001, for a complete discussion of curricular implementation). The atmosphere of the classrooms was collegial, and once the unit was completed the students all took turns displaying and critiquing their models projected on a large screen.

Model-It was introduced to these middle school students in early 7th grade, after which they used it repeatedly in 7th and 8th grade. There was no formal instruction on the software after the first introduction, however ongoing training on software use and model creation was done in one-on-one interactions between teacher (or researcher) and the student pairs. To introduce Model-It, the teachers would display their screen to the class, and the students would observe the creation of a simple model using the software. The teachers would review how to create a new model, save a model file, and find/open an existing model file. The software had no tutorial or help mode, but the interface of the software made the features obvious enough that students could begin to work effectively after the initial 30 minute introduction.

A typical day using Model-It would involve the students arriving for class and the teacher establishing the goals for the day and then sending the students to the computers. If this was the second or third day of use, the students would often proceed to the computers to begin work immediately. The teachers would then cycle through the room, checking on each pair, and asking about their progress, reviewing their model, and verifying that they were on track. Teachers would also make clear before and during class that objects, factors, and relationships needed to have explanation boxes filled in. Sometimes teacher assistance was as simple as a software tip, and other times it involved a major conceptual revision to the students' model, or instruction to correct a misconception made evident by the model. If a problem or misconception was found more than once, the teachers would generally stop the class for a moment to address the issue with the entire class. Student pairs typically worked on models for the entire class period and then saved and closed their models when they were complete or when instructed by the teacher.

The role of the researchers was primarily to set up equipment, resolve any technical difficulties, and observe. Due to the long-term nature of the data collection, the researchers inevitably became seen as extensions of the teacher and on occasion were sought out for help with content or structure questions about the models. On those (relatively rare) occasions, the researchers would attempt to provide the same style of help that the teacher gave.

3.3 Data Sources and Collection

The data corpus for this study is comprised of Process Video (PV) tapes (Krajcik, Simmons & Lunetta, 1988) of learner pairs as they used Model-It software, and the resultant Model-It model (computer) file. Each analyzable "case" consists of a student pair using Model-It for one exposure (2 to 5 class days, one PV tape per day, an average of 70 minutes of video) to produce one model artifact. Process video tapes and model files were collected from two "target student" pairs in each classroom each day that Model-It was in use. Each "exposure" to Model-It lasted from 2 to 5 days, so each case has from 2 to 5 transcribed tapes, plus the final model file.

3.3.1 Structure of Classes and Data Collection

For the first year of a cycle (seventh grade) the students were exposed to a water quality curriculum in two phases (fall and spring), and they created models both

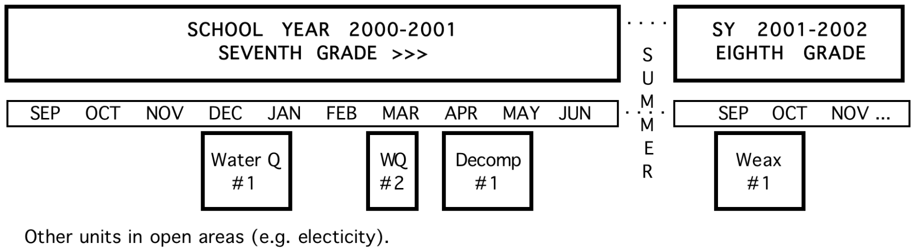

Figure 3.1: Units during school years

times. The multi-week curriculum provided classroom lectures and activities on water quality, stream ecology, and various natural science content, as well as outdoor activities such as stream-walks, and water quality testing in a stream adjacent to the school. Later in the first year, students completed a curriculum on decomposition that again contained classroom discussions and activities, centered around the construction and observation of "decomposition towers" in which each pair created one or more terrarium-style closed environments in which to observe the effects of time, temperature, and other variables on the decomposition of various types of garbage (e.g., orange peel, newspaper, bottle cap). Finally, in eighth grade, the students participated in a curriculum unit on weather, in which they learned about weather forecasting and dynamics and made observations and predictions. In every one of these units, the students were required, towards the end of each unit (and occasionally at the beginning), to create a final model that represented their understanding of, for example, water quality in the stream, and then present that model to the class for discussion and review. For the modeling phase of each unit, target students' final models were collected and their use of Model-It was recorded using a process-video setup, which captures all screen activity along with audio of learner discussion.

Figure 3.1 shows how students "re-pair" for each unit, going "over time" to the right. In year 2, class #3 was taught by Tara, for that year only. Learner pairs were not held consistent over the calendar year, in keeping with long-standing teacher practice of varying learner pairs over time for variety and fairness. While this lack

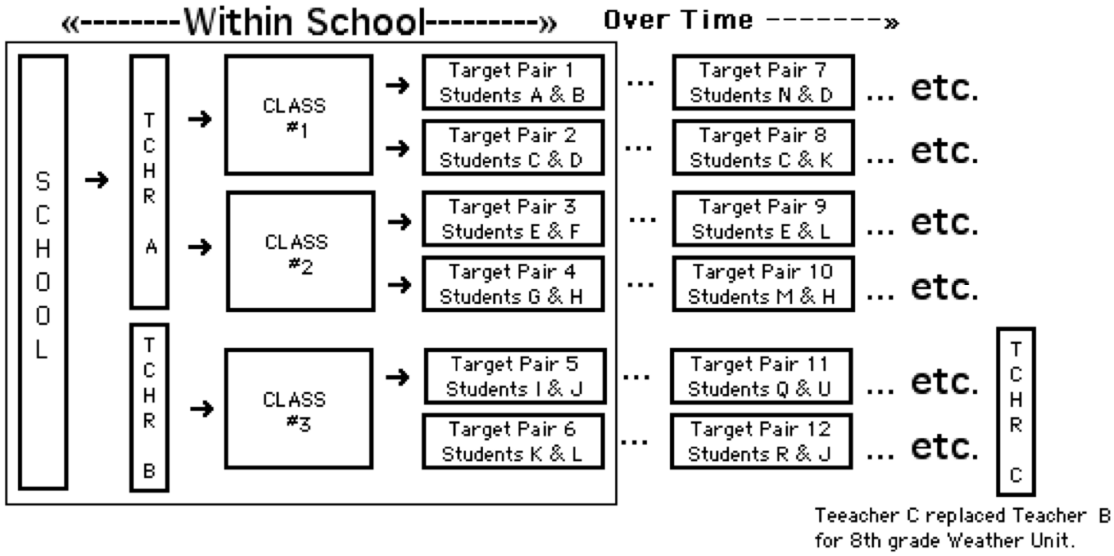

Figure 3.2: Year 2 - showing year/class/unit/target pair structure.

of consistency was not ideal because it prevented direct pair comparisons over time, this choice was made to avoid unduly disrupting the normal classroom routine. The implication of this for data analysis is that each learner pair will have to be treated as a separate instance (case) of model creation within a curriculum exposure, and that arguments about change over time will have to be about the learner pairs in general, as opposed to tracking the change of a particular pair.

As seen in Figure 3.2 each class can produce up to two analyzable cases. Although the two years of data collection produced dozens of potential cases, some cases are not usable due to model file corruption or inadequate video files. Such technical difficulties are not uncommon in research using similar protocols (e.g., Sins, et al., 2005). The research team repeatedly fielded upgraded versions of the Model-It program (due to stability and function upgrades, not changes to scaffold design). In some cases, this resulted in corrupted model files being saved. In other cases, video tape failed to record, or the recording unit became unplugged. In still other instances, a student might have worked alone for multiple days due to a partner's absence, resulting in completely silent work, which could not be effectively analyzed. In order to be an

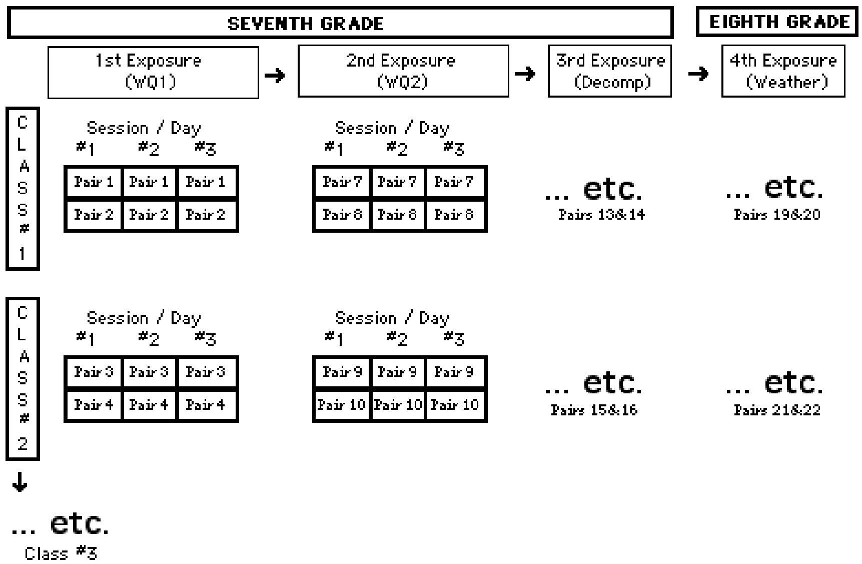

Figure 3.3: Data Gathering
Showing how cases are gathered from classes. Essentially, each class (with two pairs of target students) produces two analyzable cases in each curriculum exposure.

analyzable case, all the tapes and the model file for a given pair had to be clear and uncorrupted. Out of a possibility of 24 cases, a total of 20 cases are developed across the four units: five cases for Water Quality1, four cases for Water Quality2, five cases for Decomposition, and six cases for Weather.

Each case reflects the substantive content of one to four PV tapes and one digital model file. The tapes follow the learner pair in their creation of the model from start to finish. The use of process video has a long history since the technique was first described (Krajcik, et al., 1988; Jackson, 1999; Zhang, 2003). It is important to acknowledge the limits of the source material (Merriam, 1998). These tapes do not include all classroom context, although sometimes teacher interactions with the entire class are captured. The tapes also do not capture student gestures other than their on-screen gestures with the cursor. On a daily basis, the PV tapes were collected after students were finished with the computer. Tapes were labeled and logged into a master database. Tapes were later converted to digital video files for ease of analysis and storage.

The students' final models can be rich sources of evidence for their understanding of complex scientific phenomena, as was noted in chapter 2. The Model-It files are a dynamic representation of the learners' understanding of the science content. Each file contains all of the chosen graphics, the objects-factors-relationships making up the model itself (to include all required descriptions), and it remains a working simulation that can be opened and examined. Each day the learner pairs were to save their file under a new name on the local hard drive, and all model files, including the final completed model, were saved to the teacher server on a daily basis. At the end of each unit, all model files were collected onto digital media for storage and later analysis.

3.4 Data Reduction

The reduction of the data corpus was accomplished in several steps. First, the video tapes were digitized to allow easy storage, retrieval, and review. Second, the dialog on the tapes was transcribed, and then the digital video was reviewed as many times as necessary to add to the transcript all relevant time marks, researcher comments, and observed screen gestures. An example of a transcript can be found in appendix B. The transcript records all verbal interactions as well as noting the non-verbal screen gestures, which are required to fully understand how the students are constructing meaning (Roschelle 1992). The use of transcripts is a cornerstone method for qualitative analysis (Miles & Huberman, 1994), and the quantification of qualitative and verbal data is well described by Chi (1997). All substantive student interactions about the modeling tool or science content were transcribed, but lengthy off-task and social discussion were noted without full transcription. In particular, a full explication was made in the record of relevant screen activity (e.g., cursor movements, use of interface options) to ensure as much of the students' meaning as possible is captured. Third, the model files were opened (this required the use of various hardware and system configurations due to software revisions during the study) and a digital photo / screen capture (JPEG format) was made of all aspects of the model to allow review of the model in contexts outside of the unique hardware/system/software on which it was created. Using the screen captures and the tape dialog, a "model summary" text document was created that captured the number, type, and contents of every component in the model (i.e. Objects, Factors, and Relationships). Each model was tested and run, and observations about model function were recorded in the model summary text document. Fourth, all these digital text sources were imported into an NVIVO database for subsequent coding and analysis. NVIVO is a fifth generation, industry-leading software package that has been broadly used in qualitative analysis for over a decade. The NVIVO program allows for dynamic "in vivo" coding of various media content, facilitating qualitative analysis. A precursor version was used successfully in multiple studies (e.g. Fretz, et al., 2002a/2002b; Zhang 2003).

Each case's transcripts and model were coded using the coding schemes detailed below. NIVIVO qualitative analysis software was used to store and manage all source documents, coding schemes, and codes. The use of NVIVO (and the underlying database engine that drives it) allows the codes tagged to any given document or portion of text to be viewed, searched for, counted, compared and contrasted. Once all transcripts were imported and coded, the result was a set of cases in an NVIVO database that can be subjected to iterative analysis and further coding. The process of coding was iterative in the sense that the a priori coding scheme developed through the literature and prior research was modified and clarified as a result of coding initial cases. For example, additional demographic codes, such as learner age or gender, could be added. An existing code could be split into two codes if additional differentiation becomes obvious, or a new code could be added for unexpected themes that show up repeatedly, such as, hypothetically, student groups interfering with each other's progress. The resulting large database of coded data could be subjected to various queries in support of the research questions. Queries performed are discussed below in relation to each research question.

3.5 Coding for Research Question One

For question one, "What types of student and teacher interactions occur with and around scaffolds in a dynamic modeling tool, and do they change over time?" the primary data source analyzed was transcripts of videotapes that capture conversation and on-screen use of the software. Model files were referred to on occasion when verifying what the students were typing into the various articulation boxes.

Research Question	Data Source(s) Used	Coding Scheme Used	Hypothesis (I predict that...)
1A	Process Video (PV) transcripts, Model Files (to verify what was typed)	Primarily section 2 of appendix A	With variation by pair, learners do articulate their thinking in writing, but it may degrade over time as they gain experience and put less emphasis on the scaffold.
1B	PV Transcripts	Primarily section 4 of appendix A	Learners engage in high quality discussion around the articulation tasks, and this will improve over time as they become more comfortable with collaborative construction of their models.

Table 3.1: Coding for question 1A. Research question one concerned the nature of articulation that occurs in the scaffolds designed into the program. Articulation was coded at four levels: Robust, Simple, Ignore, and Incorrect. Specific exemplars of each type, taken from the data, can be seen in table 3.2.

OBJECT	
Correct and Robust	Learners make a complete and scientifically accurate text entry. Example: for Object "Clouds" they enter "visible moisture in the air"
Correct and Simple	Learners make scientifically accurate but simple text entry. Example: for Object "Decomposition" they enter "rotting"
Ignore or Gibberish	Learners type nothing into the box, or type random text. Example: "asdfhjhk"
Incorrect	Learners enter scientifically inaccurate or unrelated text. Example: "spoon" or for Object "Stream" they enter "nice"
FACTOR	**Same as for OBJECT** (text entry box is similar)
RELATIONSHIP	
Correct and Robust	Learners make a complete and scientifically accurate text entry, including the "because" statement. Example: "As Sun : Amount of sun increases then Air : Relative Humidity decreases by about the same BECAUSE – sun increases temperature and warmer air can hold more moisture"
Correct and Simple	Learners make a scientifically accurate but simple text entry, including the "because" statement. Example: "As Stream : Dissolved Oxygen increases then Fish : Population increases by more and more BECAUSE – fish need to breathe"
Ignore or Gibberish	Learners type nothing into the box, or type random text. Example: "asdfhjhk"
Incorrect	Learners enter scientifically inaccurate or unrelated text. Example: "because of shiny awesomeness" as the reason why Stream:Temperature increases with Sunlight:Amount.

Table 3.2: Codes for articulation in program scaffolds

Ignore is the simplest code and was assigned when the learners simply failed to enter any data in the required articulation box. Incorrect was a code used when the description or text was scientifically incorrect or nonsensical. Simple and Robust codes were used when the description or text was correct, with the essential distinction being that Robust requires something beyond simple description. The ideal instructional goal for Model-It use would be to have no incorrect or ignored articulation scaffolds, and then to have the majority of the articulations be correct and robust.

The coding was an iterative process that was top-down as well as bottom-up (Chi, 1997, Miles & Huberman, 1994), and the a priori coding scheme (appendix A) was modified to accommodate emerging patterns or themes that were observed. This subset of coded data (transcript and model summaries) was presented to a second trained coder, to assess reliability of the coding scheme. Using the coding scheme and a basic explanation of the codes, the second coder matched 90% of the time, and where codes did not match they were similar ("correct & simple" vs. "correct and robust"). After discussion and training on three models, three additional models were scored and agreement was 95%.

3.5.1 Coding for Question 1A

The codes for scaffold use are similar to those used in previous research (Fretz et al., 2002a). Borrowing the concepts of use, accuracy, and reflectiveness from the studies that directly assessed scaffolds in Scaffolded Work Environments (SWEets), (Luchini, 2003; Quintana, 2000), this set of codes covers each instance of the articulation scaffold. Articulation scaffolds are coded for the manner in which the learners use them (or ignore them, as listed in table 3.2). The quality of their scientific statements (in terms of what is specifically entered in the articulation box) are assessed with these codes. Based on prior research (Fretz, et al., 2002a) codes were designated for interac-

TYPE	CODE DESCRIPTION	ACTUAL EXEMPLAR (case/tape) (typed articulation in bold)
Ignore or Gibberish	Learners type nothing into the box, or type random text. Example: "asdfhjhk"	Any instance of empty articulation box. (Case 07 / tape 197 Student pair of two males on first day of decomposition model, had one doing all the work, and his practice was to create an object and then immediately create an associated factor without discussion and without any articulation for the object. 100% of the objects in this model had empty articulation boxes, but the overall model was Average.)

Table 3.3: Exemplar of Ignore/Gibberish Articulation

tion with each tool scaffold (4 master codes: Object, Factor, Relationship, Dynamic testing), and for interpersonal scaffolds (3 master codes: Teacher, Researcher, Peer). Each master code contained sub-codes that related to the nature of the learners' interactions with the scaffold, which characterized the interaction in terms of whether the scaffold was ignored or used, and if used, the accuracy and depth of the input (articulation). To further clarify how the coding occurred on actual source material, tables 3.3, 3.4, 3.5, and 3.6 provide actual samples of coded text as well as the context from the case in question.

3.5.2 Data Analysis for Question 1A

The queries in support of research question 1-A, and sub-questions 1 and 2, "Do tool scaffolds for articulation lead to learners articulating their thinking in writing, does this change over time, and if so, how?" involved the codes from section 2 of the coding scheme (appendix A). To answer the question of how often and well articulation scaffolds were used, the coding for *all* use of articulation scaffolds for Objects, Factors, and Relationships (codes 2.1.1, 2.1.2, and 2.1.3) can be grouped by the four sub-codes: Robust, Simple, Incorrect, and Ignore. The NVivo software

TYPE	CODE DESCRIPTION	ACTUAL EXEMPLAR (case/tape) (typed articulation in bold)
Incorrect	Learners enter scientifically inaccurate or unrelated text. Example: "because of shiny awesomeness" as the reason why Stream:Temperature increases with Sunlight:Amount.	*(Case 02 / tape 235 Student pair of two males on first day of 2nd Weather model who were frequently off task and creating the only model to be scored as Poor. They are focusing more on amusing themselves than on model creation. In this instance, they are working with an object icon that looks like a tree or a shrub which provokes a discussion of a well known comedy movie, and movie dialog gets entered as the description.)* *(Case 04 / Tape 232 Student pair of two females on the second day of 1st Weather model, creating relationships for a model that would be scored as Good overall. In this case the discussion brings up a scientific misconception that gets entered as a description. This is a rare mistake for this pair.)* "Temperature affects humidity, it increases it, because it" "Wait temperature increases humidity by, yeah" "Because you need it to be, you need it to be, it's always warm when it is humid, it's never like cold humid." "All right. 'Cause when it is warm, does it, I don't care if it is warm or cold..." **"because if it's, you can't have humidity when it's cold."**

Table 3.4: Exemplar of Incorrect Articulation

TYPE	CODE DESCRIPTION	ACTUAL EXEMPLAR (case/tape) (typed articulation in bold)
Simple	Learners make a scientifically accurate but simple text entry, including the "because" statement. Example: "As Stream : Dissolved Oxygen increases then Fish : Population increases by more and more BECAUSE – fish need to breathe"	(Case 10 / Tape 200 Student pair of two males in second day of Decomposition model, creating a factor for a model that would be scored as Excellent. They are finishing up their factors and discuss, in simple but accurate terms, how to describe the factor "amount of decomposers". They offer a straightforward statement of what decomposers require to survive as the description.)
		"The rate, remember oxygen and water, the rate of decomposition affected by the amount of decomposers there's a period." **"Decomposers depend... on oxygen and water to survive"**

Table 3.5: Exemplar of Simple Articulation

TYPE	CODE DESCRIPTION	ACTUAL EXEMPLAR (case/tape) (typed articulation in bold)
Robust	Learners make a complete and scientifically accurate text entry, including the "because" statement. Example: "As Sun:Amount of sun increases then Air : Relative Humidity decreases by about the same BECAUSE – sun increases temperature and warmer air can hold more moisture"	*(Case 10 / Tape 191 Student pair of two males in first day of Decomposition model, making their first round of objects for a model that would be scored as Excellent. Even for these first simple objects, they discuss detailed descriptions and enter thorough and accurate sentence in the articulation box.)*
		"Decomposers break down Decomposers" **"Decomposers feed on the wastes of dead bodies of other living or once living materials"** "All right, this is a direct definition, it's pretty disturbing, but.."

Table 3.6: Exemplar of Robust Articulation

allows coded segments of all transcripts to be gathered and grouped or contrasted with Boolean operators such as "and", "or", and "nor". Thus a request to find all text coded for "Robust" (codes 2.1.1.1 and 2.1.2.1 and 2.1.3.1) provides a summary of all instances of Robust articulation across all transcripts, broken down by type. This can be repeated for codes Simple, Incorrect, and Ignore.

In order to answer the sub-question about changes over time, the same count can be run four separate times, on four subsets of the data, one for each curriculum exposure. Non-parametric statistical analyses are performed to determine the significance of any differences observed. For the question 1A, a Kruskal-Wallace H test, the non-parametric equivalent of a one-way ANOVA, is performed to detect any differences amongst the results for object, factor, and relationship. Then Mann-Whitney U tests are preformed between all possible combinations to detect specific relationships. These non-parametric assessments are required because the data are nominal and ordinal and parametric assumptions cannot be met. For question 1B, the matrix of articulation changes over four curriculum units is tested with a Friedman test, the equivalent of a repeat-measures ANOVA, because the ordinal data does not meet parametric assumptions.

3.5.3 Coding for Question 1B

Question 1B concerned the nature of dialog that occurs around the use of scaffolds in the tool, asking "While using tool scaffolds for articulation, what is the quality of discussions learners engage in, does this change over time, and if so, how?" These codes are to characterize the type and substance of learning conversations, the sense-making the learners engage in *as they are confronted by each type of scaffolding.* Essentially, how much did the learners discuss and was what they discussed scientifically accurate? The codes for both type and substance of learner dialog are justified below and defined in table 3.7.

Prior research has noted that learner responses often fall into three or four categories, and these codes draw on that research. Roschelle (1992) found that learner responses fell along a continuum from "continuation without response" to "simple acknowledgement" to "verbatim recitation" to "mutually acceptable elaboration", in order of desirability. As discussed in chapter 2, "the key process for maximizing and expanding cognitive resources was connected discourse... in which participants acknowledged, built, and elaborated upon others ideas" (Hogan, et al., 2000 p. 426). Hogan similarly noted three patterns of peer interaction, as she counted and analyzed conversational turns:

- CONSENSUAL - one speaker ran the discussion, the other agreed, passively accepted, and/or repeated back.
- RESPONSIVE - two or more respondents responsible for building discussion.
- ELABORATIVE - all speakers, multiple contributions, linking to or building on prior comments.

Additionally, Zhang (2003, p.194) found incidentally in his study of modeling practices that learner pairs seemed to have three types of collaboration: "good rapport/full communication", "good rapport/ one dominant - less communication", and "lack of rapport & low respect/communication". In terms of coding for substance, Guzdial (1994) also scored learner statements on 3 levels of complexity/sophistication while investigating the effects of an ILE.

Analysis of iterative turn taking was central to Hmelo's (2000) analysis of learners using a scaffolded computer tool. She counted the number of conversational turns, and coded them for types of collaborative interaction. The assessment of learner verbal interactions has long been relied upon in qualitative research generally and the learning sciences specifically (Fretz, et al., 2002; Miles & Huberman, 1994; Reiser, 2004; Roschelle, 1992; Tabak, 2004). Taking the software interactions and cursor gestures into account helps develop a more robust picture of learner thinking (Roschelle, 1992). Tabak (2004) uses transcribed dialog along with screen shots and contextual descriptions of software use to elucidate her theory of differentiated scaffolding.

The concept of dialog quality focused on the scientific content of the discussion, but the pattern of the discussion also matters. As noted in other research, including prior Model-It studies, conversations between members of a pair of learners using the tool fall generally into three types. The types of dialog increase in degree of involvement and number of conversational turns, and cover the spectrum from short discussion dominated by one learner to a longer discussion with both learners putting forth and/or critiquing ideas until consensus is reached.

Verbal interactions or discussions that occurred while learners were using the scaffolds were coded on two dimensions: Quality (in terms of scientific understanding shown) and Type (in terms of the type and amount of conversational turn-taking). Original coding definitions and actual exemplars from the data are shown in table 3.8 and table 3.9.

3.5.4 Data Analysis for Question 1B.

The queries in support of research 1-B, and sub-questions 1 and 2, involved the codes from section 2 and 4 of the coding scheme (appendix A). To answer the question about the frequency and type of dialog used by learners around the scaffolds, a count can be run for the overlap (co-occurrence) of the two dialog codes. For example, the query could ask how many times the robust quality dialog code occurred at the same time as the elaborative type dialog code (presented visually in figure 3.4). Since there were a possible nine combinations of the type and quality codes, this count would be run a total of nine times.

For sub-question b regarding changes over time, these nine counts would be run

Type	
Consensual	One learner drives discussion, the other does not respond, assents non-verbally, or acknowledges agreement with simple phrases or repeat-back. Example: " How about a Stream object, for Water Quality?" "Sure"
Responsive	Both learners discuss through at least four conversational turns. Minor elaboration of the original idea may occur. Example: "How about a Stream object, for Water Quality?" "Do we need one?" "Well, lake is not the whole system, and our measurements were in the stream." "OK, let's do it"
Elaborative	Both learners actively construct a collaborative understanding and response, through at least six conversational turns. Prior comments are built upon and elaborated with each turn. Example: "How about a Stream object, for Water Quality?" "Do we need one?" "Well, lake is not the whole system, and our measurements were in the stream." "Well, it is the same water, right?" "No, there can be stuff that enters the stream after it leaves the lake, and stream aeration can increase D.O, so..." "Right, right, and the stream is too small for fish" etc.
Undetermined	Learners engage in conversation that is not off-task, but cannot be coded in the categories above.
Off-Task	Learners engage in conversation that is not off-task, but cannot be coded in the categories above.
Substance	
Robust	The scientific content of the discussion is robust and correctly stated. This code will often, but not always, occur with 4.1.3. (e.g., learners could have an elaborative conversation about misunderstood science). Example: Learners have a four turn conversation that contains detailed rationale for using a bell curve relationship for how pH affects fish population.
Simple	The scientific content of the discussion is simple, but not incorrect. This code will often, but not always, occur with 4.1.2. (e.g., learners could have a longer elaborative conversation that never gets very specific). Example: Learners could have a seven turn conversation that only in the end arrives at a basic statement like "run-off increases turbidity".
Inaccurate	The scientific content of the discussion is wrong. Example: "So, increasing moisture increases the decomposition of plastic" "yeah, it's like rusting" "OK".

Table 3.7: Codes for Type and Substance of learner dialog

TYPE	CODE DESCRIPTION	ACTUAL EXEMPLAR (case/tape)
Consensual	One learner drives discussion, the other does not respond, assents non-verbally, or acknowledges agreement with simple phrases or repeat-back. Example: " How about a Stream object, for Water Quality?" "Sure"	"Alright, so my barometric pressure um the" "Force" "Force of air coming down on" "on the earth" (237) *(Case 04a / Tape 237 Student pair of two females on the first day of 2nd Weather model, creating relationships for a model that would be scored as Excellent overall. They are adding numerous objects and factors in starting their model. They are following a repeated pattern of finishing each other's sentences as they work.*
Responsive	Both learners discuss through at least four conversational turns. Minor elaboration of the original idea may occur. Example: "How about a Stream object, for Water Quality?" "Do we need one?" "Well, lake is not the whole system, and our measurements were in the stream." "OK, let's do it"	"Now what do we do? Do anything else affect humans?" "Well, humans affect animals" "How? They kick them out of their habitat" "Yeah, Animals they decrease" "Decrease, Humans decrease animals" "By a lot or a little? Less? More and more?' "More and more."(139) *(Case 17 / Tape 139 Student pair of male and female on the second day of Water Quality 1 model, creating relationships for a model that would be scored as Average overall. They are looking at their model icons and creating several relationships between factors created on day one, interspersed with off-task conversation.)*
Elaborative	Both learners actively construct a collaborative understanding and response, through at least six conversational turns. Prior comments are built upon and elaborated with each turn. Example: "How about a Stream object, for Water Quality?" "Do we need one?" "Well, lake is not the whole system, and our measurements were in the stream." "Well, it is the same water, right?" "No, there can be stuff that enters the stream after it leaves the lake, and stream aeration can increase D.O, so..." "Right, right, and the stream is too small for fish" etc.	"Number of cars" "So, number of cars would affect the conductivity because of the salt" "Um Hm. Affect the conductivity and the pH." "Yeah" "So wait a second" "Stream conductivity increases by" "Because?" "Conductivity increases, yeah, increases" "Decreases? Increases. By a little, how's that? Oh maybe like, well, wouldn't it be like a lot cause of the salt and the stuff that could be leaking? Yeah, but it's kind of far away, isn't it? Yeah, that's true, OK, you're right. So, a little because um, salt Because Salt that is put on the road equals higher conductivity. *(Case 16 / Tape 140 Student pair of male and female on the second day of 1st Weather model, creating relationships for a model that would be scored as Average overall. In this case the discussion centers on how cars might affect stream conductivity since they have learned that road salt runoff can affect conductivity in streams.*

Table 3.8: Exemplars of Dialog Type

TYPE	CODE DESCRIPTION	ACTUAL EXEMPLAR (case/tape)
Robust	The scientific content of the discussion is robust and correctly stated. This code will often, but not always, occur with 4.1.3. (e.g., learners could have an elaborative conversation about misunderstood science). Example: Learners have a four turn conversation that contains detailed rationale for using a bell curve relationship for how pH affects fish population.	““It would be bell shaped because all the, cause they have excess plant growth and then there's too many plants so they can't like live and stuff and then this, so “OK, so excess plant growth” “At the start there” “Then it'll happen all over again” “Because like at the beginning they all go up like if there's excess plant growth, so there's a lot of plants growing. But then when there's too many plants like they all, there's, they take up too much oxygen, use all the oxygen so they can't breathe so they just die off.” *(Case 15 / Tape 189 Student pair of two males in the fourth day of Water Quality 2 model, making for a model that would be scored as Good. They are discussing the relationships between stream factors and plant growth during eutrophication as they make final adjustments to their model prior to presentation, they have received some earlier scaffolding from the teacher.)*
Simple	The scientific content of the discussion is simple, but not incorrect. This code will often, but not always, occur with 4.1.2. (e.g., learners could have a longer elaborative conversation that never gets very specific). Example: Learners could have a seven turn conversation that only in the end arrives at a basic statement like “run-off increases turbidity”.	“Variable, how hot?” “Description, it gets very hot when the sun is shining.” “It gets” “It gets very often.” “Okay? Okay.” (136) *(Case 19 / Tape 136 Student pair of male and female in the first day of Water Quality 1 model, making for a model that would be scored as Average. They are creating factors for the object sun as part of building up the pieces of their model.*
Inaccurate	The scientific content of the discussion is wrong. Example: “So, increasing moisture increases the decomposition of plastic” “yeah, it's like rusting” “OK”.	“Because you need it to be, you need it to be, it's always warm when it is humid, it's never like cold humid.” “All right. “ “Cause when it's warm, does it, I don't care if it's warm or cold.” “Because if it's, you can't have humidity when it's cold.“ *(Case 04 / Tape 232 Student pair of two females on the second day of 1st Weather model, creating relationships for a model that would be scored as Good overall. This dialog is part of a discussion that reveals scientific misconceptions that do not get corrected by either student, and the misconception in the dialog is entered into the articulation scaffold as well.*

Table 3.9: Exemplars of Dialog Quality

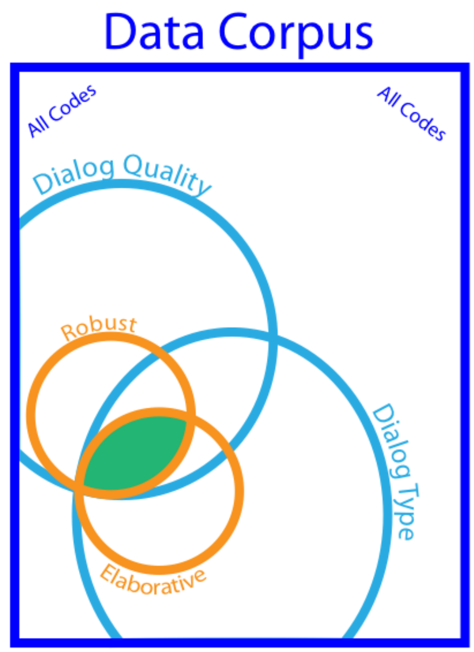

Figure 3.4: Showing visual structure of a query for the co-occurrence of Robust and Elaborative dialog codes.

on four separate sub-sets of the data, segregated by curriculum unit. Non-parametric statistical analyses are performed to determine the significance of any differences observed. For the co-occurrence of dialog type and quality, a chi-square test is used to assess if there is a significant relationship between the variables. The chi-square test is required because the data are ordinal and parametric assumptions cannot be met. For the matrix of changes in dialog over time across four curriculum units a Friedman test, the non-parametric equivalent of a repeated measures ANOVA, is used because the data are ordinal and parametric assumptions cannot be met.

3.5.5 Coding for Interpersonal Scaffolding

As discussed in chapter 2, interpersonal sources of scaffolding in classroom environments are inextricable elements of the learner process of using a scaffolded tool. Scaffolding behaviors from peers and from teachers both can provide redundant or synergistic effects. Because these sources of scaffolding are likely to affect how the learners progress, and are of general interest to this investigation of scaffolding, they are assessed using the five codes in table 3.10, drawn directly from the summarized research on interpersonal scaffolding in chapter 2 (these codes are also found in the coding scheme, appendix A). Each instance of learner dialog with a teacher or researcher is coded and this data is used to inform discussions of learner model creation and as part of the master scaffolding score which is used to answer research question 2C, more fully detailed below.

3.6 Coding for Research Question Two

For question two, "what is the quality of learners' models and how does that quality change over time?" the primary data source analyzed was the model artifact produced by the student pairs. This digital file was opened with the relevant version of Model-It and the individual components recorded as JPG graphic files and the specifics of each component (what text was written, what values were entered) were recorded in a text document for later analysis and coding. By scoring each model in detail, the data was reduced to numerical scores/codes which were used to investigate the research questions on model quality.

3.6.1 Coding for Question 2A

The scoring of written artifacts, such as concepts maps or models also has a long history beginning with Novak and Gowin (1984) and continuing to the present day (Derbentseva, et al., 2007; Novak, 1999). Model-It models have been scored and evaluated by several researchers (Jackson, 1999; Stratford, 1996) and their work informs the scoring rubric. Scoring of science learning artifacts for quality is not limited to models or concept maps. In their examination of distributed scaffolding, Puntambakar and Kolodner (2005) assessed 7 aspects of learner dialog and journal

Change Task	Teacher helps learners select task, modifies task for learners by reducing degrees of freedom, prompts/helps learners maintain direction on task. Example: Teacher specifically requires first time users of the tool to create ONLY A FEW Objects and Factors on their first model and then proceed to test mode.
Model Behaviors	Teacher demonstrates or models the desired action or process, or invites learners to demonstrate for teacher. Example: Teacher asks one pair to show their model to another as she watches.
Enhance Motivation	Teacher attempts to develop shared interest and goal, makes explicit attempts to control frustration or risk, provides subtle constructive criticism or praise, gives encouragement in support of goal. Example: After review of model and testing, teacher says, "This is good work so far! Your testing showed that all of your relationships work as you would expect, but you have only one independent variable. What do you think about that?"
Share Knowledge	Teacher provides content or task knowledge using Socratic questioning techniques or tailored assistance. Explanations may be offered using metaphor or PCK. Learner understanding is verified. Example: After observing a pair having trouble working with the pH variable, and trying to use two separate factors for "acid" and "alkaline", the teacher says, "You know, the pH scale is a unitary scale from 1 to 14, so is there a way to capture a relationship that runs from high to low, but the peak value is in the middle?"
Share Metacognition	Teacher provides guidance on thinking or practice. Encourages reflection on prior work and/or problems that are occurring. Assists learners in making generalizations about their thinking or practice. Example: Teacher suggests that a team with dozens of factors but who has not yet moved to testing might want to focus their model and try testing a smaller subset of their ideas first, or a teacher who asks a learner pair to reflect on what they observed in their stream walk when they are confused about a proposed relationship.

Table 3.10: Codes for Interpersonal Scaffolding

Research Question	**Data Source(s) Used**	**Coding Scheme Used**	**Hypothesis (I predict that...)**
2A	Model Files	Model Evaluation Rubric, appendix D	As a group, the models will show a positive relationship between the amount and quality of scaffold use and the quality of the final model.
2B	Model Files	Model Evaluation Rubric, appendix D	As experience with the software and modeling tasks increases, the quality of scaffold use and model artifact will also increase.
2C	Model Files & Transcript Database	Model Evaluation Rubric, appendix D Coding Scheme (Scaffolding)	There will be a relationship between the learner pairs' degree and quality of scaffold use and the quality of their final model.

Table 3.11: A breakdown of research questions, data sources, and hypotheses.

entries using a four point scale.

Scoring of models is informed by research on the conceptually similar (though static) concept maps (Novak & Gowin, 1984). Concept maps consist of propositions, which are two concepts and their relationship. As theories about concept maps have advanced, a new emphasis on Cyclical concept maps has emerged (Safayeni, et al., 2005), with the key additions to cyclical maps being the structure (cyclical versus hierarchical) and the inclusion of rate or effect information in relationship links (which moves this type of concept map that much closer in function to a Model-It model).

3.6.2 Justification for Scoring Rubric

Model-It models, and a similar form of artifact the concept map, have been scored in a variety of ways. Novak and Gowin (1984) addressed the scoring of concept maps in their seminal work on concept mapping. The central tenets of their evaluation scheme remain valid today. Table 3.12 summarizes a range of research on the evaluation of models or concept maps that is discussed below.

In early research on concept maps, Novak (1984) focused on enumerating core components, specifically the number of propositions and the links between them. Additionally he assessed the degree of cross linking and the number hierarchical levels, if present. The validity and specificity of the propositions and links were also considered in the scoring. In early research on assessing dynamic computer models Jackson (1994) rated models on a four point rubric in three areas: Accuracy, Complexity, and Completeness. Rating was done by raters who knew the instructional task associated with the created model, and they applied a holistic standard for completeness that was based on their knowledge of the curriculum goals. Greater numbers of accurate components and relationships rated higher scores. In a literature review of 21 studies that evaluated concept maps, Ruiz-Primo and Shavelson (1996) found the most common forms of assessing concept map quality included counting nodes and links, as well as levels of hierarchy, and the use of an "expert" concept map for comparison. These aspects were used in over 50% of the studies reviewed. The accuracy or scientific validity of nodes and links, and the amount and quality of any associated description were also assessed in many studies. In studies examining Model-It models, Stratford (1996) and Singer and colleagues (2000) counted components (e.g., factors, relationships) and assessed their accuracy, as well as the quality of the explanations embedded in them. Stratford went into greatest detail, assessing the overall "craft" of the model in terms of the number of incorrect components or sloppy construction, as well as developing a set of 7 basic model structures with which to categorize learner models. In later research, Singer and colleagues counted total numbers of factors and relationships versus the number of correct factors and relationships, while also categorizing factors and relationships that were inappropriate for the model, duplicate, irrelevant or off task, or had errors in scientific accuracy (i.e.

incorrect degree or direction of relationship). In more recent studies of concept maps, the was continued focus on counting the number of components and their accuracy as the core scoring variables, with additional points awarded for increased complexity of structure and numbers of interlinks between components (Derbentseva, et al., 2007; Yin, et al., 2005).

As can be seen in the research summarized in table 3.12 the original concepts of structure and counting of components and propositions remain of consistent interest [in Model-It terms, Objects and Factors would be components, and Relationships would be propositions]. Structure was assessed in five of the studies reviewed, and the tendency was to assess levels of hierarchy, and in one case (Stratford, 1996) the actual type of structure. Enumeration of components was used in almost all of the studies, but what was counted varied. The numbers of concepts, items, or links were the most common items counted. A special case of this counting is for links, where an item could be linked multiple times with multiple other items. Accuracy was a complicated assessment to make, and while almost all of the studies accounted for it in the overall model or the components, only two (McClure, et al., 1999, Sins, et al., 2005) used a reference or ideal model (criterion map) to score against. In most cases, the scientific accuracy of the particular item or link was what was assessed, as in this study. Relevance is sometimes (Singer, et al., 2002; Sins, et al., 2005; Stratford, 1996) considered separately from accuracy, particularly in studies where accuracy is more granular and assessed at the component level. In that case, then an item could accurate but have nothing to do with the model's other components, and thus be reflective of poor modeling practice. Lastly, in themes related closely to accuracy in an overall sense, some studies used criterion maps or a holistic judgment as to the quality of the overall model (McClure, 1999; Sins, et al., 2005). Table 3.13 notes the central themes that cut across the various assessment strategies. How these themes are accounted for in model scoring for this study is detailed in the next section.

Researcher / Year	Artifact	Scoring Method	Notes
Novak (1984)	Concept Maps	1xPropositions (if valid), 5xHeirarchy (if increasing in specificity), 10xCrosslinks (if meaningful) Specific examples (if valid) +1 pt.	Original development of Concept Maps
Jackson, et al., (1994)	Models	Rated 1-4 (poor – excellent) for: Accuracy, Complexity, and Completeness. Poor would be for few or no accurate relationships, Excellent was for an error free model.	Did not give standard for complexity. Used a holistic rater standard of completeness based on instructional task.
Ruiz-Primo & Shalvelson (1996)	Concept Maps	21 prior studies reviewed, out of 21 which used: Use of criterion/expert map = 9 Accuracy of Nodes and/or links = 7 Counting of Nodes = 6 Counting of Links = 10 Amount or quality of description = 3 Holistic judgment = 2 Levels of hierarchy/size = 11 Counting of cross-links = 6	Meta-Analysis
Stratford (1996)	Model	Craft (skill at doing it) – incorrect or sloppy - Conceptual (content) - Explanations (accuracy, depth [causal, corr, restate], integrity [rich/high/mid/fair.poor] - Structure (coherence [unified, partial fragmented, fragmented], form [7 basic, + combos] - Factors (accuracy [arbitrary, default, realistic, no criteria], breadth [fair, complete, over, unknown] * as compared to guide scenarios. - Relationships (accuracy [non-causal, backwards, nonexistent, incorrect], depth (# steps from indep to dep, both shortest and longest], integrity [ratio of factors to relationships, and # of relationships] Behavioral (function) - Fidelity (to reality [low-med-high]) - Over time (constant / straightforward complex / sophisticated complex)	50 Pairs ** had a type of relationship (delayed rate) not found in this version of Model-It - Objects not assessed.
McClure, et al., (1999)	Concept Maps	Varied by condition. Half used master map, half did not. Holistic condition used rater judgment. Relational used Number and quality of propositions. Structural used Number and quality of propositions AND bonus points for higher level structures (interlinks and extra levels)	Comparison of six Concept Map scoring methods. Use of master map lowered reliability of scoring in all cases. Relational with master map provided best results for teacher use.
Singer, et al., (2002)	Models	Total factors, valid factors, Total relationships, valid relationships. Also looked for: inappropriate objects, duplicates, irrelevant/off task object, directional relationship errors, effect relationship errors, and irrelevant/off task relationship errors.	Found most models had at least one error.
Yin, et al., (2005)	Concept Maps	Number of propositions, accuracy of propositions (total and individual). Additional variable for structure, better scores for increased complexity and interlinks	
Sins, et al., (2005)	Models	5 point scales for conceptual validity as compared to idealized model, and 5 point scale for structure, with points taken off for invalid items and links	
Derbenstseva, et al., (2007)	Cyclical Concept Maps	Number of quantified concepts, Number of dynamically specified propositions. Accuracy of specification increased score for propositions.	Also divided into structure: Tree, Cross-Link, and Cyclical

Table 3.12: A summary of Model and Concept Map scoring systems

ITEM	DESCRIPTION	SCORING METHOD	IN THIS STUDY
Structure	The visible design, e.g., a "tree" hierarchy, or a ring, or a web.	More levels of hierarchy are better. More complex inter-linked structures are better. A model with a 5-level tree would be better than one with a 3-level tree.	Yes, while the specific shape or structure is not assessed or categorized, as program does not preserve screen presentation of model, the components can be examined for complexity of structure. (a model with one central factor linked out to four others would be a simple "star", whereas the same five components each multiply interlinked would be more complex) [FUNCTION]
Number (of X)	A simple count of the number of components, object, or factors (e.g., "stream", "trees", "sun"), or relationships/propositions (e.g., "as sun increases trees increase")	More is better, provided they are relevant (see relevance). A model with a total of 30 relevant components would be better than one with only 15.	Yes, the number of relationships is assessed. [FUNCTION]
Interlinks	Instances where there are more than one-to-one links between objects, and/or relationships that cross between hierarchical structures.	More is better, provided they are relevant (see relevance). A model with a total of 20 relevant links between components would be better than one with only 10.	Yes, the number of relationships is assessed. [FUNCTION]
Accuracy	Is the item accurate in its name, description, and defining characteristics (e.g., factors "fish" and "dissolved oxygen" could be accurate, but a relationship that showed fish increasing when dissolved oxygen decreased would be scientifically inaccurate.	Generally presumes a standard of accuracy per curriculum delivered as well as a general scientific plausibility.	Yes, the accuracy of the description for each component (object/factor/relationship) is assessed. [ACCURACY]
Relevance	Does the Object, Factor, Relationship/Proposition belong in the map or model. (e.g., a factor of "number of bees" would be irrelevant in a model of stream water quality.	Presumes a vague (or specific) standard of what is expected for a given artifact. Often used standard of "scientifically correct".	Yes, the appropriateness of each component (object/factor/relationship) is assessed, with missing, illogical, or extraneous components reducing the score. [COMPLETENESS]
Criterion Map	Comparison to a standard CM or model that reflects the desired content	Generally scored for each aspect that matches the master/ expert model, with points deducted for missing items, and no extra points for possibly valid items that don't show up on the master model.	No, there is no single standard model for the curriculum areas related to each model. Learners might choose to focus on different aspects of weather, or example.
Holistic Quality	Subjective "is this model good"	Presumes evaluator has mastery of the content that is being modeled.	Yes, each model is evaluated or fidelity in terms of how well it replicates what it intends to model. [FUNCTION]

Table 3.13: Common Scoring Criteria for Maps or Models

3.6.3 Model Scoring

In developing the scoring rubric for Model-It models, the above research was taken into consideration. The three main areas of the rubric are Accuracy, Completeness, and Function. Accuracy (in the rubric) captures the "number" and "accuracy" aspects in table 3.13, in that more is better, but the components also have to be scientifically correct and well described. Completeness captures the "relevance" aspect of table 3.13, in that missing, illogical, or extraneous components are penalized. Function captures the "structure", "interlinks" and "holistic quality" aspects of table 3.13, in that structural complexity and the fidelity of function of the model are carefully examined.

A very similar version of this study's rubric was used once before (Fretz, et al., 2002a), and was successful in scoring a small set of models from middle school science classes. A summary of how the models are scored in terms of Accuracy, Completeness and Function follows, and further information can be found on the scoring rubric itself in appendix D.

3.6.3.1 Accuracy.

Accuracy is assessed by examining all Objects, Factors, and Relationships and scoring each one on the scientific quality of the information articulated in the item. Blank or one word descriptions get a low score (i.e., an object "Bugs" has a text description with just one word, "bugs"). For increasing detail more points are awarded, with the full four possible points going to descriptions that are both correct and unusually detailed and specific. For factors the same rules for objects apply, as well as assessing the range specified for the factor. (i.e., a factor labeled "number of worms" that has a range of "good-better-best" would score poorly). Relationships have the most specific description options, with the requirement for a "because statement". In addition to the rules above for objects (that the description should be filled in, and with a full and relevant explanation), relationships are also assessed or the direction and degree specified for the relationship. Thus, a relationship with a good description of how increased sunlight causes increased temperatures, that had a specified relationship showing “as sunlight increases, temperature will decrease by a little”, would score only average because the relationship has an incorrect direction.

3.6.3.2 Completeness.

Completeness is assessed for objects, factors and relationships. An assessment is made for components that are missing, illogical, or extraneous. For objects, factors, and relationships the same four point scale is applied, and maximum points are awarded for having no missing, illogical, or extraneous components. An example of a missing object would be for a model of lake water quality, there is no object for “water”. An example of an illogical factor would be in a model of creek water quality there is an object creek and a factor of “Creek – amount of candy”. An example of an extraneous relationship would be in a model of water quality there is a relationship of acid rain to car roof damage.

3.6.3.3 Function.

Function is assessed by actually running the model in test mode and manipulating all independent variables. The models are assessed for Fidelity, Structure Complexity, and Overall Quality. Fidelity captures the overall function of the model as compared to the structure and behavior of the system it was designed to represent. The model is run in test mode, and each independent variable is manipulated to examine model function. Thus, a model of a stream ecosystem that contains major errors like sewage output affecting acid rain, and reversed relationships of acidity to fish population would score poorly on fidelity. A model with very realistic components and function with only minor discrepancies would score the full 3 points.

Structure complexity is also assessed with function score. A sophisticated and complex model will have both a larger number of components and also a greater number of relationships. Higher numbers of factors per object and longer chains of relationships are also better. Finally, the holistic category of overall quality captures all the aspects of the model that might not be assessed elsewhere in the rubric. If the model just barely meets the assigned class criteria, and is contains no major errors, but has mostly weak descriptions and seems expedient, it would be of only medium quality. However, an extremely well crafted but parsimonious model would gain points here, for the attention to detail. For example, a model may have only 5 factors, but each one might have custom numerical ranges and each factor is linked with multiple, complex, accurate relationships, like a bell curve for pH.

3.6.4 Data Analysis for Research Question Two

The queries in support of research question 2, and sub-questions a, b, and c involves the codes from section 5 of the coding scheme (appendix A). The analysis software facilitates the coding of the model components, but the master scores are relatively straightforward to gather and present. For the question on the quality of the models, a simple distribution of the final model scores can display the range and frequency of scores. For the question about changes over time the models scores would be broken out into four groups according to curriculum exposure, and then presented in the same manner as question 2A. For question 2C, the master model score is modified as detailed below and then compared with a master scaffolding score, also detailed below. Parametric and non-parametric statistical analyses are performed to assess the significance of any relationship.

3.6.4.1 Scaffolding Use versus Model Quality.

Comparison of scaffolding use with model quality for each pair can provide an interesting, but limited, view of possible benefits associated with effective scaffold use. To facilitate this comparison, a master score for scaffold use are created for each learner pair, and this is compared to a modified master model score. The model score has to be modified, because part of the model quality score directly assesses the quality of the articulations typed into the model's articulation boxes, which is the same content assessed for certain scaffold codes. Unless these concepts are broken out of the model score, there is a methodological short-circuit that guarantees one will be strongly associated with the other, because they are looking at similar things. However, model quality was based on several factors, so removing one factor will not make the quality score useless.

The master scaffolding score is determined by collapsing the overall counts of tool and teacher scaffold use, and the % of robust articulations, for each pair. Based on typical modeling tasks, an average number of articulations is 10-15, with 30-40% of that number being robust, and learners engage in 3-4 instances of interpersonal scaffolding. Using these baselines allows the generation of a rubric that captures the range of models in this study. Of the 20 points in the master scaffolding score, ten come from the total amount of articulation, five from the percentage of that articulation that was robust, and the final five points come from the amount of interpersonal scaffolding observed with that pair. Thus, a pair that had over 20 articulations, with 80% of them being robust, and having 9 or more instances of interpersonal scaffolding across the multi-day model creation task, would score a perfect 20 points. Table 3.14 shows how the three areas are scored to yield a master scaffolding score for each pair.

The modified model quality scores are then plotted with their associated scaffolding score for each pair and parametric and non-parametric post-hoc analyses (Pearson and Spearman correlations) are used to determine significant relationship, if any. Be-

Total Articulation (10 pts)	% Robust (5 pts)	Total Interpersonal (5 pts)	Master Score
< 10 = 2.5	< 20% = 1	1 or 2 = 1	Out of 20, sum of three scores to the left.
10 to 15 = 5	20-39% = 2	3 or 4 = 2	
16 to 20 = 7.5	40-59% = 3	5 or 6 = 3	
> 20 = 10	60-79% = 4	7 or 8 = 4	
	>80& = 5	9 or more = 5	

Table 3.14: Breakdown of master scaffolding score system

cause in this analysis all data are interval and the sample size is larger, the Pearson correlation is the stronger test, but is weak at detecting relationships that are not linear. The weaker non-parametric equivalent Spearman correlation is performed as well, and the two results together will inform question 2C.

This chapter has laid out the methods for this study. A total of 46 learners are examined, in pairs, as they use dynamic modeling software during four curriculum units over two school years. Process video tapes, which combine computer screen video with learner discussion audio, are made of each model creation session. Transcriptions of these model creation sessions, and scored written summaries of the models created, are coded with the assistance of a qualitative analysis software program. Detailed review and counts of coded articulations and dialog form the basis for answering the research questions of this study. Where applicable, non-parametric statistical analyses are performed to assess the significance of differences found. In the next chapter, these results are presented.

CHAPTER IV

Findings

4.1 Introduction

In this chapter I discuss the findings of the analyses of the data corpus as described in chapter 3. My data analyses address two specific questions regarding scaffold use and model quality. I discuss the findings as they relate to each research question and sub-question. A variety of textual and graphical representations of raw and summarized data are provided in support of assertions made. Excerpts from transcribed discussion and contextual information are provided to ground examples in complex classroom environment. Screen shots and example content from student models are provided. Finally, I summarize the findings and assertions in preparation for a discussion of implications in chapter 5.

4.2 Research Question One

My first research question asks - *What types of student and teacher interactions occur with and around scaffolds in a dynamic modeling tool, and do they change over time?*

This question focused on the idea of assessing both the use of scaffolds in the tool (e.g., what did they express, in written or verbal form, as a result of scaffolds designed to get them to type in, or articulate, their understanding of a model component), and assessing the articulation around the tool (e.g., the sort of conversations that occurred because of the scaffold, as learners constructed a mutually agreeable representation of their understanding to enter into the program). This leads to two sub-questions, each with two parts:

- Do tool scaffolds for articulation lead to learners articulating their thinking in writing, does this change over time, and if so, how?
- While using tool scaffolds for articulation, what is the quality of discussions learners engage in, does this change over time, and if so, how?

4.2.1 Frequency of Articulation Types

As discussed in chapter 3, the responses to scaffolds in the articulation boxes (in Object, Factor, and Relationship windows) would be coded for four types of articulation. Objects are the initial model components created, such as "sun", "stream" or "fish". Factors are the various measurable qualities of each object such as "sun-hours per day" or "stream – temperature" or "fish – number". Objects and Factors have articulation boxes that are blank and in which the learners are to enter written descriptions. For example, the factor "stream – temperature" could have a description of "this is the temperature of the stream behind our school, measured in degrees Fahrenheit". Counts of each type of articulation were gathered for each model component. The overall totals of each type of articulation across all cases is shown below in Table 4.1 and Figure 4.1 It was relatively rare for articulation scaffolds to be ignored or used incorrectly (between 1-10% of the time, depending on the scaffold), however, the most common score for articulation was at the simple level, occurring in 72% of the instances of scaffold use. Robust articulation was much less common, representing only 21% of the total responses.

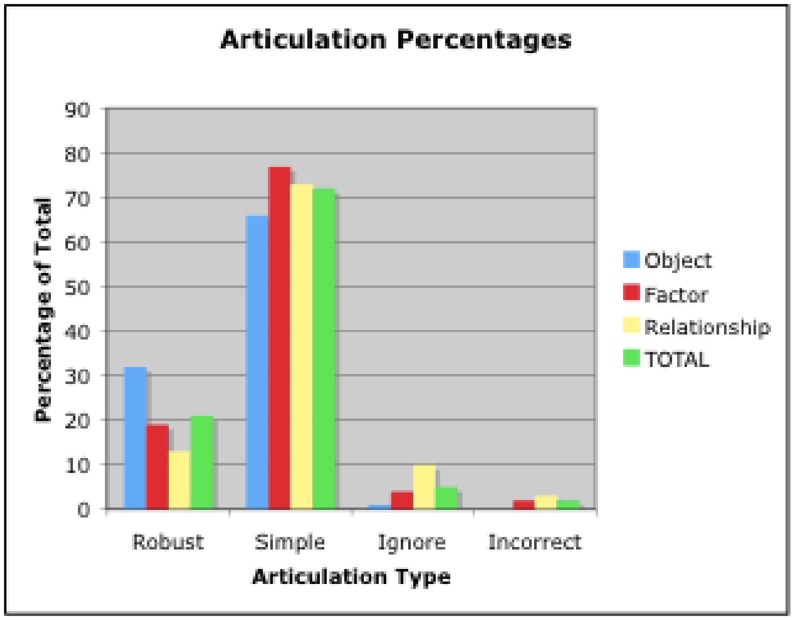

Figure 4.1: Articulation scores for the three types, by percentage.

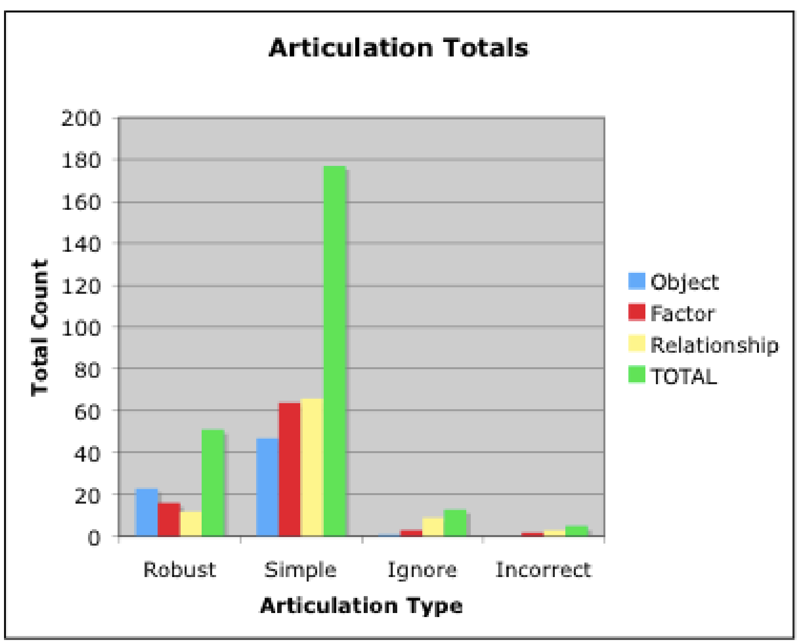

Figure 4.2: Articulation scores for the three types, by total count.

	Total Robust **Count** [# sources] % of total	Total Simple **Count** [# sources] % of total	Total Ignored **Count** [# sources] % of total	Total Incorrect **Count** [# sources] % of total
Object	**23** [12] 32%	**47** [23] 66%	**1** [1] 1%	**0** [0] 0%
Factor	**16** [7] 19%	**64** [30] 77%	**3** [2] 4%	**2** [2] 2%
Relationship	**12** [7] 13%	**66** [27] 73%	**9** [3] 10%	**3** [3] 3%
Total	**51** [26] 21%	**177** [80] 72%	**13** [6] 5%	**5** [5] 2%

Table 4.1: Articulation Summary Chart ("# sources" is the number of unique source documents that these codes are found)

The total of articulations scored as "simple" was four times greater than those scored as "robust" and "simple" articulations were the most common for all types (object, factor and relationship). As seen both in figures 4.1 and 4.2 and Table 4.1, "robust" articulation was most often seen in objects, with a slightly lower percentage for factors, and the lowest percentage for relationships. This drop in percentages may reflect the increasing complexity of the task (as learners progress from describing objects to describing factors to describing relationships between them), since describing an object (the initial building block of the model, such as "stream") in a robust and detailed way is easier than providing a robust articulation of a relationship (e.g., why "stream: nutrients." increases "stream plants: amount" only to a certain point due to eutrophication). The more desirable robust articulation was seen often enough to show it is clearly possible for learners to articulate this way in the scaffold, but it was not the modal response. Even so, it is important to note that the articulation scaffolds as a whole were used, in simple or robust ways, over 93% of the time. The count data from Table 4.1 were reformatted with Object/Factor/Relationship as nominal data and Incorrect/Ignored/Simple/Robust as ordinal data and a Kruskal Wallace test found the differences in the counts to be significant (2, 246) = 14.13, p = .001. This indicates that the differences in the amounts of the four types of articulations between the three groups are larger that would be expected by chance, and that further investigation is called for to refine the differences between the three groups. Mann-Whitney tests between the three nominal groups found differences between Object and Factor, (156) = 2521.50, p = .025 and between Object and Relationship (161) = 2335.50, $p < .001$ to be significant, but differences between Factor and Relationship were not significant : (175) = 3400.50, p = .098. The types of articulations for Objects were more often robust, whereas the counts for Factors and Relationships were similar, so this analysis confirms what is noticeable in Table 4.1, specifically that robust articulation occurred more often for Objects.

4.2.2 Issues Related to Lack of Robust Articulation

Analysis revealed a number of issues that may potentially explain the lack of robust articulation: skipped articulation boxes, learners' misunderstanding of the task, and learners hurrying through the modeling task. First of all, there were cases where a learner is using the software alone (either actually alone, or functionally alone with a partner who is not on task), and in these cases, particularly when there was no dialog about what was being created, the step of typing into the articulation box was skipped. In case one case (case 7), during 7th grade Decomposition unit, a learner who is controlling the software creates objects quickly and without articulation. After labeling the object he skipped ahead immediately to create a related factor and put description in the factor. The model from this pair has blank articulation boxes in 100% of their objects, but they did fill in every other articulation box (factors and relationships). These articulations are of mixed quality, however, and many of this pair's model components score poorly (with a final score of "average"), so this habit of skipping object descriptions might have been part of a generally less engaged modeling effort.

Secondly, and similarly, if the pair was in a hurry, (such as trying to design an entire new model in a single class period because they changed their plan), then the pair would often skip the detail work of typing in descriptions while focusing on creating objects and factors and linking them quickly to make the model functional in a short period of time. For example, on tape 242 in a final weather model, the learner pair was adding several new objects like relative humidity, and one of the pair was anxious to move on to building relationships, and thus tried to get the other to skip the articulation, as one asked for help with the description, *"Come on, what should I say?"* and gets a reply of, *"Oh, who cares, just leave it, it doesn't matter."*

Again with this pair, near the end of the class period the two learners have added a relationship as they finalize their model before doing demonstrations of their models to the class. The pair was looking at the "because" statement (the part of the articulation scaffold that provides a partial sentence for them to complete), and one of the learners wanted to finish the model and stop making modifications, while the other wanted to work until demonstrations start. In this sort of situation, articulations are often rushed or skipped:

> *S1 Are the, the amount of clouds, no, the amount of clouds decrease and the weather gets nicer*
>
> *S2 No, because we are doing clouds*
>
> *S1 So, you say weather or..*
>
> *S2 NO, just do decreasing, get out of here you don't have, like, any time at all!*

Thirdly, learners in several cases had one or both members of the team under the impression that articulation was not required. To have learners that remain unsure about this requirement is almost shocking, given the number of times teachers mention it to the class and to individual pairs. For example, in tape 229 as a learner pair discusses an object for Barometric Pressure for their weather model, they have the object description window open, and one has asked how to spell barometric, to which the other replies:

How am I supposed to know? You don't have to write the definition of it though. You don't write the definition do you?

In this case, the learners do eventually decide to put in a definition, however it is a "relationship" description of how X affects Y, as opposed to a scientific description of barometric pressure and how it is measured. Again in tape 235, as a pair creates several factors quickly for their 8th grade weather model, they repeatedly dismiss the articulation boxes without entering any text. First they offer:

We don't need a description at all. We just need the variable. What's the variable?

And then later they again, while the learner controlling the mouse dismisses articulation boxes after only entering the name, state:

We don't need descriptions for our variables

This pair may have also been feeling time pressure, as discussed above, because prior to these comments, they stated *"we really have to get to work now"*, and thus might have been making a conscious decision to ignore teacher guidance in support of their goal to get the components of the model in place quickly, even if not fully described.

Even when the scaffold is used, misunderstandings can reduce the chance of getting a robust articulation. Often, learners would struggle with how to describe objects (a seemingly simple task) and then default to putting in simplistic statements. Such as in tape 157, where the learners end up with an object of "plants" and a one word description of "green":

> *S1 So, plants, objects. Plants, type. I don't know, um, description, well, how.. what's a description? How do you make a description of a plant? You know how it says, like, here?*
>
> *S2 Green?*
>
> *S1 What do you write?*

And again in tape 225 in a weather model, a similarly simplistic description of "warm" for an object:

S1 Sun?

S1 I never know what to write here!

S2 Like, warm

Additionally, when making descriptions of objects and factors where details about the object or factor would be appropriate (e.g., D.O. – "dissolved oxygen is the amount of oxygen in the water available to plant and animal life and is measured in ppm"), the learners will instead put relationship information such as statements of how increased D.O. increases fish population. In tape 229 in a weather model, learners propose a description of temperature that keys on how air pressure relates to it:

You write how it affects that, so like, in this case you take Does air pressure affect temperature?

This description gets scored as simple because, while correct, it addresses only one of many possible relationships and does not actually fully describe the object or factor. A robust for a factor air:temperature would, for example, be: "this is the temperature of the air in degrees Fahrenheit".

In summary, the answer to "Do tool scaffolds for articulation lead to learners articulating their thinking in writing, does this change over time, and if so, how¿' is yes but not particularly well. Learners do articulate their thinking while using the articulation scaffold correctly over 90% of the time the scaffold is present. Their articulations are predominantly simple, with robust articulations occur more often in scaffolds where the task demands less (e.g., describing an object like "stream" vs describing a relationship like "amount of fungus to decomposition").

Next I discuss the changes in articulation over time and across the four curricula.

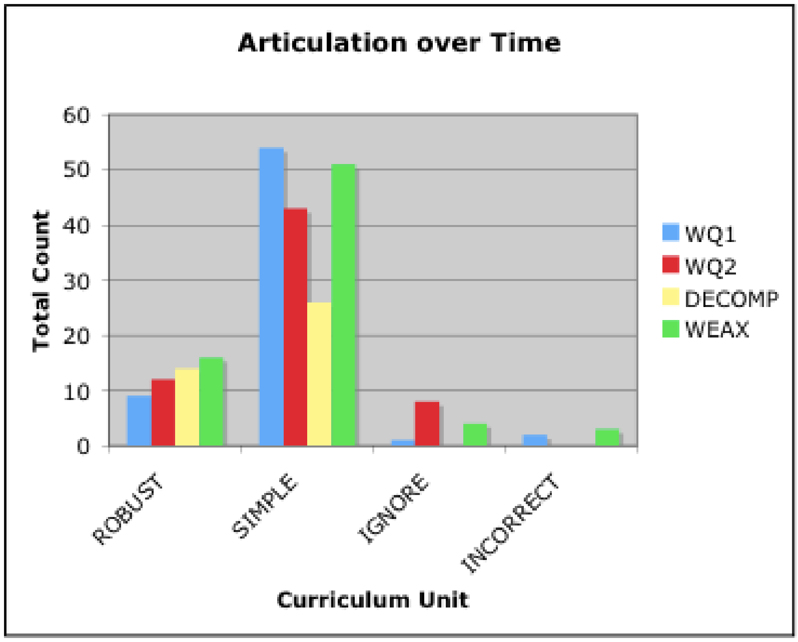

Figure 4.3: Articulation by type (total count) through the four curriculum units.

4.2.3 Changes in Articulation Over Time

Examining articulation within and across the four curriculum exposures, we can see changes over time. Three curricula occurred in 7th grade (WQ 1 & 2, Decomposition) and one more unit at the start of 8th grade (Weather). A detailed breakdown of the frequency of Robust, Simple, Ignored, and Incorrect articulations in each curriculum is presented in Table 4.2 and Figure 4.3.

	Total Robust **Count** [# sources] % of total	Total Simple **Count** [# sources] % of total	Total Ignored **Count** [# sources] % of total	Total Incorrect **Count** [# sources] % of total
WQ1	**9** [3] 14%	**54** [14] 82%	**1** [1] 2%	**2** [2] 3%
WQ2	**12** [6] 19%	**43** [11] 68%	**8** [2] 13%	**0** [0] 0%
DECOMP	**14** [5] 35%	**26** [8] 65%	**0** [0] 0%	**0** [0] 0%
WEAX	**16** [7] 22%	**51** [13] 69%	**4** [2] 5%	**3** [3] 4%

Table 4.2: Articulation Over Time

Articulations at the simple level again make up the majority of all instances in all four curriculum exposures. However, the percentage of articulations at the robust

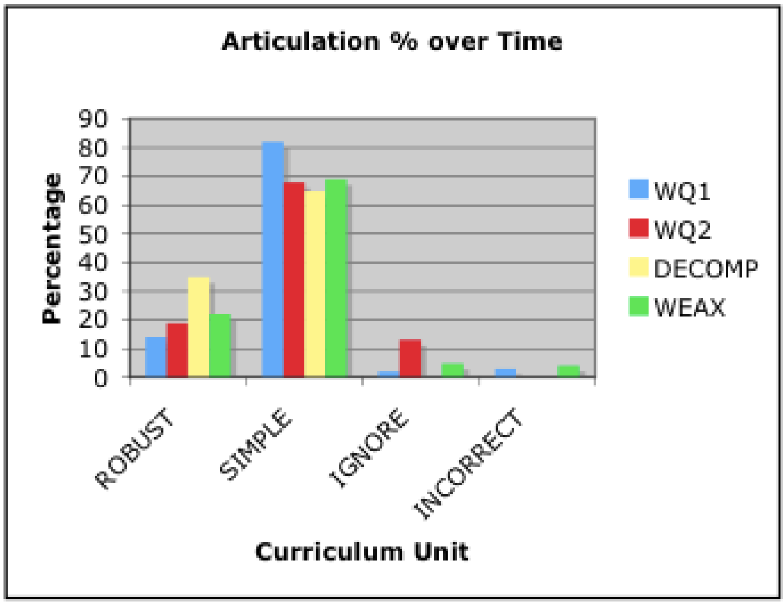

Figure 4.4: Articulation by type (percent) through the four curriculum units.

level increases through all three curricula in the 7th grade, and then drops slightly for the final curriculum in 8th grade. The steady increase in robust articulations, both in raw numbers and percent, across the three curriculum units in 7th grade is indicative of both greater facility with the modeling process and the use of the software. Percentage counts from Table 4.2 were reformatted with the curriculum units over time as nominal variables, and the percents of articulation as ordinal variables. However, a Friedman test showed no significant differences (3, N = 4) = 3.39, p = .497. This indicates that the changing percentages are not different from what is expected by chance. The lack of statistical significance of these results must be considered carefully. While the upward trend is clearly observable for robust articulations, the overall assessment of differences among all the counts found no significant differences. Using non-parametric statistical tests on small sample sizes makes it difficult to detect significant effects unless they are very large. Thus these results are equivocal and any conclusions drawn from them must be made cautiously.

4.2.4 Factors Related to Increased Articulations

There are a number of factors that may account for the small but steady increase in articulations: repeating content, narrowly focused units, increasing learner modeling skill, and ongoing teacher scaffolding. For example, in terms of curriculum, Water Quality 1 is followed by a deeper review of the same content area in WQ2, so one would expect an increasingly robust level of learner understanding. This deeper understanding would likely be seen in more robust articulations. In the subsequent decomposition unit, there is a high degree of hands-on activity in the creation and daily monitoring of the decomposition towers, and a more restricted domain of terms and concepts, which could also contribute to the increase in robust articulation. In terms of classroom context, the teachers do review each final model for a grade, and students may gradually become more aware that a complete model is considered to be one with filled in description (articulation) boxes. In addition to this feedback and motivation, the students also receive feedback on what a good model is during group demonstrations. There is some evidence that students internalize this requirement, for example, in tape 124, students are finishing the first phase of their model, having been given the task of creating several objects and factors and then saving. As they complete the last factor, one reminds the other to attend to the articulation box:

S1 Okay, so now we save

S2 You forgot to fill in the box, remember?

S1 Oh(types)... Save.

In tape 136 the students again prompt each other, after creating object they are working on associated variables. The student controlling the keyboard ignores the description box for the factor they have created, and moves the mouse to close the window and start a new factor. Their partner protests:

S1 Description!

S2 I'm sorry? . We already did a description

S1 I know but we have to do a description for the variables

and again in tape 133, when the student controlling the keyboard skips the articulation box and the other student points out: *"You have to put in a description, remember?"* To which the other student replies by clicking in the description box and saying, *"Description..."*

4.2.5 Teacher Scaffolding in Support of Articulations

Another contributing factor could be that teachers also remind students during classes to remember to fill in their descriptions as they create their models. This is over and above their stated requirement at the start of the unit that final models have all descriptions filled in. The standard practice of all three teachers was to regularly circulate through the classroom during model creation classes, stopping to visit each learner pair at least once. In tape 136 we see the teacher reviewing a relationship with the students and then offering an encouraging reminder:

S1 There, they have a relationship with the house.

(Teacher) Cool, make sure you fill in all the boxes, all right?

Again in tape 137, as another student pair discusses the improvement of their model, a similar reminder from the same teacher:

S1 Now we should do trees.

S2 No, remember we can do another variable

(Teacher) Make sure you fill in all the boxes, all right?

S2 Yeah

Even when just stopping for 10 seconds to look over a shoulder as in tape 160, with no prior discussion of content or the model, the one question asked by the teacher is: *"You're filling descriptions?"* Similarly in tape 142, the teacher is reviewing a model with a pair of students and goes out of her way to ask a researcher for help in reminding the students (referring to the articulation boxes), *"if you could, when you go around and look, and make sure they have something in every area."* To a pair at the start of the 8th grade weather pre-model who are unsure of their content knowledge and ask what to put in the description boxes, the teacher offers: *"Put in exactly what you might know"*. These types of procedural reminders focus just on the act of filling in the articulation boxes, but teachers also (when asked and on their own) conduct more in depth discussions about content in articulation boxes.

During more extensive discussions with students about content and structure, teachers still work in reminders, as on tape 157 as the teacher and students review model components on screen: *"all right, so let's click on that to see what you wrote"* insisting that the icons be opened up so as to see and assess the articulations. Teachers also offer discussions of why the articulations are required, as in tape 166, emphasizing the need to make their thinking explicit, so a stranger can fully understand what they are modeling for their water quality model. This pair is struggling with what to write in the description boxes, a problem discussed earlier. They complain repeatedly that they *"don't understand"* how to describe an object or factor, even though this is their second exposure to Model-It. To students who have made the mistake of putting relationship information into an object description, the teacher offers:

> *you don't want to talk about how your model is going to work here, what you're doing is, you're putting a description of that object, so I'm a stranger coming into this model, I want to know" (later) "maybe your object is, um, stream life, right, or aquatic life, and then in your description you can be more specific. You could say, this is stream life specifically macro-invertebrates. That way, when I look at your model I know what it really means.*

And still later in the tape the discussion continues:

S1 So, like, if you do a car, it's just this, I mean, everyone knows that's a car, like?

(Teacher) Correct, so for your description you could just say these are automobiles.

S1 Cars.

(Teacher) Yeah, these are, and maybe you want to be more specific because somebody could ask, 'do you mean electric vehicles? My town has a lot of electric vehicles. Is your pollution model still the same if all the cars electric?' Somebody could be silly like that, so that's where the description forces you to be specific, what is this object? Because maybe you have factories, you know, would you say there's a difference, there's a factory in one town that does nothing but make cars, there's a factory in another town that makes paper, and there's a factory in another town where they make computer chips. So you think all those factories are all the same in terms of how they contribute pollution?

S2 No, no.

Teacher and peer reminders, and discussions of model requirements, as described in this section, occur consistently across the curricula, whereas the changes in unit content and focus are unique to each modeling exposure.

Based on the analyses presented, the answer to the question "Does learner articulation of their thinking change over time?" is yes, but with qualifications. As noted previously, simple articulations were the modal response, but robust articulations were made in substantial numbers even in the first curriculum unit model. The percentage of robust articulations increased with each curriculum unit in 7th grade, and then dropped slightly in the final model the following year at the start of 8th grade. While increasing facility with the tool and the usability of the scaffolds may account for part of this trend, a number of other factors were discussed that could be contributing to the observed effect.

Having reviewed the student interactions with the scaffolds in the tool, I next review the interactions around the tool, in the form of dialog between learner pairs, other students, and teachers. The goal of this analysis will be to develop an answer to the second part of research question one.

4.2.6 Quality of Discussions Around the Scaffold

Learner dialog around articulation scaffolds was coded on two dimensions in keeping with prior research. One set of codes captured essentially the length and degree of interaction between the pair, while the other captured the scientific quality of the dialog. Consensual was the shortest and least collaborative, Responsive was a bit longer and both learners contribute at least something, and Elaborative involves both learners discussing in depth through multiple conversational turns. Inaccurate is discussion that is not scientifically correct, Simple is discussion that is correct but not of much depth or substance, and Robust is discussion that is scientifically correct and meaningful. Using these two sets of codes, every instance of dialog (that was on-task and intelligible) was evaluated for quality and type. The number of co-occurring instances of each resulting possible coding pair throughout the data corpus is shown below in Table 4.3.

Table 4.3 shows the vast majority of the interactions were characterized as simple in quality, and of those most were consensual or responsive where learners were elaborating on each other's ideas. As expected, there were no instances where elaborate dialog was coded as inaccurate. The shared understanding that results from the longer elaborative pattern means that if one of the learners does have a misconception, it gets corrected. Also, robust quality discussions were most often of the elaborative type, with a smaller number being responsive and almost none being consensual.

Simple quality discussions were almost always associated with the consensual or responsive dialog patterns, where conversations were shorter. The length of discussion would appear to be related to the quality of the dialog. The counts in Table 4.3 were reformatted with both sets of codes as ordinal variables. A Chi-Square test revealed that the data for the two sets of codes were related $\mathbf{c}^2(4, \mathrm{N} = 116) = 33.63$, $p < .001$. This means, as the pattern in the chart would seem to indicate, that two variables are related in a way not ascribable to chance. Very short or single learner discussions were overwhelmingly Simple and almost never Robust, whereas longer, dual learner discussions were mostly Robust, never Inaccurate and rarely Simple.

4.2.7 Examples of Dialog Around Articulation Scaffolds

Dialog that occurs during learner pairs' use of articulation scaffolds takes three primary forms, as discussed above. Simple and Consensual is just short quips such as one learner suggesting what to put for Object "sun", and saying "hot?" the reply is "yes". The next form is Simple and Responsive, which is discussion conducted at a simple level, but for a greater number of conversational turns and with more participation from each learner. For example, in tape 136, students are creating objects for their water quality model, and are in the early phase of model construction, creating numerous factors. They have selected several icons, including a building and a tree, and are exploring plausible relationships between them. They propose a relationship between people building houses and cutting down trees, which is simplistic because this is a water quality model and the actual effect of building a house on local tree populations is negligible. They show a Responsive dialog pattern, where one student provides the idea and justification, and the other mainly agrees but also adds to the discussion. So, while not Consensual (where one learner simply confirms), it is also not long or substantive enough of a conversation to be Elaborative. In this case, one student's simplistic understanding of how local trees are affected by local house construction is not challenged by the other student:

S1 The pollution would affect the trees.

S2 The people affect the trees.

S1 How?

S2 Because they could cut them down to make more houses, which means the houses also affect the trees. Decreases, by, no, by a lot.

S1 No, no.

S2 By a little, okay. Trees will

S1 Yeah, here, by a little.

S2 They'll chop down the trees.

This interaction contrasts sharply with the example below, where a more robust and elaborative discussion occurs.

The most desirable form, as students work with the scaffold, is to have dialog that is both Robust and Elaborative. In the example below from tape 140, we can see two students working on their model of a stream and discussing the addition of trees and how these might affect various qualities of the stream. The consistent turn-taking and debate between the students shows a desirable degree of cognitive conflict and involvement in the model construction process, illustrating how they co-construct a mental, and then computer, model of how leaves get in the stream and impact water quality. They are proposing adding trees to the model, and are discussing how the number and proximity of the trees, and the resultant leaf litter, could affect the turbidity of the stream:

S1 Did we already do number of trees and leaves? There or there?

S2 Yeah so wait, the number of trees. Equals number of leaves.

S1 Wait doesn't, wait doesn't it seem like we already did this?

S2 No we did size didn't we?

S1 Yeah you're right.

S2 Okay. (they are looking at a relationship articulation box)

S1 By a lot? Yeah it increases.

S2 By a lot, sure.

S1 A lot cause like the more, there, if there's a lot more trees there's a lot more leaves.

S2 Yeah cause it like multiplies each time right?

S1 Yeah.

S2 So. Because the more trees not that hard to understand.

S1 Yeah. Okay so um the location of the tree is gonna affect the...

S2 Would affect, well it wouldn't affect the number of leaves.

S1 No.

S2 But the number of leaves in the water. Location of the tree will affect the turbidity right?

S1 Yeah cause if it, cause if it's like not by, over the stream.

S2 Yeah.

S1 The uh, yeah.

S2 Turbidity. Increases by a little, not that much.

S1 Yeah. Oops.

S2 Because the closer the tree is to the the more chance that there is. More chance, how do you spell chance?

S1 See, yeah. That, for leaves to fall in.

S2 For leaves to fall into the water. Well and then hold on.

S1 Okay.

S2 Which raises the turb. (idity)

S1 There we go. This is cool.

S2 This is really cool. Okay so let's see

In this Elaborative pattern, we can see two learners working together, adding to each other's ideas, to plausibly link trees and leaves to the turbidity of the stream in their model.

The answer to the question, "While using tool scaffolds for articulation, what is the quality of discussions learners engage in, does this change over time, and if so, how?" is complex. Learners did regularly engage in discussions around model content, but they also frequently were silent or engaged in off-task discussions while using the scaffolds. The scaffolds can only guide, not control, student cognition, and any time you put two students together you will have some amount of off-task discussion, given normal social needs.

4.2.8 Changes in Discussions Over Time

Since discussions were determined to be predominantly simple coupled with consensual or responsive, we can now investigate if these overall averages hold true across each individual curriculum exposure and model. Matrix counts were made for the co-occurrence of dialog codes in each of the three 7th grade curricula and the one 8th grade curriculum.

These tables show very little change in the types of dialog over time. On summary counts, the units have nearly identical numbers and patterns of distribution. The counts were reformatted with dialog type and quality, as well as curriculum unit, as ordinal variables. A Friedman's ANOVA was performed which found no significant differences: (3, N = 9) = .13, p = .989. This means that differences between the counts over the four curriculum units are that which we would expect by chance, reinforcing the observation that the numbers are nearly identical.

While there are not substantive changes between units to discuss, there are several interesting observations to be made when examining the instances of dialog throughout the data corpus. In both the Weather and WQ1 units there were several cases

	Consensual	Responsive	Elaborative
Robust	3	10	18
Simple	37	35	8
Inaccurate	2	3	0

Table 4.3: Total Occurrences of Dialog Quality vs. Type

	Consensual	Responsive	Elaborative
Robust	0	1 (3%)	4 (14%)
Simple	9 (32%)	10 (36%)	2 (7%)
Inaccurate	0	2 (7%)	0

Table 4.4: Occurrences of Dialog Quality vs. Type in WQ 1 Total = 28

	Consensual	Responsive	Elaborative
Robust	1 (3%)	4 (14%)	5 (17%)
Simple	9 (31%)	9 (31%)	0
Inaccurate	1 (3%)	0	0

Table 4.5: Occurrences of Dialog Quality vs. Type in WQ 2 Total = 29

	Consensual	Responsive	Elaborative
Robust	1 (4%)	2 (8%)	4 (17%)
Simple	7 (29%)	6 (25%)	0
Inaccurate	4 (17%)	0	0

Table 4.6: Occurrences of Dialog Quality vs. Type in Decomp Total = 24

	Consensual	Responsive	Elaborative
Robust	1 (3%)	3 (8%)	5 (13%)
Simple	12 (31%)	12 (31%)	4 (10%)
Inaccurate	1 (3%)	1 (3%)	0

Table 4.7: Occurrences of Dialog Quality vs. Type in Weather Total = 39

of Simple dialog coded with the Elaborative pattern, which is otherwise uncommon. The longer Elaborative pattern was overwhelmingly seen in conjunction with Robust scientific discussion. However, it is possible to have a learner pair engage in a long running discussion that despite taking considerable time, results in very little substantive discussion. By way of example, in tape 223, a pair is early in the process of creating a weather model and discussing how ozone might relate to weather. They consider customizing the values for the factor as well as how to define it. They have a long conversation which they both contribute to, but they do not engage in Robust discussion. They raise issues like the ozone layer, the possibility of a hole, or holes, in it, and the greenhouse effect (which is related to, but not the same as the ozone layer), but they do not engage these ideas or each other.

S1 Okay uh, ozone layer.

S2 Okay.

S1 Maybe we should probably say like thin right? Thin in places.

(they are discussing how to customize the values)

Thin.

S2 It's like thin.

S1 Maybe a hole too.

S2 Yeah hole.

S1 What, oops sorry.

S1 Okay so hole, (Laughs) thin, and thick.

S2 Sure.

S1 Yes. Well that's how it is, okay.

S2 But it's like in different places. Ozone layer far North of America and then we begin to do ozone layer core Antarctica.

S2 Okay you know what? No.

S2 Exists, and then we're gonna have something called global warming so.

Do you know what global warming is?

S1 Yes, yes.

S2 Okay.

S1 Really

S2 No what I did is, is, okay sure. But um

S2 The greenhouse, the green.

(they are typing in description)

Greenhouse.

S1 Yeah the greenhouse.

S2 The greenhouse effect.

S1 Yeah that thing, yeah doesn't that affect like the ozone layer?

S2 It melts the ice and the ice goes whoooo, and it fills up the ocean with ice.

S2 Okay never mind then, okay shut up. If there is a hole above a land-mass.

S1 There's a hole just like a circle in the sky.

The discussion above (coded elaborative but simple), ends up being another discussion at a simple level, it just plays out in an elaborative format where they never firmly engage relevant scientific content related to the description they are making.

When considering the question, "Does the quality of discussions learners engage in change over time?" the answer would seem to be that there is no noteworthy change. The dialog was predominantly simple and consensual, or simple and responsive, with a significant minority of responses as robust and elaborative. This overall average pattern held true when each curriculum unit was examined alone, and when the units were compared the pattern looked almost identical. Robust dialog was almost always associated with elaborative responses. This pattern held true across all four units with no clear trend of change.

4.2.9 Discussion and Articulation Considered Together

Given the preceding discussion of learner articulations using scaffolds and the two aspects of learner discussions that occur around those scaffolds, it is interesting to consider how the two are related. By developing a matrix of when certain codes co-occur in the data, a picture can be developed of how often certain types of dialog occur in conjunction with a certain type of articulation. Because the three types of scaffolds all similarly encouraged articulation, scaffold articulations were collapsed across all scaffolds (object, factor, relationship) to a single total and dialog was collapsed to the four most frequent combinations as discussed earlier (Robust content with either Elaborative or Responsive pattern, and Simple content with either Responsive or Consensual pattern). This allows a comparison of articulations in the scaffolds and the type of dialog that co-occurred with that articulation. Table 4.8 below shows the results of this analysis.

Articulations	Simple & Consensual Dialog **Count**	Simple & Responsive Dialog **Count**	Robust & Responsive Dialog **Count**	Robust & Elaborate Dialog **Count**
IGNORE % of total artic.	**3** 100%	**0** 0%	**0** 0%	**0** 0%
INCORRECT % of total artic.	**0** 0%	**1** 100%	**0** 0%	**0** 0%
SIMPLE % of total artic.	**35** 44%	**34** 43%	**1** 1%	**9** 11%
ROBUST % of total artic.	**5** 15%	**7** 21%	**9** 26%	**13** 38%

Table 4.8: Articulations as they co-occur with types of Dialog

In a continuation of earlier patterns, we see that simple articulations and simple dialog dominate the count, and co-occur the most frequently. Almost 90% of simple articulations are associated with simple dialog. The percentages for robust articulations and dialog are not quite as strong, but still over 60% (26% + 28%) of robust articulations were associated with robust dialog. Counts from Table 4.8 were reformatted with articulation and dialog codes as ordinal variables. A Chi-Square test found a significant relationship between type of dialog and quality of articulations $\mathbf{c}^2(9, \text{N} = 118) = 41.01$, $p < .001$. This indicates a relationship between the types of articulations and the types of dialog that cannot be accounted for by chance. Simple dialog was overwhelmingly associated with simple articulations, and robust dialog was most commonly associated with robust articulations.

In summary for research question one, it was found that articulation scaffolds are used successfully the majority of the time, and that the percentages of robust articulations tended to increase over time. The types of dialog engaged in while using articulation scaffolds were mostly simple and sometimes robust, and this did not change over time. A comparison of the most common types of dialog with articulations showed that robust dialog was associated with robust articulations and simple dialog was associated with simple articulations.

4.3 Model Quality and Scaffold Use

Research question two asks: *Does the nature (quality) of students' scaffold use affect the quality of their final model?* This question focused on the artifact, the final model (digital file) created by the learner pair and submitted to the teacher as a representation of their learning in fulfillment of a class requirement. Drawing on, and integrating, established systems for evaluating artifacts in general, and concept maps or dynamic models specifically, student models were scored across detailed areas grouped in nine sub-topics, summarized by three main sections and combined into one final score as a representation of model quality. The model evaluation rubric can be found in appendix D. Given that the 20 final models analyzed came from a set of cases distributed across four unique curriculum exposures over the course of a year, this led to two sub-questions: What is the quality of learner pairs' final model artifact? Does the quality of the model change over time?

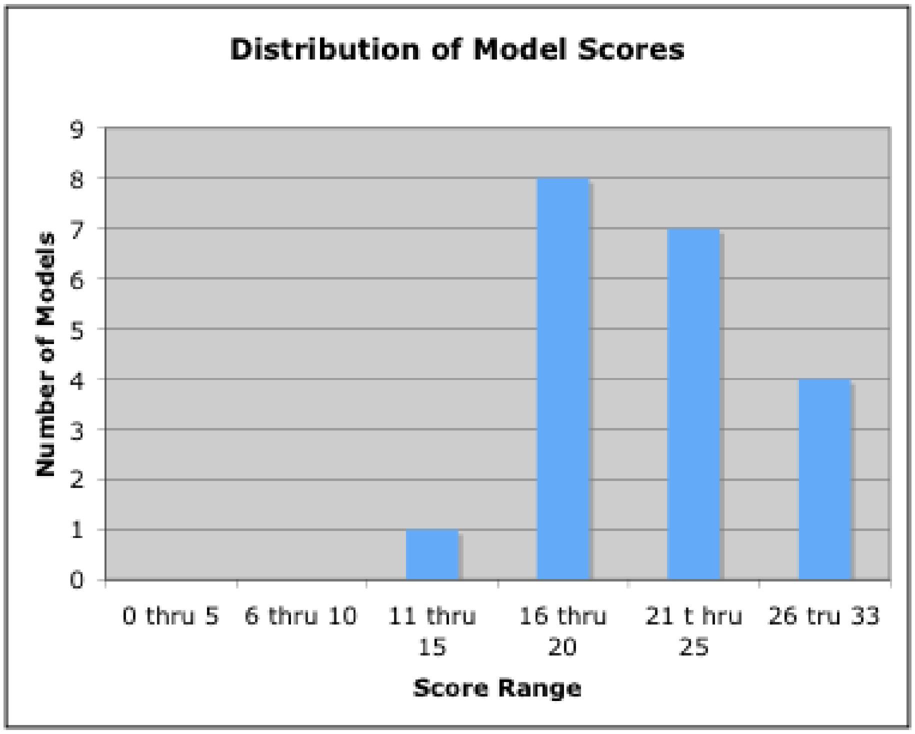

Figure 4.5: Grouped Range of Model Scores

4.3.1 Quality of Model Artifacts

Model artifacts were coded using a detailed three-part rubric. The maximum total points for a "perfect" model was 33. For the purposes of this analysis, any score of 26 or above was considered Excellent, 21-25 Good, 16-20 Average, 11-15 Poor, and 10 or below Unacceptable. Model scores ranged from 13 to 29 with models most frequently falling in the 16-20 range. Only one model scored as poor and none were unacceptable. Figure 4.5 shows the distribution of scores.

The overall average quality of models was good. The average score overall was 21.3 which equates to a "good" model. In this summary view, the data confirm prior research (Spitulnik, 1999; Stratford, 1996) that showed students can use Model-It to create models of substantial scientific quality.

Range of model scores. To illustrate the types of model creation efforts seen, I now present three cases. The first case will be of the highest scoring model, the next

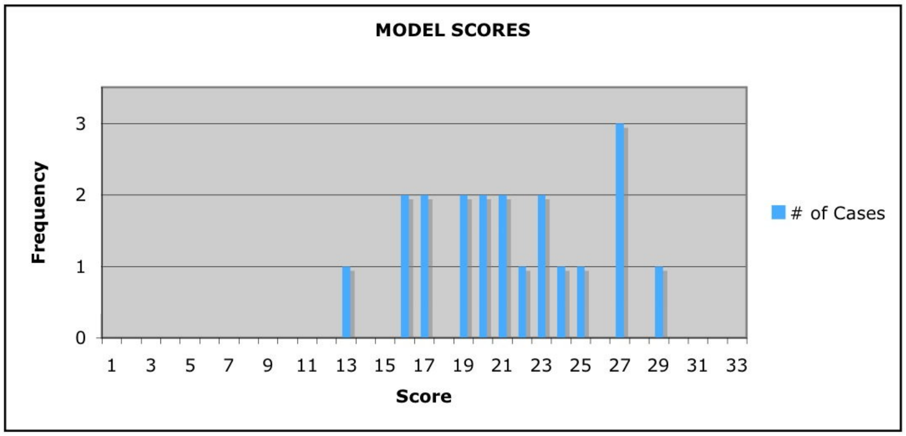

Figure 4.6: Frequency of Model Scores

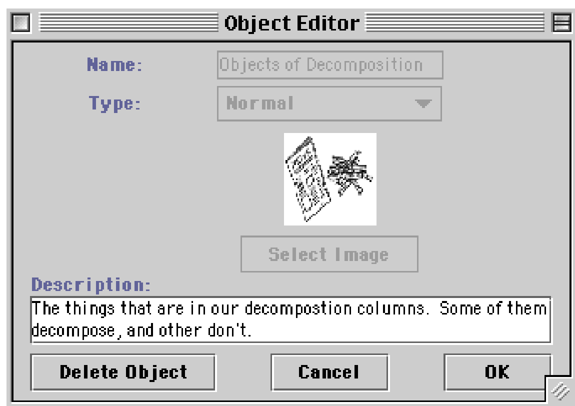

Figure 4.7: High Scoring Model Object

will be of an average model that scored 20 on the overall rubric and the last will be the lowest scoring.

An excellent model. Case 09, which had the highest scoring model, was an excellent example of model building done right. This pair, a male and female student, was making a model for the Decomposition unit, the third exposure to Model-It in the 7th grade. Their model scored 29 and had articulations in 100% of their scaffold boxes, and scored very highly for a model that had significant complexity of structure and content without any obvious missing or illogical components. They captured the key variables from their decomposition tower (temperature, water, bacteria, and worms) and linked them together in a parsimonious model that was carefully crafted. They

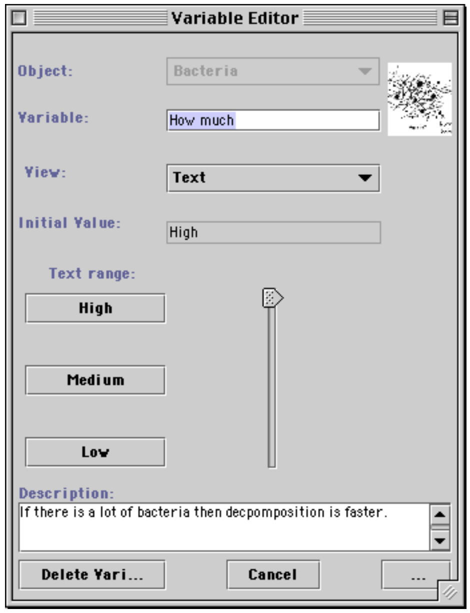

Figure 4.8: High Scoring Model Factor

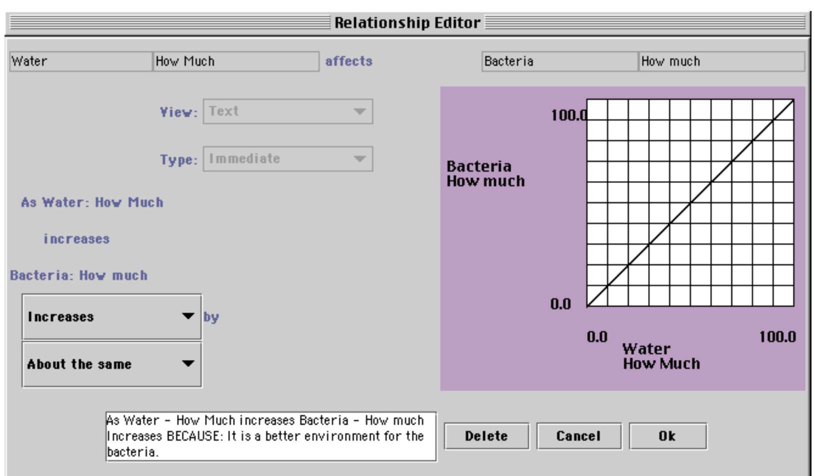

Figure 4.9: High Scoring Model Relationship

incorporated robust features such as a bell-curve relationship between moisture and worms to capture the idea that if the worms become immersed in water (or completely dry) they will die off. Examples of their model components can be see in Figures 4.7, 4.8, and 4.9. This pair had a very positive attitude, focusing well on the model creation task for long periods of time. They had almost no off-task discussion. They worked steadily and experienced no problems with the software or reasons to rush their work. They had several robust and elaborative discussions while making the model, and also had two instances of scaffolding from the teacher. This interpersonal scaffolding focused on motivating the students and in one case focused on sharing knowledge about decomposition as she helped them think about their model. The decomposition unit was also the most "experiential" in that the students created two decomposition towers and watched things decompose for weeks while learning about the processes. As with all units except weather, they made only the summative model at the end of the unit. Thus, this pair was not only working well as a team within a richly constructivist unit, but had support from the teacher, and was working in the third and final exposure of 7th grade when we would expect the highest level of experience and ability with the program.

An average model. In case 11 we see an average model, scoring 21, which was the approximate mean for all models. In this case a male and female learner are creating a second Water Quality model. This is their second exposure to Model-It in seventh grade and a deeper review of content they were first introduced to earlier in the year. This model was somewhat parsimonious, with only six objects, and six factors which mirrored the objects exactly. Examples of their model components can be see in Figures 4.10, 4.11, and 4.12. These six factors were linked with nine relationships. Their stated goal was to model how humans affect water quality. The model components scored well overall, with no descriptions left blank. This model failed to include any common water quality indicators (the subject of the unit), such

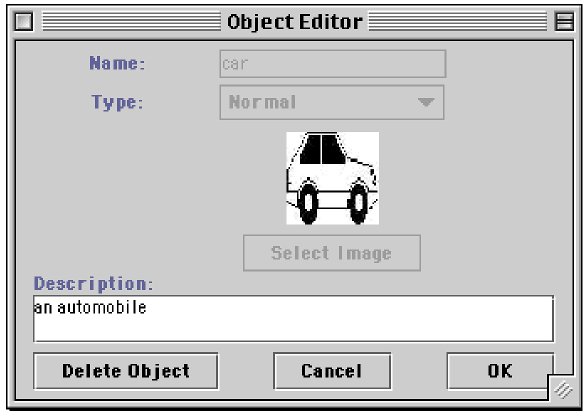

Figure 4.10: Average Model Object

as Dissolved Oxygen or Turbidity or pH (despite specific prompting to this pair to use water quality indicators like pH during class by the teacher, this pair used objects like "road" "humans" and "buildings"). They also structured their model so humans were the only independent variable, and they drove all the other variables up, sometimes at extreme rates, so when running the model, and increasing the number of humans, there was an immediate drop in water quality and not much else could be seen. This pair was struggling with how to describe objects and factors which may account for the parsimony. There were two instances of teacher scaffolding where the pair was given in-depth explanation and analogies to clarify the task of articulating descriptions. Even after this scaffolding, the pair continued to express confusion and frustration, although they do, in fact, complete articulations for each model component. The pair is off task for one short length of time during the first day, but stays on task overall and progresses methodically. Overall, this pair shows some skill in using Model-It and average content knowledge with their model, but clearly they have not mastered the modeling task nor did they demonstrate in their model the water quality factors they had just spent weeks working with. We might speculate that the challenges faced by this pair in articulation contributed to their struggle to build a complex model with multiple water quality factors.

A poor model. In case 02, which produced the lowest scoring model, there were a

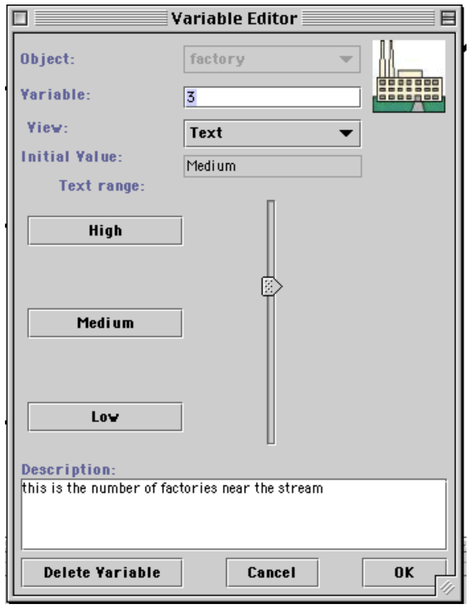

Figure 4.11: Average Model Factor

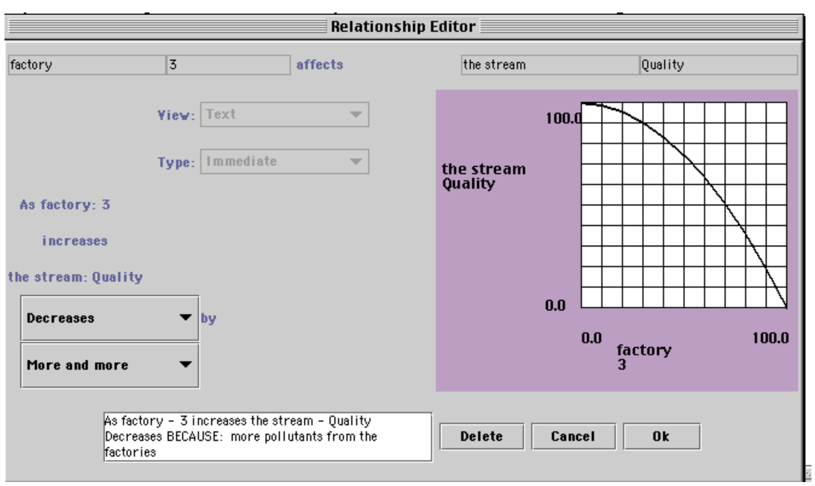

Figure 4.12: Average Model Relationship

number of issues that may account for the low performance. This pair, two males, was making a model for the weather unit, which was the fourth exposure to Model-It, and depending on the students' past experience it could have been the twelfth to fifteenth class they used the program. They were returning to Model-It for the first time as 8th graders, after approximately five months which included summer break. Their model scored 13, which was the lowest of any model analyzed. The model had serious problems in terms of structure and content. Examples of their model components can be see in Figures 4.13, 4.14, and 4.15. Despite the previously discussed efforts by teachers to make expectations of articulation clear, this pair entered minimal description for some objects and then entered no description (articulation) at all for nearly 100% of their factors and relationships. This decision alone ensured their model would score very poorly. Furthermore, when they did create factors they would choose arbitrary names so an object would be "we iker this image" and the associated factor would be called "Kenny", which they found to be an amusing typo on their statement that they liked the image, and the name of one of the pair. While they had seventeen relationships, over half involved nonsensical factors (e.g., the relationship between "Shrubbery" and "Barometric Pressure").

Figure 4.13: Low Scoring Model Object

As an additional example, this pair selected an icon that looked like a small tree and chose to name it "shrubbery" and then make the description based on a humorous

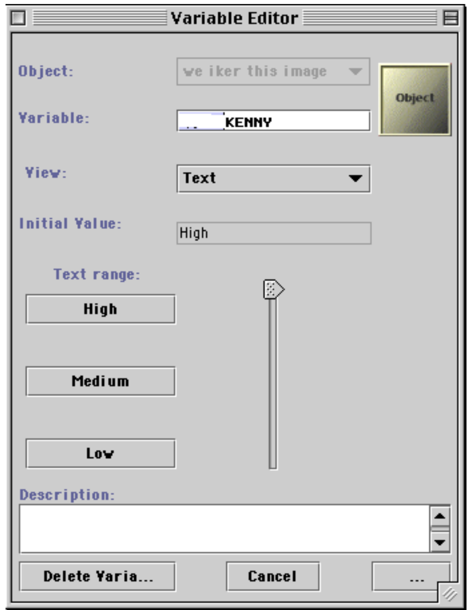

Figure 4.14: Low Scoring Model Factor

quote from a Monty Python movie. Even when specifying their relationships (which, as noted above had no explanatory articulation) they would make scientifically implausible relationships such as having relative humidity affect wind speed, or having a factor in a relationship for "where rain happens" which had no reasonable association to the model purpose. The only "output" factors in this model were "Shrubbery" and "Precipitation – where it happens" (as opposed to amount, or frequency), both of which have only dubious connection to the curriculum content. There were some relationships, such as one showing lower Barometric Pressure relating to increased Precipitation, which were relevant and well designed, but these were in the minority.

The pair for case 02 voiced negative opinions from the very start of their first day. Complaints about the program and or the computer hardware were heard, such as *"I hate this Mac keyboard"* and *"Awww, I hate this program"*. Once they began creating objects and factors, they showed very little serious interest in using the program as they had been taught in the previous year. There was a high degree of "off-task" discussion about TV shows and current events or classmates. This pair had no scaffolding interactions from the teacher, which is relatively unusual. Lastly,

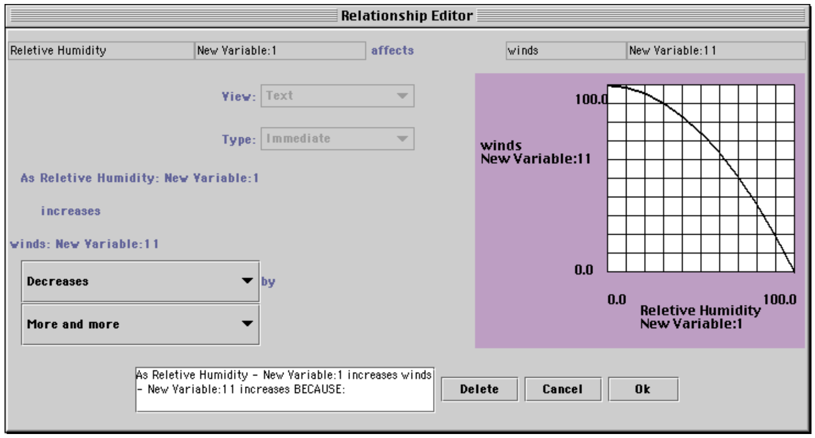

Figure 4.15: Low Scoring Model Relationship

this pair may have lost both motivation and ability in terms of using the program due to the summer break.

In summary, this pair was both displeased with the program and task, and mostly off-task, so as a result they produced a model that scored very low according to the rubric. This pair also had no teacher interaction, which was very unusual. While they did in fact make a number of scientifically correct relationships and choose some factors related to weather, their overall lack of attention to detail resulted in a very poor model.

The answer to the question, “What is the quality of learner pairs’ final model artifact?” is a complex one. Clearly the fact that all but one model scored average or better is heartening, and an examination of the best model reveals a high quality artifact with rich evidence of students’ creativity and understanding of complex phenomena. This best model was balanced by one lower quality model that revealed just how far off track the modeling process can get when frustrated and unmotivated students also don’t get any teacher scaffolding. The importance of teacher involvement with the students as they use software is clearly indicated by the low scoring model case, where students familiar with the software and the content got badly off track and mostly used the software to amuse themselves and create a model only to check the box' and complete the assigned task. However, if we ignore that one case for a moment, we see that the majority of students do use the software to create models that are good or excellent representations of their scientific understanding.

4.3.2 Changes in Model Quality Over Time

Models from each curriculum exposure were averaged and this average score is shown over time in Figure 4.16. There were four models in Water Quality 2, Decomposition and Water Quality 1 had five, and Weather had six.

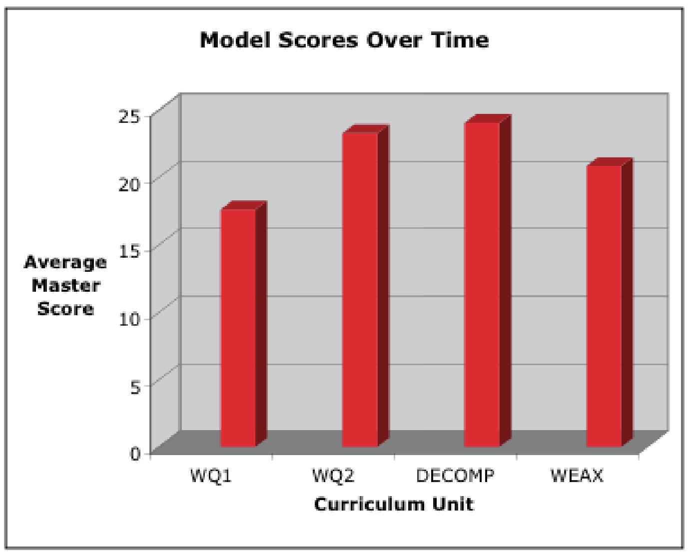

Figure 4.16: Model Scores Over Time

The average model score is lower for the first use in WQ1, and then increases steadily throughout the 7th grade exposures for WQ2 and Decomp. For the final exposure in the Weather curriculum in early 8th grade, the score drops slightly. If the one very low scoring model is dropped, the score for Weather is 22 and even closer to the scores for WQ2 and Decomp. It is possible that some loss of motivation and program skill may have occurred over the summer break before the 8th grade Weather unit. The graph of average model scores over time without the outlier case can be seen below in Figure 4.17.

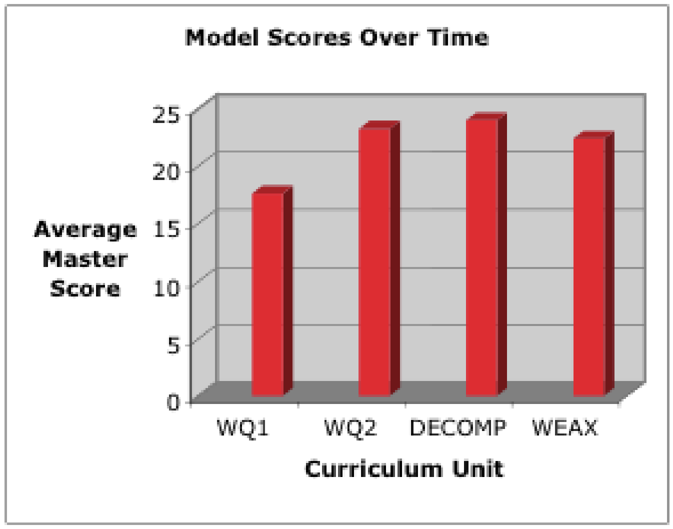

Figure 4.17: Model Scores Over Time, excluding lowest scoring model

Average model scores peak or plateau with the third exposure in Decomposition. The repeated exposures during 7th grade steadily increase in average model score. The very best scoring model occurred in the third curricular exposure and many of the best models occurred in the last two exposures. In fact, when looking at only models scored Good or better, there were *no* such models in the first exposure (WQ1) and then four such models in WQ2, DECOMP, and WEATHER. Thus, the number of high quality models increases after the first exposure and remains high over time.

In summary, the answer to the question, "Does the quality of the model change over time", is yes and in a positive direction. The fact that all models in the first exposure were graded Fair or better indicates that the scaffolded dynamic modeling tool Model-It supports successful modeling without requiring extensive training or support. This is in keeping with prior single-exposure research. That average model scores, and the number of high scoring models, increase over time shows that students can increase their model quality with repeated use of Model-It in a supportive context. The fact that the worst model was created in the last exposure cautions against overconfidence. Clearly some learners continue to require consistent support from the teacher to have an optimum modeling effort.

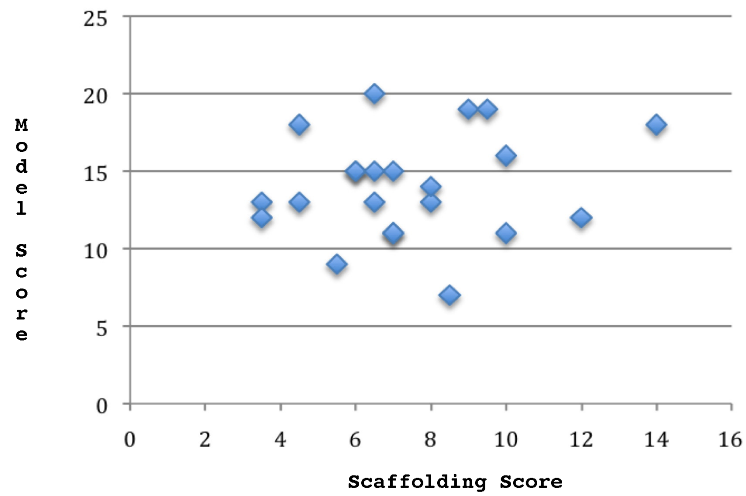

Figure 4.18: Plot of Scaffolding Scores vs. Model Scores

4.3.3 Model Quality and Scaffolding Use

The question of model quality's relation to scaffolding use is complex. Because the articulations evaluated for scaffolding use were part of the models, and thus part of model scores, they had to be factored out of the model score for comparison purposes. Specifically, one of three sub-components in the original master model score (the one that assessed the quality of written descriptions) was dropped. When this was done, the range of scores that was 13-29 changed to 7-20. The highest and lowest scoring model remained the same. These model scores were matched with the related master scaffolding score for the learner pair in question. The matched scores for the 20 cases were examined with Pearson and Spearman (parametric and non-parametric) tests to determine relationship between the variables, if any. No difference was found, r(20) = .167, p = .46 [Pearson], r(20) = .107, p = .64 [Spearman]. Conceptually this means that there is almost no relationship between revised model quality and scaffolding use. As can been seen in the plot, Figure 4.18, the distribution is quite broad, and a low degree of relationship would be expected.

4.4 Summary of Findings

In the most broad and general sense, the findings were supportive of the hypotheses behind the research questions. Articulation scaffolds were used successfully and with consistency, but most often at a basic or simplistic level. The nature of discussions around the scaffold was also consistently at a simple and responsive level. However, robust and elaborate discussions did occur for some learners, and this happened more often over time. Learner models scored over a wide range, but were of good to high quality overall. Model quality was seen to increase over time through the four curriculum exposures, with a slight drop in the final exposure. Contrary to expectations, no significant relationship between scaffold use and model quality was found. Next, in chapter five, I discuss the implications of these findings as they relate to the larger context of science learning, human-computer interaction, and the design of software tools.

CHAPTER V

Discussion

5.1 Discussion

At an educational research conference panel discussion about scaffolding in science education, an audience member once queried, "If scaffolding is the answer, what is the question?" The larger question is and has always been how to best promote robust learning in the resource-constrained classroom environment. Providing and improving tailored support to learners in their Zone of Proximal Development is one strong answer. Technological tools have a role to play, and their design has been the subject of ongoing study. The use of scaffolded technological tools in ILE's is a challenging process and raises numerous questions about the design of software interface elements as well as the interaction of teachers, tools, and learners in the busy classroom environment.

The results from this study suggest that particular simple tool scaffolds designed to promote articulations, even in a general sense, succeed in that goal most of the time. Scaffolds were almost always used and the articulations were almost always correct, but did not always reflect robust understanding. The frequency of related scaffolding provided by teachers for the same goal highlights the importance of distributed (Putambekar & Kolodner, 2005) and synergistic (Tabak, 2004) scaffolding. The quality of written articulations supported by the scaffolds improves over time through the first three exposures in one school year, but this trend was not found to be statistically significant. Learner dialog that took place around scaffold use was varied. While a large amount of total learner discussion was off-task, there were frequent discussions between learner pairs in which they shared or refined their scientific understandings. This occurred most often around the use of articulation scaffolds as they had to argue about what to type in. The types of dialog observed were similar to those observed in earlier research (Zhang, 2003) and as with articulations, were most commonly Simple. In contrast to the articulation quality, dialog quality and frequency did not change over time. The quality of learner model artifacts was good overall, which is in keeping with previous research. Artifact quality improved over time which extends previous research. Model quality was compared with scaffold use for each pair, but no significant relationship was found.

This chapter reviews and elaborates on these findings and discusses the limitations and implications of this study. Four major findings are presented, along with their relationship to the literature and general implications, as well as specific implications for the design and use of scaffolded software tools like Model-It. A conclusion then suggests possible directions for future research.

5.1.1 Articulation Scaffolds, Even Simple Ones, Work

The answer to the question whether scaffolds for articulation lead to learners articulating their thinking in writing was "yes". While learners articulated their descriptions of Objects, Factors, and Relationships 93% of the time they encountered the scaffold, the vast majority of those articulations were of the Simple type. These Simple articulations were usually very short or one word comments, or the articulation failed to describe the Object and rather provided information about how the Object was related to some other Object or Factor (which is information that belongs in the Relationship articulation box). The Robust type of articulation occurred 21%

of the time, and was more likely for articulations in Objects and Factors than in Relationships. Direct assessment of these unsophisticated instantiations of Quintana and colleagues (2004) design guideline [for the facilitation of articulation and reflection], showed that these scaffolds are effective. Despite the minimalist design of these scaffolds, they are used successfully almost all the time. In addition to the scaffold design, the synergistic effects of teacher scaffolding (both before and during modeling tasks) play a major role in this successful use. The results of this study also show, however, that there is room for improvement.

The large number of simplistic descriptions or putting Relationship information in Factor descriptions creates the impression that the scaffolds could be better designed to guide learners in their articulations. The scaffolds do not contain direct prompts or examples, and only the Relationship articulation box contains a partial sentence learners must fill in. While learners have had several chances to observe the teacher creating each model component, and receive ongoing teacher scaffolding and guidance during each modeling session, there could be more. Taking a cue from McNeill (2006), who found distributed scaffolding helped middle school science learners develop their scientific discourse skills, scaffolds for articulations could be embedded into curriculum documents, or provided in a handout. The fact that robust articulations are not the most common type is disappointing in light of the designed goals of the tool and the scaffolds. This is even more noteworthy given the frequent and clear guidance and incentives the teachers provided to use the scaffolds.

In addition to tool design changes (discussed later in this section), this finding suggests possible modifications to the classroom use of a scaffolded tool like Model-It. A more detailed review of each type of articulation during early demonstrations from teachers might help. For example, a teacher might put up an Object "Stream" with a description of "wet" and contrast it with an Object "Stream" with a description of "the small stream behind our school that drains our parking lot and field".

Additionally, detailed critiques and feedback during the creation or review of the models (e.g. asking not "did you fill in your descriptions?" but "are your Object descriptions complete and do they describe the Object itself and how it is measured?", etc), could improve success. While relying on related tool and teacher scaffolds is in keeping with the concept of distributed scaffolding (Putambekar & Kolodner, 2005) and synergistic scaffolds (Tabak, 2004), the teacher cannot supervise every modeling action of every pair, so an improved tool interface is important. An interface that allowed easy observation and sharing over the network would facilitate more teacher and peer review of a given model.

Learner articulation changed over time, with an increase in robust articulations over the first three curriculum units. This is clearly a desirable effect, but it did not persist into the fourth unit that occurred in the start of 8th grade after summer break. Additionally, these changes were not found to be statistically significant (across the first three, or all four units), possibly due to the small sample size and the low power of the non-parametric analyses required. Increased facility with the scaffolds may account for some of the improvement, however there are several other possibilities. Changes in curriculum content and related activities (e.g. the repeat of Water Quality or the tight content focus of Decomposition) could account for some of the increase in robust scores. When learners have increasing mastery over their content (Water Quality), or are dealing with a simpler and specific experiment (decomposition towers) and more limited domain of terms and concepts (Decomposition), then they might be expected to provide more robust descriptions of the facets of their models. The cumulative effect of ongoing reminders and prompts from teachers and peers alike could also account for some of the increase. The importance of this synergy with the tool scaffolds is further discussed later.

The slight drop in the final (weather) unit at the start of 8th grade was unexpected. While the "summer break" factor surely accounts for part of the drop, it may again also be related to curriculum in that the weather unit returns to a very broad content area where student content mastery may vary widely. Given that these scaffolds do not fade, it might also be the case that the learners (on their fourth exposure to the tool) are ready to create models focusing on function rather than detailed internal description, and they may be performing the sort of "self-fading" reported by other researchers (Jackson, 1999; Luchini, et al., 2003) where they simply engage the scaffold less often or less enthusiastically. Self-fading would only be undesirable to the degree that it affected performance, so this is an interesting area for further study.

Earlier studies of Model-It (Jackson, 1994) and Symphony (Quintana, 1999) (a similarly scaffolded tool) found that learners did engage all scaffolds with equal success, in contrast to the general success and improvement over time seen in this study. Quintana (1999) found that scaffolds, included prompted activity text fields, were not consistently used by all learners, and that some were not used at all. Jackson (1994) noted that several scaffolds, including those for articulation, were sometimes ignored by learners (both early on and as they gained facility with the tool). An extremely important difference in these studies is that they were focused narrowly on the tool interface and interpersonal interaction around the use of the tool (from teacher or researcher) was deliberately minimized. Theses studies thus provide an interesting contrast to this study, in which the synergistic effects of teacher scaffolding may account for the different utilization of articulation scaffolds.

5.1.2 Teacher Scaffolding Was an Important Factor in Successful Scaffold Use and Model Creation

While not directly assessed in its own research question, interpersonal scaffolding was a topic addressed in chapter 2 and was accounted for in the coding scheme as part of the holistic assessment of scaffolding use. Teachers provided scaffolding of some kind to almost every learner pair analyzed, in most cases engaging them multiple times with both low level reminders and higher level discussions about the details of their descriptions and model structure. As noted in chapters 3 and 4, teachers not only set up the modeling task with the requirement that articulation scaffolds be used, but they also were quite disciplined about continually reminding the class and the individual pairs to ensure they were using the scaffolds. In the examination of the lowest scoring model, there was no teacher scaffolding for that pair. In contrast, the highest scoring model had multiple instances of high level teacher scaffolding. While this study was not specifically designed to test the influence of teacher scaffolding, and there is insufficient statistical power to substantiate this Relationship, it is noteworthy.

This finding would be expected given the emphasis in recent literature on the importance of scaffolding from multiple sources (Puntembakar & Kolodner, 2005) and the synergistic effect of multiple scaffolds supporting the same learning goal (Tabak, 2004). Additionally, McNeill and Krajcik (2009) found that teachers' instructional practices interacted with context-specific scaffolds to allow students to make the greatest improvement, which suggests a synergy between related interpersonal and tool scaffolds. There are several implications for the use of Model-It (and similar tools) in the classroom. First, the teachers and curriculum could emphasize even further the initial review of Model-It functions, and perhaps even add a demonstration and component by component review of a high quality sample model. This specific demonstration of how the scaffold is best used, perhaps contrasted with a poor use, would be helpful in setting learners up for successful use. In order to foster greater opportunity for interaction with "more knowledgeable others", the teachers and curriculum could emphasize a new task of demonstrating the model to another learner pair, after some set period of work time.

5.1.3 Dialog Did Not Improve Over Time Like Articulation, But High Quality Articulation Was Most Often Associated With High Quality Dialog

The question about the quality of the discussions learners have while using tool scaffolds for articulation is complex. While there was regular dialog and it usually had to do with the model content, it was not always a clear reflection of the learners' thinking and learners had a large number of off-task discussions while creating their models. Dialog that occurred while learners were using articulation scaffolds showed an association, in that shorter and simplistic dialog most often was associated with shorter and simplistic articulations. Consensual and Responsive patterns (with one learner leading, or short/basic comments back and forth) were almost always associated with the Simple quality of scientific discussion. The Elaborative pattern, with multiple conversational turns was most commonly associated with Robust discussion. This finding is similar to prior research showing patterns of dialog during modeling practices. Zhang (2003) noted the three types of dialog patterns but did not assess quality. Novak and Krajcik (2006) noted that scaffolds seemed to enable high level discourse, but the direct association of these patterns with the quality of scientific discussion is new. Model-It articulation scaffolds, in this regard, are passive (text boxes) and dialog is not directly scaffolded. As discussed later, scaffolding learner dialog directly may be a productive direction for future interface designs.

The sheer frequency and volume of off-task discussion may argue for need to make the task, and the scaffolds, more engaging. Of the dialog that did occur during scaffold use, the majority of the instances discussed scientific content in a Simple way and in a Responsive or Consensual style, where the conversation was neither deep or long. It was relatively rare to have a pair of learners who were similarly enthusiastic about the task, and who both tended to verbalize to advance their ideas, productively collaborating. Far more often, a single member of the pair controlled the keyboard input and thus the dialog around the model creation task. The other learner would offer suggestions but not truly collaborate with the one controlling the keyboard and mouse. Researchers examining a tool similar to Model-It similarly found learner dyads often dominated by one learner (Ergazaki, et al., 2007). Since the student controlling the keyboard had de facto control of the articulation (and often the dialog), the teachers actively sought to share keyboard time equally by having students switch positions during the class, or emphasizing at the start of a class that whoever had the keyboard the day before should not have it the next day. Although this helped ensure that each learner has a chance to directly engage the model, it did not solve the problem of having one learner "lead" and the other passively "follow". While the longer robust and elaborative discussions did occur, it is clearly desirable to have them occur more often for all pairs.

There was very little change in dialog over time. The types of dialog seen around tool scaffold use in the first unit are repeated across all units, with roughly the same frequency and quality. This lack of change in discussions over time stands in contrast to the pattern of improvement in the articulation seen over time. It may be that while dialog is important for learners to share their ideas, and they experience valuable cognitive conflict, other factors may be in play over time that result in improvement in the articulations that actually get typed into the description boxes (as discussed previously). Ergazaki and colleagues (2007) found the amount and type of dialog varied depending on the stage of model creation, but this was a single study of only three dyads. The type of and quality of dialog occurring around tool scaffold use had not been directly assessed for Model-It in previous investigations and thus represents an important contribution of this work. Since part of the case for articulation rests on the idea of cognitive conflict and supporting the learners as they struggle with sense-making, the lack of change in dialog quality with concomitant improvement in articulation quality found in this study is worth further investigation.

Given the association between robust dialog and robust articulation found in this study, it is possible that increasing the amount of robust discussion would lead to even more increases in robust articulation. The interesting question is how this might be accomplished in the classroom. As mentioned earlier, more frequent interaction with other learner pairs might help, with each period ending with a demonstration of the model to peers, similar to the "gallery walks" of Kolodner and colleagues (2003). These interactions could be scaffolded with prompts like "my model differs from yours because." or " the most important thing to improve about your model is ." Or perhaps it would follow naturally if the tool scaffolds and the teacher scaffolding behaviors increased emphasis on articulation quality, where instead of asking "did you fill in the boxes?", they are asked to share and critique what they wrote. Teachers might give an example of what productive discussion between learners might look like. Furthermore, with the lack of dialog and feedback associated with the "solo modeler" condition, teachers might productively focus extra attention on those learners who are alone on those relatively rare occasions.

5.1.4 Model Quality Improves Over Time, With Room for Improvement

The answer to the question about the quality of learner pairs' final model artifacts was that the models were almost always of average or better quality. The majority of models were Average or Good, with one Poor and four Excellent. This is very much in keeping with prior research (Stratford, 1996; Singer, 2000; Zhang, 2003) that found students were able to use the scaffolded Model-It tool to create models of substantial quality. The fact that there was one Poor model, and that Average models were the most common is also in keeping with prior research [albeit in an urban school setting) that shows students require support to avoid making common errors, such as invalid Relationships and inappropriate Objects (Singer, 2000). There is room to improve and particularly to support novices' initial model making attempt.

As noted in the example model cases from chapter four, while part of the recipe for an excellent model is the motivation and focus of the learner pair, it is also clear teachers play an important role. Teacher provided scaffolding was present for nearly all modeling sessions, but tellingly was almost entirely absent from the one model that scored Poor. Synergy between multiple sources of scaffolding is currently theorized to be important in the busy classroom environment (Tabak, 2004), and requires a consistent unified theory of design. McNeill and Krajcik (2009) found that teachers who provided a variety of instructional supports (interpersonal scaffolds) aligned with scaffolds in the curriculum produced a significant gain in tested scientific reasoning/explanation ability. Teachers who did not provide this support did not see a similar effect from the curriculum scaffolds alone. A formalized review of what scaffolds are provided to the learners, by which method, and why, could assist the classroom use of a tool like Model-It. In attempts to improve learner models, Singer (2000) used specific teacher demonstration, paper handouts that forced planning and articulation, and teacher review to reduce learner errors. The evidence from the example cases indicates such efforts might have been similarly helpful in these classrooms.

In the highest quality model, the learner pair was rarely off-task in their discussion, they used nearly 100% of their tool scaffolds, and they had numerous Robust and Elaborative discussions while using the tool scaffolds. There were also two instances of teacher scaffolding discussions with this pair. In the lowest scoring model, the pair was usually off-task in their discussion, regularly ignored or mis-used articulation scaffolds, and had no direct scaffolding interactions with the teacher. These two examples illustrate the value of interpersonal scaffolding (peer and teacher) as part of the modeling process. These examples also show how careful support is critical to realize the full benefit of such technologies (Novak & Krajcik, 2006), and provide support for Tabak's (2004) argument that synergy between scaffolds provides the most effective support to learners. Because the task of motivating and monitoring learners is so complex, keeping them focused, and critiquing their work cannot be handled by the tool alone.

Model quality did improve after the first curriculum exposure and in subsequent units over time. In the first Water Quality unit, there were no models that scored Good or higher. In all later units there were 3 or 4 such models. Thus both average and absolute model quality improved in the second exposure and remained at roughly that level for the next three exposures. This finding extends previous research on successful use of modeling tools (Stratford, 1996; Zhang, 2003) which did not assess model quality over time. Previous research had shown that Model-It, as a scaffolded tool, was able to support learners in creating models of substantial quality. This result confirms and extends that knowledge, by showing that model quality improves with repeat exposures, and remains high. The "learning curve" for Model-It is low, as desired by the designers (Quintana et al, 2001), and this along with learners norming their performance as a result of seeing the first round of models demonstrated, probably accounts for the improvement from the first to second exposure. Future research could examine if repeat exposures to the tool without such feedback and critiquing resulted in similar gains.

5.2 Summary of Contributions

This study makes several contributions to the field, addressed in general terms above, and clarified here. First, this study has extended our understanding of how dynamic modeling tools are used successfully over time, which could be useful in both classroom application and future redesign of this or similar tools. Second, in a more general sense, this study has provided a specific examination of articulation scaffolds as called for in the scaffolding design guidelines of Quintana and colleagues (2004). In their guideline 7c, they refer to "providing reminders and guidance to facilitate articulation during sense-making", and they directly include "even simple reminders" such as those found in Model-It. By examining the verbal and written evidence of learners' articulations around these scaffolds, insight has been gained into how these scaffolds can be best designed and employed. This will be further discussed later in this chapter. Third, this study, while not directly designed to assess scaffolding synergy, nonetheless found significant evidence in support of this important concept. Tabak (2004) describes synergistic scaffolding as multiple scaffolds from multiple sources interacting in support of one or more learning goals, and collectively providing greater effect than the scaffolds alone. She speaks to the diverse nature of learner needs and the tension between scaffolding agents (such as teacher and tool) that support the learner. From the examples in chapter 4, and as discussed in this chapter, this study provides multiple examples of the key role teachers play in the successful use of technology tools. This will also be further discussed later in this chapter.

5.3 Design Implications for Technology Tools

These findings suggest several possible modifications to how technology tools (such as Model-It) are designed and employed. It may be the case that providing specific examples, and/or generic prompts within articulation scaffolds would make things more clear for the learners and result in more robust articulation. For example, instead of the blank articulation boxes labeled "description", an example could be provided above the box, or a specific sentence stating declaratively what a high-level use of the articulation box would contain. Additionally a persistent field where the learners specify their "driving question" could assist in guiding learner elaboration. It is also well within current programming and computer power limits to employ rudimentary "artificial intelligence" engines to provide a sort of high level feedback when each learner input is provided. For example, after a learner clicks on "done", a blank articulation box would get a prompt for "please fill in a description for this", and a one or two word articulation would get a query such as "your description seems short, please ensure your description fully and specifically describes (Object or Factor)". This sort of contextual feedback from software agents has been attempted with some success in recent work (Forbus, et al., 2005), along with software based comparison of student predictions to actual program outcomes. This comparison idea could also be employed in tools like Model-It. For example, prior to testing their model, the learners could be asked to name their 1 or 2 key input (or independent) variables, and their 1 or 2 key output (or dependent) variables, and then specify values for them. They might also predict (in text) how raising or lowering one will affect the value of the other. Then, after the model is run, the results could be compared to the predictions by the software, and notable discrepancies could be highlighted for the learners for discussion and correction.

Continuing with this "artificial intelligence" idea, software tools could be designed to better support learner dialog. By monitoring learner progress, the tool could provide a contextual prompt to "share and demonstrate your current model (or work) with others" would pop up after, say, two cycles of testing, or the creation of "x" number of components and Relationships. The association found between robust dialog and robust articulation suggests that redesigning the scaffolds to encourage or require discussion might help. Prompts could periodically, (after a set amount of minutes, or a set amount of progress creating the model components) suggest that learners demonstrate their model or seek feedback from peers and/or the teacher. Dialog was examined, but not directly scaffolded, in this study, but the findings of this study indicate it might be productively scaffolded. The software might also be configured to suggest the learner pairs switch keyboarders at certain intervals, so that no one learner dominates the process.

For Model-It in particular, the interface might productively incorporate the many common errors uncovered by this and other research (e.g. Singer, et al., 2000), where learners make models with very flat structure, or only one link to and from each component, or have links to and from every component, or have "orphan" components with no connections at all. These sorts of models are not always erroneous, but the software might easily identify many of these common discrepant conditions and provide corrective/metacognitive prompts (e.g. "Your model's longest chain of linked Factors is TWO, and most high scoring models in this curriculum have chains of FOUR or more. You may be missing some of the complexity of the system you are modeling. Check with a teacher or peer if you are unsure."). All of these sorts of prompts could be keyed to custom variables for each curriculum, to account for differences such as decomposition curriculum having fewer concepts and more parsimonious models in general. Design changes such as these would make the task more faithful to the actual practice being taught, and likely increase the quality of models created.

These design implications partly address how a technological tool could be purposely designed to account for how it will be employed within the dynamic classroom environment. There are additional implications related to scaffolding from agents other than the software tool.

5.4 The Case for Synergy

In the larger vision of scaffolding throughout the learning environment (however it is defined), tool design is not enough. Even the best designed software tool is part of a larger system of supports enabling some larger learning goal. Explicitly taking account of the interactions between scaffolding sources (e.g. teacher and tool) will increase the likelihood of successful scaffold use and learner understanding. As discussed previously, teacher support around the use of the tool scaffolds in this study was both persistent and considerable. Evidence was presented for the value of this support in terms of the quality of the final student model. Yet the teachers in this study did not receive specific training on the employment of this particular technology tool, nor were their curriculum units designed around the use of this particular tool. Neither was the tool designed with a specific conception of what scaffolding teachers could provide, and what the teachers provide can be very important. McNeill and Krajcik (2009) found that teachers who provided instructional supports aligned with curricular scaffolds produced greater student success on measures of scientific reasoning and argumentation. That the teachers integrated the tool into the units, and consistently addressed the need for articulation did partly capture Tabak's (2004) idea of an "intentionally designed package" with "shared features", but a more fully integrated approach would almost certainly have been even more effective. Still, this study speaks to the validity of the Synergistic scaffolding concept, the importance of integrating teachers into the design and implementation plans for technology tools, and offers some suggestions of how this might be accomplished.

5.5 Limitations

There are a number of constraints and limitations to be discussed in relation to these findings. These limitations are related to classroom context, data collection, and the nature of the learners.

The classroom context (that is, the normal practice of the teachers) did not allow for learner pairs to remain the same over time, so no pair remained together across any two units. This impairs the analysis of cases over time, since any argument has to be made using aggregate comparisons by unit. The fact that one teacher was replaced by a newer teacher for one class in the final curriculum exposure adds an additional confounding Factor to the analysis, although this teacher was also highly qualified, followed the same lesson sequence, and engaged in similar interactions with the learners.

While the longitudinal design allowed for collection of data across four curriculum units, there were differences between the units that complicate analysis. The curricula were not standardized in format, size, or structure, so there were significant differences in the length of the units, and the amount of content (terms, activities, experiments) the students were exposed to. For example, since the teachers found Water Quality to be worthy of more emphasis, two of the units were on that one topic, and having one "repeat" in a group of four units raises a host of variables that complicate comparisons across units.

Problems with data collection also limited analysis in various ways. The technological limits of screen capture hardware make it impossible to read what is being typed on the screen, which sometimes impaired a full understanding of what was typed (e.g. for a model component that was later deleted and does not show up in the final model). Additionally, as new versions of the software were fielded, multiple versions of the program were required to read old model files, and some files were corrupted. This reduced the number of analyzable cases available.

The nature of the learners also resulted in some limitations. Given that learner discussion formed one cornerstone of the analysis, any lack of or lessening of discussion was problematic. Some learner pairs were simply poor conversationalists, and rarely verbalized their thinking. Occasionally a dominant learner would take control of the process and work without speaking at all. On days when one of the pair members was absent a similar effect was seen, as the learner worked alone. When learners felt rushed for time, such as on the final model creation day before presentation, they would work very quickly and speak telegraphically if at all. Learner pairs were often "off task" in their discussion, even while actively constructing aspects of a high scoring model.

In regards to generalizability, this sample is only one well-resourced independent school and three classrooms supported by experienced tech-savvy teachers, which impairs the ability to generalize the results to other settings. However, the results of this focused longitudinal examination of how middle school science learners interact with scaffolds in a dynamic modeling tool has provided some interesting results and suggested some directions for future research.

5.6 Future Directions

Future studies could extend this work in a variety of ways. Another longitudinal study with consistent pairs over multiple exposures would allow for descriptions of how specific pairs of learners evolve in their scaffold use and modeling practice. The addition of an outcome measure, such as a post-test on curriculum content could provide better evidence of successful scaffold use and model creation being related to test performance.

If variations on scaffold design were to be tested, a randomized trial where learner pairs were given different versions of the software would allow more direct assessment of what implementations of design principles work best. For example, a version with specific prompts (or other design changes suggested earlier in this chapter) and also reminders triggered by short (1-2 word) descriptions could be compared to a version with generic guidance written into the top of the articulation box, and perhaps a third group that got both features.

Future studies can broaden the investigation of this and other design guidelines, for example, embedding expert guidance. Continued longitudinal investigations could tease apart the "areas of best effort" for teachers versus tools as distributed and synergistic scaffolds become the standard. Additionally, studies might productively investigate the difference between scaffolding higher and lower demand cognitive tasks (e.g. describing an Object vs. describing a complex Relationship).

5.7 Conclusions

This study contributed several findings in support of existing research and several findings that extended existing research. While a more detailed assessment, this study reaffirmed prior research that showed learners could successfully engage scaffolds designed into a dynamic modeling tool, and then use that tool to create models of substantial quality. Focusing specifically on articulation scaffolds, and the scaffolding design guideline for supporting the articulation of ideas and sensemaking, this study characterized more specifically the frequency and quality of learners' use of those scaffolds. Findings also supported the importance of the teacher as a source of scaffolding, and the need to attend to the synergistic nature of scaffolding in ILE's. Most importantly this study showed that learners improve the quality of their articulations over time as they use the dynamic modeling tool repeatedly through multiple curricula during a school year. In terms of model quality, this study showed that model quality improves across multiple uses of the scaffolded modeling tool over time, but that this improvement plateaus in the fourth use. Finally, this study attempted to contrast learner pairs' use of scaffolds with the quality of their final model and found no significant relationship. Scaffolding remains a term, and a question, worthy of continued study.

APPENDICES

APPENDIX A

Coding Schema

1. Administrative

1.1 School

1.1.1 Name

1.2 Teacher

1.2.1 Sandra

1.2.2 Tara

1.2.3 Ursula

1.3 Unit

1.3.1 Water Q 1

1.3.2 Decomposition

1.3.3 Water Q 2

1.3.4 Weather

1.4 Period

1.4.1 A

1.4.2 B

1.4.3 C

1.5 Case #

1.5.0

1.5.1

1.5.2 etc to case 16

1.6 Grade

1.6.1 Seventh

1.6.2 Eighth

2. Scaffolding Use (Tool)

2.1 Articulation

2.1.1 Create/Modify OBJECT

2.1.1.2 Correct and Robust

2.1.1.3 Correct and Simple

2.1.1.4 Ignore or Gibberish

2.1.1.5 Incorrect

2.1.2 Create/Modify FACTOR

2.1.2.1 Correct and Robust

2.1.2.2 Correct and Simple

2.1.2.3 Ignore or Gibberish

2.1.2.4 Incorrect

2.1.3 Create/Modify RELATIONSHIP 2.1.3.1 Correct and Robust

2.1.3.2 Correct and Simple

2.1.3.3 Ignore or Gibberish

2.1.3.4 Incorrect

3. Scaffolding Use (Interpersonal)

3.1 Teacher

3.1.1 Change Task

3.1.2 Model Behaviors

3.1.3 Motivate

3.1.4 Share Knowledge

3.1.5 Share Metacognition

3.2 Researcher

3.2.1 Change Task

3.2.2 Model Behaviors

3.2.3 Motivate

3.2.4 Share Knowledge

3.2.5 Share Metacognition

3.3 Peer (rare)

3.3.1 Change Task

3.3.2 Model Behaviors

3.3.3 Motivate

3.3.4 Share Knowledge

3.3.5 Share Metacognition

4. Dialog (not scaffolding)

4.1 Type

4.1.1 Consensual

4.1.2 Responsive

4.1.3 Elaborative

4.1.4 Undetermined

4.1.5 Off-task

4.2 Substance

4.2.1 Robust

4.2.2 Simple

4.2.3 Inaccurate

5. Model Scores *5.1 Accuracy*

5.1.1 Section 1-A score

5.1.2 Section 1-B score

5.1.3 Section 1-C score

5.2 Completeness

5.2.1 Section 2-A score

5.2.2 Section 2-B score

5.2.3 Section 2-C score

5.3 Function

5.3.1 Section 3-A score

5.3.2 Section 3-B score

5.3.3 Section 3-C score

2. Scaffold Use (TOOL)	
2.1 ARTICULATION	This is a master code for Articulation
2.1.1 Create/Modify OBJECT	**This is a master code for use of the OBJECT scaffold**
2.1.1.1 Correct and Robust	Learners make a complete and scientifically accurate text entry. Example: for Object "Clouds" they enter "visible moisture in the air"
2.1.1.2 Correct and Simple	Learners make scientifically accurate but simple text entry. Example: for Object Decomposition they enter "rotting"
2.1.1.3 Ignore or Gibberish	Learners type nothing into the box, or type random text. Example: asdfhjhk
2.1.1.4 Incorrect	Learners enter scientifically inaccurate or unrelated text. Example: spoon or for Object Stream they enter nice
2.1.2 Create/Modify FACTOR	**This is a master code for use of the FACTOR scaffold**
2.1.3.1 Correct and Robust	Learners make a complete and scientifically accurate text entry, including the because statement. Example: As Sun:Amount of sun increases then Air : Relative Humidity decreases by about the same BECAUSE sun increases temperature and warmer air can hold more moisture
2.1.3.2 Correct and Simple	Learners make a scientifically accurate but simple text entry, including the because statement. Example: As Stream : Dissolved Oxygen increases then Fish : Population increases by more and more BECAUSE fish need to breathe
2.1.3.3 Ignore or Gibberish	Same as 2.1.1.3
2.1.3.4 Incorrect	Same as 2.1.1.4

Table A.1: Summary of Scaffold (Tool) coding.

Scaffold Use (Interpersonal)	
3.1 Teacher	*This is a master code for Teacher Scaffolding*
3.1.1 Change Task	Teacher helps learners select task, modifies task for learners by reducing degrees of freedom, prompts/helps learners maintain direction on task.
3.1.2 Model Behaviors	Teacher demonstrates or models the desired action or process, or invites learners to demonstrate for teacher.
3.1.3 Enhanced Motivation	Teacher attempts to develop shared interest and goal, makes explicit attempts to control frustration or risk, provides subtle constructive criticism or praise, gives encouragement in support of goal.
3.1.4 Share Knowledge	Teacher provides content or task knowledge using Socratic questioning techniques or tailored assistance. Explanations may be offered using metaphor or PCK. Learner understanding is verified.
3.1.5 Share Metacognition	Teacher provides guidance on thinking or practice. Encourages reflection on prior work and/or problems that are occurring. Assists learners in making generalizations about their thinking or practice.
3.2 Researcher	*This is a master code for Researcher Scaffolding*
3.2.1 Change Task	Same as 3.1.1
3.2.2 Model Behaviors	Same as 3.1.2
3.2.3 Motivate	Same as 3.1.3
3.2.4 Share Knowledge	Same as 3.1.4
3.2.5 Share Metacognition	Same as 3.1.5
3.3 Peer	*This is a master code for Peer Scaffolding (rare)*
3.3.1 Change Task	Same as 3.1.1
3.3.2 Model Behaviors	Same as 3.1.2
3.3.3 Motivate	Same as 3.1.3
3.3.4 Share Knowledge	Same as 3.1.4
3.3.5 Share Metacognition	Same as 3.1.5

Table A.2: Summary of Scaffold (Interpersonal) coding.

4. Dialog (not scaffolding)	
4.1 Type	*This is a master code for Dialog (generally between peers)*
4.1.1 Conceptual	One learner drives discussion, the other does not respond, assents non-verbally, or acknowledges agreement with simple phrases or repeat-back. Example: How about a Stream object, for Water Quality? Sure
4.1.2 Responsive	Both learners discuss through at least four conversational turns. Minor elaboration of the original idea may occur. Example: How about a Stream object, for Water Quality? Do we need one? Well, lake is not the whole system, and our measurements were in the stream. OK, lets do it
4.1.3 Elaborative	Both learners actively construct a collaborative understanding and response, through at least six conversational turns. Prior comments are built upon and elaborated with each turn. Example: How about a Stream object, for Water Quality? Do we need one? Well, lake is not the whole system, and our measurements were in the stream. Well, it is the same water, right? No, there can be stuff that enters the stream after it leaves the lake, and stream aeration can increase D.O, so.. Right, right, and the stream is too small for fish etc.
4.1.4 Undetermined	Learners engage in conversation that is not off-task, but cannot be coded in the categories above.
4.1.5 Off-Task	Learner engage in conversation unrelated to science or model construction. Example: are you going to the dance Friday?
4.2 Substance	*This is a master code for Dialog Substance*
4.2.1 Robust	The scientific content of the discussion is robust and correctly stated. This code will often, but not always, occur with 4.1.3. (e.g. learners could have an elaborative conversation about misunderstood science). Example: Learners have a four turn conversation that contains detailed rationale for using a bell curve relationship for how pH affects fish population.
4.2.2 Simple	The scientific content of the discussion is simple, but not incorrect. This code will often, but not always, occur with 4.1.2. (e.g. learners could have a longer elaborative conversation that never gets very specific). Example: Learners could have a seven turn conversation that only in the end arrives at a basic statement like run-off increases turbidity.
4.2.3 Inaccurate	The scientific content of the discussion is wrong. Example: So, increasing moisture increases the decomposition of plastic yeah, its like rusting OK.

Table A.3: Summary of Dialog (not scaffolding) coding.

5. Model Files	
5.1 Accuracy	*This is a master code for MODEL ACCURACY*
5.1.1 1-A Score	This code is obtained by using model scoring rubric
5.1.2 1-B Score	This code is obtained by using model scoring rubric
5.1.3 1-C Score	This code is obtained by using model scoring rubric
5.2 Completeness	*This is a master code for MODEL COMPLETENESS*
5.2.1 2-A Score	This code is obtained by using model scoring rubric
5.2.2 2-B Score	This code is obtained by using model scoring rubric
5.2.3 2-C Score	This code is obtained by using model scoring rubric
5.3 Function	*This is a master code for MODEL FUNCTION*
5.3.1 3-A Score	This code is obtained by using model scoring rubric
5.3.2 3-B Score	This code is obtained by using model scoring rubric
5.3.3 3-C Score	This code is obtained by using model scoring rubric

Table A.4: Model file coding.

APPENDIX B

NVIVO Example Case File

EXAMPLE TEXT FILE FROM NVIVO

CASE 16 7th Grade Water Quality 1 C - Tina Tape 1 of 3

Summary-

(02:15)

Okay so were gonna click water quality. Okay. Okay so we first need to make something right?

Yeah.

Right. Lets make a person.

Wait just a second, okay.

Okay.

So we need trees.

Trees, I dont see a tree. (they are looking at icons in icon bar, then choose a new blank object and search for images) Oh um.

It looks like a tree, new object.

Trees, trees yeah, yeah. (no descriptions entered) Okay background or just normal?

Um.

Normal right?

Normal yeah.

Zero variables.

Now we need to add variables to it.

Lets move it up here first. Okay so what are our variables? So how do we do this, new variable?

Yeah. Um go to.

Okay.

New object, click new object.

New object zero right? Now what should our variable name be?

Wait just one second. Okay um and the like new type, click type, like a little arrow by type and go normal and then wait. No go, I think, yeah Im not quite sure Nick did it.

Number of leaves right?

Yeah.

Okay. View as (inaudible 03:44) or number, because theres numbers? (they are choosing how to customize their factor range) Yeah.

Well we dont know the actual number do we?

Oh thats right.

So lets do text okay?

Yeah.

Okay so this tree should have high amounts of leaves.

Okay.

Hows that? Description, description of the variable or description of the trees?

Um Im not sure.

Okay you raise your hand and Ill start typing the variable thing Okay?

Okay. Um do we out in every single variable on this or do we keep clicking to add a new variables? Like..

(Researcher) What do you mean, you want more than one variable?

Yeah we have a lot, we have lots of variables for trees; the number and the type and stuff.

(Researcher) Yeah you can just click on the tree and it.

Oh so I keep clicking over and over again?

Description is it just like, our thing is number of leaves and so we do it on number of leaves?

(Researcher) Um yes, so the object is trees, and you name....

Oh yeah okay. How do you give it a name?

(Researcher) No you cannot change it right now so.

Okay.

(Researcher) So yeah after that you just click a numeric variable and click on the tree again.

Okay.

Okay number of leaves on a tree will affect how many

Will fall.

Hoe many leaves will fall into the water.

And, and wait, and affect turbidity.

And that.

Thats all we need to put I think.

Okay.

Okay.

Three variables, there we go.

Uh huh.

How many more do we need?

Um yeah like it would like what were your objects? Just like the waterfall? (Talking to another student) I think so.

Hold on how do we do this?

Um oh wait.

For location.

For location of tree. Um.

Should we just not do that?

Lets just, wait lets, wait lets put location to the left just leave the hi, high, medium, low thing alone.

Okay, okay the location of the tree or trees?

Um trees, whatever, of the trees.

Whatever. Will aff, amount of leaves falling into or out of water. And is there anything else?

Um not really.

Okay.

Okay lets see we have one more the life that supports.

Okay. Life in tree.

Yeah how much? Like medium?

Yeah medium.

Okay. Um will affect, what should it do?

What? Um I dont know. Wait we said like erosion.

APPENDIX C

Model Summary from NVIVO

EXAMPLE MODEL SUMMARY FROM NVIVO

Decomp 7th tapes 191 200 209 Sandra

Number of Obj TEN

Number of Fac ELEVEN

Number of Relationships EIGHT

OBJ

FOOD custom these include our bread, apple, orange, and banana = 3

PAPER custom this is a line paper for our project = 3

POPSICLE STICK no icon This is our popsicle stick. It is covered in varnish and denser than our paper = 4

COTTON TERRYCLOTH custom this is our cotton terrycloth. It is directly made from an organic substance. = 3

DECOMPOSERS custom decomposers feed on the wastes or dead bodies of other living or once living materials. = 4

PRODUCTS THAT ARENT BIOTIC custom Human made products that are not biotic. The arent made from any living or once living materials. = 4

BIOTIC no icon everything that can decompose = 2

RATE OF DECOMPOSITION no icon The rate of decomposition is affected by the material being decomposed and the amount of decomposers = 3

OXYGEN custom Oxygen is a need for decomposers = 2

WATER custom Decomposer need water = 2

FAC

FOOD amount of decomp custom depending on the amount of decomposers present and the amount of time allotted, the amount of decomposition increases or decreases. = 3

COTTON T Amount of Decomp, Potentially custom The amount of decomposition, potentially is affected by the amount of decomposers and if the material is organic. = 3

POPSICLE STICK Amount of Decomp, Potentially custom The amount of decomp, potentially is affected by the amount of decomposers and if the material is organic. = 3

RATE OF DECOMP Rate of decomp std The rate of decomposition is affected by the amount of decomposers present. Decomposers depend on oxygen and air to survive. = 3

BIOTIC amount of biotic substances being decomposed. Custom The amount of organic substance being decomposed is affected by the amount of decomposers present. = 3

PAPER amount of decomp, potentially custom The amount of decomposition, potentially is affected by the amount of decomposers present and if the material is organic. = 3

WATER amount of water std Water is essential for decomposers to survive= 2

OXYGEN amount of oxygen std Decomposers need oxygen to survive and decompose. Unfortunately, decomposers use up the oxygen and leave less for the other organisms. = 4

HUMAN MADE PRODUCTS THAT ARENT BIOTIC Amount custom Materials that are non-organic do not decompose or take many years to do so.” = 3

FOOD new var std NO DESCR EXTRANEOUS

DECOMPOSERS how many decomposers are present custom Decomposers break down organic matter. By doing so, they enrich the soil. The four main types of decomposers are bacteria, fungi, insects, and earthworms. = 4

REL

Biotic-amount decomposed increases, Popsicle stick amount of decomp, P increases BECAUSE the popsicle stick is organic material. = 3

Biotic amount decomposed increases, Paper-amount of decomp P, increases BECAUSE paper is made directly from organic material = 3

Biotic amount of decomposed increases, Food- amount of deomposition increases BECAUSE foods are organic substances. = 3

Biotic amount decomposed increases, cotton Terrycloth-amound of decomp P, increases because it is made directly from an organic material, cotton. = 3

Biotic amount decomposed increases, Rate of Decomp rate increases BECAUSE Biotic stuff decomposes = 3 (redundant)

HUMAN MADE PRODUCTS THAT ARENT BIOTIC Amount of human produced non-organic matierals increases, rate of decomp-rate DECREASES BECAUSE non-orgainc materials do not decompose. Therefor there would be less for the decomposers to do. = 3

Decomposers -how many present increase rate of decomp-rate increases BECAUSE decomposers break down organic matter. The more decomposers present, the more decomposition. = 4

Oxygen-amount of oxygen increases, Decomposers -how many present increase BECAUSE decomposers depend on oxygen to survive. = 3

Water amount of water increases Decomposers -how many present increase BECAUSE there can be more decomposers. (what about immersed wood, like Venice pilings?) = 1

TEST

All relationships are straight linear, and all but one are positive. This produces a model with pretty clear function, as decomposition increase based in independent variables, the objects(amounts) drive to zero.

OVERALL SCORES

ACCURACY (0-4 for each OFR)

Objects = 30/10 = 3

Factors = 31/10 = 3

Relationships = 26/9 = 2.9

Overall = 9

COMPLETENESS (0-4 for OFR in terms of M/I/E)

Objects = 3 missing TEMP

Factors = 3 one extraneous

Relationships = 3 one illogical/redundant

Overall = 9

FUNCTION (1-3 for Fidelity, Structure complexity, Overall quality)

Fidelity = 3

Structure = 3

Quality = 3

Overall = 9

OVERALL MODEL SCORE (sum, out of 33 total) = 27

APPENDIX D

Scoring Rubric

CHAPTER THREE **METHODS** – DRAFT K 4/5/10 46

1 - ACCURACY **Instructions**

Based on what is presented in the model, and taking into account any differences in content the modelers were exposed to – how accurate is this model? That is, how "correct" are its individual parts and the descriptions they contain? This will generally focus on Object, Factors, and Relationships, but can also reflect the holistic accuracy of the model, the attention to detail in its component parts, and adherence to assigned requirements. (Make sure you keep "completeness" and "function" ideas out of this section, and in their respective sections below). Overall scores for each section are figured using the scoring boxes.

If possible, determine the Driving Question or assigned goal for the model, to assist in assessing the model. It is best to return to sections 1&2 after using section 3 and testing the model function, as this can highlight items that may be relevant for scoring section 1&2.

1-A OBJECTS

An Accurate model would, in terms of objects:
Have all descriptions filled in, with more than a simplistic restatement of the name (i.e. "these are flies" for Object-Flies), provide object description rather than factor or relationship comments (e.g. WORMS – "the more worms the less garbage") and avoid non-scientific or scientifically incorrect descriptions. If the information is scientifically correct, you should try to give the model/student credit for it, even if, for example, they put relationship information in the title of a factor.
Blank descriptions are coded in the Zero/One column.
0 – Unacceptable 1 – Below Average 2 – Average 3 – Above Avg 4 – Superior

EXAMPLES:
0 - No Description
1 – Description is one word or same as name, ex: Object BUGS – Description "bugs"
2 – Description is correct but simplistic or describes less than fully, ex: Object BUGS – Description "black and crawly"
3 – Description is correct, thorough and relevant, ex: Object BUGS – Description "these are crawling insects that eat leaves, like caterpillars"
4 – Description is correct and unusually detailed and specific.

1-B FACTORS

An Accurate model would, in terms of factors:
Have descriptions filled in, with more than a simplistic restatement of the name, that are appropriate for the Object, and avoid non-scientific or scientifically incorrect descriptions. Also, Ranges would be customized and relevant, not at default or illogical levels (i.e. Factor "number of worms" has range GOOD-BETTER-BEST).
Blank descriptions are coded in the Zero/One column.
0 – Unacceptable 1 – Below Average 2 – Average 3 – Above Avg 4 – Superior

EXAMPLES:
0 - No Description AND Default Range when default is illogical AND/OR Factor mismatched to Object ex: Object WORMS – Factor "amount of rain".
1 – Description is one word or same as name, that is, the description displays no additional information or understanding, OR Ranges make no sense, ex:: Factor "WORMS-Number of worms" has range of "GOOD-BETTER-BEST".
2 – Description is correct but simplistic, default ranges but still make sense. ex: Factor RAIN-Amount of Rain has default range as opposed to inches scale.
3 – Description is correct, thorough and relevant, with range that is customized for the factor ex: Factor "WATER QUALITY – pH" has range from 0-14.
4 – Description and custom ranges are correct and unusually detailed and specific.

1-C RELATIONSHIPS

An Accurate model would, in terms of relationships:
Have "because statements" filled in, that do more than simplistically restate the relationship, that avoid non-scientific or scientifically incorrect or unsupported causal explanations. Have scientifically correct degree relationships / graphs. Have little or no confusion about dependent and independent variables (example problem - amount of factories is dependent on amount of acid rain) If explanation is missing, the maximum score is 2. You must attend to both the presence/correctness of explanation AND the specification (direction, type, degree) of the relationship.
0 – Unacceptable 1 – Below Average 2 – Average 3 – Above Avg 4 – Superior
EXAMPLES
(write name of relationship in box in appropriate column)
0 – Relationship is not correctly specified, AND has no description or totally incorrect description.
1 – Has either a partly correct explanation OR a partly correct relationship specified. Ex:: relationship is "increase by a little" when it should be "by a lot", OR explanation of acid rain to pH is described as "decreases a lot because acid rain is bad for trees."
2 – Has a correct explanation, OR a correct specification, OR both description/specification are present but have errors.
3 – Has correct explanation and partially correct specification, or the reverse.
4 – Both explanation and specification are correct.

Figure D.1: Scoring Rubric, Page 1

Model File Name: ___________________ Pair: _________________ Create Date: _______ Teacher: _______ Coded by: ____________ KDI – Fretz rev M 01/25/02

1 – ACCURACY Scoring

1-A OBJECTS

0 – Unacceptable 1 – Below Average 2 – Average 3 – Above Avg 4 - Superior
(write name of object in box in appropriate column) Scores of ZERO are not tallied below, you can track Objects scored at ZERO by writing them here:

ONE	TWO	THREE	FOUR
a)This Col =	a)This Col =	a)This Col =	a)This Col =
b)Tot Obj. =	b)Tot Obj. =	b)Tot Obj. =	b)Tot Obj. =
c)% a÷b =	c)% a÷b =	c)% a÷b =	c)% a÷b =
d) % x 1 =	d) % x 2 =	d) % x 3 =	d) % x 4 =
Total of values of (d) >>>>>>>> round to X.X >>>>>>>>>>>			**1A =**

1-B FACTORS

0 – Unacceptable 1 – Below Average 2 – Average 3 – Above Avg 4 - Superior
(write name of factor in box in appropriate column) Scores of ZERO are not tallied below, you can track Factors scored at ZERO by writing them here:

ONE	TWO	THREE	FOUR
a)This Col =	a)This Col =	a)This Col =	a)This Col =
b)Tot Factor. =	b)Tot Factor. =	b)Tot Factor. =	b)Tot Factor. =
c)% a÷b =	c)% a÷b =	c)% a÷b =	c)% a÷b =
d) % x 1 =	d) % x 2 =	d) % x 3 =	d) % x 4 =
Total of values of (d) >>>>>>>> round to X.X >>>>>>>>>>>			**1B =**

1-C RELATIONSHIPS

0 – Unacceptable 1 – Below Average 2 – Average 3 – Above Avg 4 - Superior
(write name of relationship in box in appropriate column) Scores of ZERO are not tallied below, you can track Relationships scored at ZERO by writing them here:

ONE	TWO	THREE	FOUR
a)This Col =	a)This Col =	a)This Col =	a)This Col =
b)Tot Rel. =	b)Tot Rel. =	b)Tot Rel. =	b)Tot Rel. =
c)% a÷b =	c)% a÷b =	c)% a÷b =	c)% a÷b =
d) % x 1 =	d) % x 2 =	d) % x 3 =	d) % x 4 =
Total of values of (d) >>>>>>>> round to X.X >>>>>>>>>>>			**1C =**

ACCURACY MASTER SCORE & DESCRIPTION

1A ____ **+ 1B** ____ **+ 1C** ____ **=** $\mathbf{1}_{total}$ ______

Figure D.2: Scoring Rubric, Page 2

2 - COMPLETENESS Instructions

Based on what is presented in the model, and taking into account the idiosyncratic exposure to content (of the particular student) - how complete is this model? That is, how "thoroughly" does the model reflect what this student might be expected to know about the phenomena being modeled? Are there obvious Objects/Factors/Relationships implied for what is being modeled, which are missing? This section will generally focus on Object, Factors, and Relationships, but can also reflect the holistic completeness of the model and its adherence to assigned requirements. The general scoring rule from section 1 applies.

Missing – Ex: in model of Lake Water Quality, there is no object WATER.
Illogical– Ex: in model of River Water Quality, there is an object CANDY.
Extraneous– Ex: in model of Creek Water Quality, there is a factor TREES-Amount of Leaves. (where there is no tie in to shade, temperature, D.O. etc.)

2-A OBJECTS

A Complete model would, in terms of objects:
Have no obviously missing Objects (i.e. no "predator" Object in a model of wild rabbit population), or have extraneous or illogical object, Have an adequate overall number of Objects (unless model implicitly needs few Objects).

4 – Superior	(no obvious missing/illogical/extraneous objects)
3 – Above Avg	(one or two objects m/i/e)
2 - Average	
1 – Below Avg	(many objects m/i/e)
0 – Unacceptable	(most objects m/i/e)

2-B FACTORS

A Complete model would, in terms of factors:
Have no obviously missing Factors (i.e. no "Temperature" Factor for an Object "Air" in an Air Quality model), Have an adequate overall number of Factors (unless model implicitly needs few factors). Have a high Factor/Object ratio (where 1:1 implies each Object has only one Factor).

4 – Superior	(no obvious missing/illogical/extraneous factors)
3 – Above Avg	(one or two factors m/i/e)
2 - Average	
1 – Below Avg	(many factors m/i/e)
0 – Unacceptable	(most factors m/i/e)

2-C RELATIONSHIPS

A Complete model would, in terms of relationships:

Have no obviously missing relationships (i.e. acid rain not linked to lake water quality), or illogical (i.e. acid rain linked to number of houses) / extraneous (i.e. acid rain linked to car roof damage in a model of water quality) relationships.

4 – Superior	(no obvious missing/illogical/extraneous relationships)
3 – Above Avg	(one or two relationships m/i/e)
2 – Average	
1 – Below Avg	(many relationships m/i/e)
0 – Unacceptable	(most relationships m/i/e)

Figure D.3: Scoring Rubric, Page 3

Model File Name: __________________ Pair: ________________ Create Date: _______ Teacher: _______ Coded by: ___________ KDI – Fretz rev M 01/25/02

2 - COMPLETENESS Scoring

2-A OBJECTS

4 – Superior (no obvious missing/illogical/extraneous objects)
3 – Above Avg (one or two objects m/i/e)
2 - Average
1 – Below Avg (many objects m/i/e)
0 – Unacceptable (most objects m/i/e)

(write name of object in box in appropriate column) TOTAL OBJECTS: _________

MISSING	ILLOGICAL	EXTRANEOUS

2-B FACTORS

4 – Superior (no obvious missing/illogical/extraneous factors)
3 – Above Avg (one or two factors m/i/e)
2 - Average
1 – Below Avg (many factors m/i/e)
0 – Unacceptable (over half of factors m/i/e)

(write name of factor in box in appropriate column) TOTAL FACTORS: ________

MISSING	ILLOGICAL	EXTRANEOUS

2-C RELATIONSHIPS

4 – Superior (no obvious missing/illogical/extraneous relationships)
3 – Above Avg (one or two relationships m/i/e)
2 - Average
1 – Below Avg (many relationships m/i/e)
0 – Unacceptable (most relationships m/i/e)

(write name of relationship in box in appropriate column) TOTAL REL: ________

MISSING	ILLOGICAL	EXTRANEOUS

COMPLETENESS MASTER SCORE & DESCRIPTION

2A ____ + 2B ____ + 2C ____ = 2_{total} ______

____

____

____

____

____

Figure D.4: Scoring Rubric, Page 4

CHAPTER THREE **METHODS** – DRAFT K 4/5/10 50

3 – FUNCTION Instructions & Scoring

This section captures how the model works in TEST mode, and how reasonable the model is (on the whole) as a representation of what it is supposed to model. This section takes into account what the teacher reasonably expects from the student in comparison to the peer group / norm (not based on individual effort, i.e. "they worked for many hours", or potential).

3-A FIDELITY

Based on overall function (as opposed to the presence or absence of any one Object/Factor/Relationship (which is covered in section 2): The degree to which this model faithfully replicates the structure and behavior of the system it was designed to represent is:

3 – ABOVE AVERAGE (very realistic, only minor discrepancies)
EX: sun as source of "thermal pollution" for small stream
2 – AVERAGE (significant discrepancies but still in the ballpark)
EX: moisture affecting decomposition of metal at same speed mold affects apple, or amount of shade affecting number of bugs
1 – BELOW AVERAGE (major discrepancies)
EX: sewage output affecting acid rain, or leaf density peaking in Jan.

SCORE 3A = ______

3-B STRUCTURE COMPLEXITY

The structure of this model is:

3 – Sophisticated & complex
2 – Moderately complex
1 – Simple (linear relationships, "1-deep" star layout)

Sophisticated & Complex models:
Have "Interlinking", where chunks or chains within the model are also linked together (unless model implicitly requires separation)
Have long (3+) chains of relationships. Longest chain = ______
Have a high Relationship/Factor ratio (where 1:1 implies each factor is only hooked to one other factor).

SCORE 3B= ______

3-C OVERALL QUALITY

Does this model meet the assigned class criteria? Even if most things are technically correct, is it overall just barely enough to meet requirements? (Example: model has dozens of objects and factors,but overall explanations are weak, or factors contain "relationship-like" names or descriptions) Or, perhaps it is a small/simple model, but has excellent attention to detail and craft? (Example: a model with only 5 factors, but each has custom numerical ranges, and they are linked with complex, accurate relationships, like bell curve for pH)

For this model, in light of the overall impression from this model, in terms of its quality is:

3 - HIGH
2 - MEDIUM
1 – LOW

SCORE 3C = ______

FUNCTION MASTER SCORE & DESCRIPTION

3A ____ + 3B ____ + 3C ____ = 3_{total} ______

____

____

OVERALL MODEL SCORE

1 ____ + 2 ____ + 3 ____ = ______

Figure D.5: Scoring Rubric, Page 5

BIBLIOGRAPHY

BIBLIOGRAPHY

Anderson, J., Corbett, A., Koedinger, K., and Peltier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences*, 4((20):167–207.

Anderson, J. R. (1989). A theory of human knowledge. *Artificial Intelligence*, 40(3):313–351.

Azevedo, R. (2005). Using hypermedia as a metacognitive tool for enhancing student learning? the role of self-regulated learning. *Educational Psychologist*, 40(4):199–209.

Bell, P. and Davis, E. (2000). Designing mildred: Scaffolding students' reflection and argumentation using a cognitive software guide. In Fishman, B., editor, *International Conference on the Learning Sciences*, Ann Arbor, MI. University of Michigan.

Bloom, B. (1984). The 2-sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 11(13):4–16.

Blumenfeld, P. C., Marx, R. W., Patrick, H., Krajcik, J., and Soloway, E. (1997). Teaching for understanding. In Biddle, B., editor, *International Handbook of Teachers and Teaching*, pages 819–878. Kluwer Academic Publishers, Netherlands.

Bransford, J. D., Brown, A., and Cocking, R. R. (1999). Technology to support learning. In *How People Learn - Brain, Mind, Experience, and School*. National Academy Press, Washington, D.C.

Bronfenbrenner, U. (1979). *The Ecology of Human Development: Experiments by Nature and Design*. Harvard University Press, Cambridge, MA.

Brown, A., Ash, D., Rutherford, M., Nakagawa, K., Gordon, A., and Campione, J. (1993). *Distributed expertise in the classroom. In G. Salomon (Ed.), Distributed cognitions: Psychological and educational considerations (pp. 188-288).* Cambridge University Press, Cambridge, MA.

Brown, A. and Campione, J. (1994). Guided discovery in a community of learners. In McGilly, K., editor, *Classroom lessons: Integrating cognitive theory and classroom practice.*, pages 220–270. MIT Press / Bradford Books, Cambridge, MA.

Brush, T. and Saye, J. (2001). The use of embedded scaffolds with hypermedia-supported student centered learning. *Journal of Educational Multimedia and Hypermedia*, 10(4):333–356.

Chi, M. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences*, 6(3):271–315.

Chi, M. (2000). *Self-Explaining: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.).* Lawrence Erlbaum Associates, Mahwah, NJ.

Clement, J. (2000). Model based learning as a key research area for science education. *International Journal of Science Education*, 22(9):1041–1053.

Collins, A., Brown, J., and Newman, S. (1989). Cognitive apprenticeship:teaching the crafts of reading, writing , and mathematics. In Resnick, L., editor, *Knowing, learning, and instruction: Essays in honor of Robert Glaser.*, pages 347–361. Lawrence Erlbaum Associates, Hillsdale, NJ.

Council, N. R. (1996). National science education standards. Technical report, National Academy Press.

Council, N. R. (2000). Inquiry and the national science education standards: a guide for teaching and learning. Technical report, National Academy Press.

Crook, C. (1999). Computers in the zone of proximal development: implications for evaluation. In Lloyd, P. and Fernyhough, C., editors, *Lev Vygotsky: Critical assessments: The zone of proximal development.*, volume 3. Routledge, New York, NY.

Davis, E. (2003). Prompting middle school science students for productive reflection: Generic and directed prompts. *Journal of the Learning Sciences*, 12(1):91–142.

Davis, E. and Linn, M. (2000). Scaffolding students' knowledge integration: Prompts for reflection in kie. *International Journal of Science Education*, 22:819–837.

Derbentseva, N., Safayeni, F., and Canas, A. (2007). Concept maps: Experiments on dynamic thinking. *Journal of Research in Science Teaching*, 44(3):448–465.

Duschl, R. A., Schweingruber, H. A., and Shouse, A. W. (2007). *Taking science to school : learning and teaching science in grades K-8 / Committee on Science Learning*. National Academies Press, Washington, DC.

Ergazaki, M., Zogza, V., and Komis, V. (2007). Analysing students' shared activity while modeling a biological process in a computer-supported educational environment. *Journal of Computer Assisted Learning*, 23(2):158–168.

Festinger, L. (1957). *A theory of Cognitive Dissonance*. Stanford University Press, Stanford, CA.

Forbus, K. D., Carney, K., Sherin, B. L., and Ureel, L. C., I. (2005). A visual qualitative modeling environment for middle-school students. *AI Magazine*.

Fretz, E. B., Wu, H., Zhang, B., Krajcik, J., and Soloway, E. (2002a). A further investigation of scaffolding design and use in a dynamic modeling tool. In *AERA*, New Orleans, LA.

Fretz, E. B., Wu, H.-K., Zhang, B., Krajcik, J. S., Davis, E. A., and Soloway, E. (2002b). An investigation of software scaffolds as they support modeling practices. *Research in Science Education*, 32((4)):567–589.

Gilbert, J., Boulter, C., and Rutherford, M. (1998). Models in explanations, part 1: Horses for courses? *International Journal of Science Education*, 20(1):83–97.

Gobert, J. and Buckley, B. C. (2000). Introduction to model-based teaching and learning in science education. *International Journal of Science Education*, 22(9):891–895.

Gobert, J. and Discenna, J. (1997). The relationship between students' epistemologies and model-based reasoning. In *The Annual Meeting of the American Educational Research Association (AERA)*, Chicago, IL.

Gobert, J. D. (1999). *Leveraging technology and cognitive theory on visualization to promote students' science learning and literacy. In V. i. S. Education (Ed.), Visualization in Science Education*. Springer-Verlag Publishers, Dordrecht, The Netherlands.

Griffin, P., Belyaheva, A., Soldatova, G., and VHC (1993). Creating and reconstituting contexts for educational interactions, including a computer program. In Forman, E. A., Minick, N., and Stone, C. A., editors, *Contexts for Learning*. Oxford University Press, New York, NY.

Guzdial, M. (1994). Software-realized scaffolding to facilitate programming for science learning. *Interactive Learning Environments*, 4(1):1–44.

Guzdial, M. (1995). Software-realized scaffolding to facilitate programming for science learning. *Interactive Learning Environments*, 4(1):1–44.

Hmelo, C., Nagarajan, A., and Day, R. (2000). Effects of high and low prior knowledge on construction of a joint problem space. *The Journal of Experimental Education*, 69(1):36–56.

Hogan, D. and Pressley, M. H. (1997). Becoming a scaffolder of student's learning. In Hogan, D. and Pressley, M., editors, *Scaffolding Student Learning*, pages 185–191. Brookline Books, New York, NY.

Hogan, K., Nastasi, B., and Pressley, M. (2000). Discourse patterns and collaborative scientific reasoning in peer and teacher-guided discussions. *Cognition and Instruction*, 17(4):379–432.

Ingham, A. and Gilbert, J. K. (1991). The use of analogue models by students of chemistry at higher education level. *International Journal of Science Education*, 13:193–202.

Jackson, S. L. (1999). *The Design of Guided Learner-Adaptable Scaffolding in Interactive Learning Environments*. Dissertation, University of Michigan.

Jackson, S. L., Stratford, S. J., Krajcik, J., and Soloway, E. (1994). Making dynamic modeling accessible to precollege science students. *Interactive Learning Environments*, 4(3):233–257.

Jacobson, M. J. and Wilensky, U. (2006). Complex systems in education: Scientific and educational importance and implications for the learning sciences. *Journal of the Learning Sciences*, 15(1):11–34.

Kolodner, J., Crismond, D., Fasse, B., Gray, J., Holbrook, J., and Puntambekar, S. (2003). Putting a student-centered learning by design curriculum into practice: Lessons learned. *Journal of the Learning Sciences*, 12:485–547.

Krajcik, J., Blumenfeld, P. C., Marx, R., Bass, K., and Fredericks, J. (1998). Inquiry in project-based science classrooms: Initial attempts by middle school students. *The Journal of the Learning Sciences*, 7(3&4):313–350.

Krajcik, J., Simmons, P. E., and Lunetta, V. N. (1988). A research strategy for the dynamic study of students' concepts and problem solving strategies using science software. *Journal of Research in Science Teaching*, 25(2):147–155.

Krajcik, J. S. and Blumenfeld, P. C. (2002). *Project-Based Learning. From Cambridge Handbook of the Learning Sciences. Edited by Keith Sawyer.* Cambridge University Press, West Nyack, NY.

Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Particiaption (Learning in Doing: Social, Cognitive and Computational Perspectives).* Cambridge University Press, Cambridge, UK.

Lee, H.-S. (2003). *Scaffolding elementary students' authentic inquiry through a written science curriculum.* PhD thesis, University of Michigan.

Lee, H.-S. and Songer, N. (2004). Longitudinal knowledge development: Scaffolds for inquiry. *Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA*.

Lepper, M., Drake, M. F., and O'Donnel-Johnson, T. (1997). Scaffolding techniques of expert human tutors. In Hogan, K. and Pressley, M., editors, *Scaffolding Student Learning*. Brookline Books, Cambridge, MA.

Linn, M., Davis, E., and Eylon, B.-S. (2004). Specific design principles: Elaborating the scaffolded knowledge integration framework. In Linn, M., Davis, E., and Bell, P., editors, *Internet Environments for Science Education*. Lawrence Earlbaum Associates, Malwah, NJ.

Linn, M. and Songer, N. (1991). Teaching thermodynamics to middle school students: What are appropriate cognitive demands? *Journal of Research in Science Teaching*, 28(4):885–918.

Luchini, K., Quintana, C., Krajcik, J., and Soloway, E. (2003). Designed for learners: applying software engineering to the development of scaffolded educational software. *International Journal of Continuing Engineering Education and Lifelong Learning*, 13(3/4):366–386.

Mandinach, E. B. (1989). Model-building and the use of computer simulation of dynamic systems. *Journal of Educational Computing Research*, 5(2):221–243.

McClure, J., Sonak, B., and Suen, H. (1999). Concept map assessment of classroom learning: Reliability, validity and logistical practicality. *Journal of Research in Science Teaching*, 36(4):475–492.

McNeill, K., Lizotte, D., Krajcik, J., and Marx, R. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *The Journal of the Learning Sciences*, 15(2):153–191.

McNeill, K. L. and Krajcik, J. (2009). Synergy between teacher practices and curricular scaffolds to support students in using domain-specific and domain-general knowledge in writing arguments to explain phenomena. *The Journal of the Learning Sciences*, 18(3):416–460.

Merriam, S. B. (1998). *Qualitative research and case study applications in education.* Jossey-Bass, San Francisco.

Metcalf, S. J., Krajcik, J., and Soloway, E. (1999). Model-it: A design retrospective. In Kozma, M. J. and Robert, editors, *Advanced Designs for the Technologies of Learning: Innovations in Science and Mathematics Education.* Erlbaum, Hillsdale, NJ.

Miles, M. B. and Huberman, A. M. (1994). *Qualitative Data Analysis.* Sage Publications, Thousand Oaks, CA, second edition.

Novak, A. and Gleason, C. (2001). Incorporating portable technology to enhance an inquiry, project-based science classroom. In Tinker, R. and Krajcik, J., editors, *Portable Technologies: Science Learning in Context.* Kluwer, Netherlands.

Novak, A. M. and Krajcik, J. S. (2006). *Using Technology to Support Inquiry in Middle School Science. From Scientific Inquiry and the Nature of Science.* Springer-Verlag Publishers, Dordrecht, The Netherlands.

Novak, J. (1998). *Learning, Creating , and Using Knowledge: Concept maps as facilitative tools for schools and corporations.* Lawrence Erlbaum Associates, Mahwah, NJ.

Novak, J. D. and Gowin, D. B. (1984). *Learning how to Learn*. Cambridge University Press, New York.

Palincsar, A. S. (1986). The role of dialogue in providing scaffolded instruction. *Educational Psychologist*, 21(1&2):73–98.

Palincsar, A. S. (1998). Social constructivist perspectives on teaching and learning. *Annual Review of Psychology*, 49:345–75.

Palincsar, A. S. and Brown, A. L. (1984). Reciprocal teaching of comprehension fostering and monitoring activites. *Cognition and Instruction*, 1(2):117–175.

Papert, S. and Harel, I. (1991). *Constructionism*. Ablex Publishing Corporation.

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *The Journal of the Learning Sciences*, 13(3):423–451.

Penner, D. E. (2001). Cognition, computers, and synthetic science: Building knowledge and meaning through modeling. *Review of Research in Education*, 25:1–36.

Piaget, J. (1959). *The language and thought o the child*. Routledge & Keegan Paul, London, 3rd edition.

Puntambekar, S. and Kolodner, J. (2005). Toward implementing distributed scaffolding: Helping students learn science from design. *Journal of Research in Science Teaching*, 42(2):185 – 217.

Puntambekar, S. and Kolodner, J. L. (1998). The design diary: A tool to support students in learning science by design. *Proceedings of the International Conference of the Learning Sciences (ICLS 98)*, pages 35–41.

Quintana, C. (1999). Symphony: A case study in extending learner-centered design through process space analysis. In *CHI '99 - Conference on Human Factors in Computing Systems*, Pittsburg, PA.

Quintana, C. (2001). *Symphony: A Case Study for Exploring and Describing Design Methods and Guidelines for Learner-Centered Design*. Dissertation, University of Michigan.

Quintana, C., Carra, A., Krajcik, J., and Soloway, E. (2001). Learner centered design: Reflections and new directions. In Carroll, J. M., editor, *Human-Computer Interaction in the New Millennium*. ACM Press, New York, NY.

Quintana, C., Fretz, E., Krajcik, J., and Soloway, E. (2000). Assessment strategies for learner-centered software. In Fishman, B. and O'Connor-Divelbiss, S., editors, *Proceedings of the Fourth International Conference of the Learning Sciences*, Ann Arbor, MI. Erlbaum.

Quintana, C., Reiser, B. J., Davis, E., Krajcik, J., Fretz, E. B., Duncan, R., Kyza, E., Edelson, D. C., and Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *The Journal of the Learning Sciences*, 13(3):337–386.

Reimann, P., Thompson, K., and Weinel, M. (2007). Collaborative learning by modelling: Observations in an online setting. *Paper presented at the Australasian Society for Computers in Learning in Tertiary Education.*

Reiser, B. J. (2004a). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences*, 13((3)):273–304.

Reiser, B. J. (2004b). A theoretical note on concepts and the need for cyclic concept maps. *Scaffolding complex learning: The mechanisms of structuring and problematizing student work.*, 13(3):273–304.

Resnick, M. (1997). *Turtles, Termites, and Traffic Jams: Explorations in Massively Parallel Microworlds (Complex Adaptive Systems).* MIT Press, Boston, MA.

Roehler, L. R. and Cantlon, D. J. (1997). Scaffolding: A powerful tool in social constructivist classrooms. In Hogan, K. and Pressley, M., editors, *Scaffolding student learning: Instructional approaches and issues.*, pages 6–42. Brookline Books, Inc., Cambridge, MA.

Rogoff, B. (1990). *Apprenticeship in Thinking: Cognitive development in social context.* Oxford University Press, New York, NY.

Roschelle, J. (1992). Learning by collaborating: Convergent conceptual change. *The Journal of the Learning Sciences*, 2(3):235–276.

Roscoe, R. D. and Chi, M. T. (2007). Understanding tutor learning: Knowledge-building and knowledge-teling in peer tutors' explanations and questions. *Review of Educational Research*, 77(4):534–574.

Ruiz-Primo, M. and Shavelson, R. (1996). Problems and issues in the use of concept maps in science asessment. *Journal of Research in Science Teaching*, 33(6):569–600.

Safayeni, F., Derbentseva, N., and Caas, A. J. (2005). A theoretical note on concepts and the need for cyclic concept maps. *Journal of Research in Science Teaching*, 42(7):741–766.

Salomon, G. (1996). *Studying novel learning environments as patterns of change. In S. Vosiniadou, E. De Corte, R. Glaser & H. Mandl (Eds.). International Perspectives on the design of Technology Supported Learning.* Lawrence Erlbaum Associates, Mahwah, NJ.

Salomon, G., Globerson, T., and Guterman, E. (1989). The computer as a zone of proximal development: Internalizing reading-related metacognitions from a reading partner. *Journal of Educational Psychology*, 81(4):620–627.

Salomon, G., Perkins, D. N., and Globerson, T. (1991). Partners in cognition: Extending human intelligence with intelligent technologies. *Educational Researcher*, 20:2–9.

Scardamalia and Bereiter (1991). Higher levels of agency for children in knowledge building: A challenge for the design of new knowledge media. *Journal of the Learning Sciences*, 1(1):37–68.

Schwarz, C., Reiser, B., Davis, E., Kenyon, L., and et al. (2009). Developing a learning progression for scientific modeling: Making scientific moedleing accessible and meaningful for learners. *Journal of Research in Science Teaching*, 46(6):623–654.

Science, A. A. f. t. A. o. (1993). *Benchmarks for Scientific Literacy*. Oxford University Press, New York, NY.

Shrader, G., Lindgren, R., and Sherin, B. (2000). A pedagogically-oriented analysis of the modeling process. In *American Educational Research Association*, New Orleans, LA.

Singer, J., Krajcik, J., and Marx, R. (2000). The design and evaluation of classroom supports for seamless integration of a dynamic modeling tool. In Fishman, B. and O'Connor-Divelbiss, S., editors, *Proceedings of the Fourth International Conference of the Learning Sciences*, Ann Arbor, MI. Erlbaum.

Sins, P. H. M., Savelsbergh, E. R., and Joolingen, W. R. v. (2005). The difficult process of scientific modelling: An analysis of novices' reasoning during computer-based modelling. *International Journal of Science Education*, 27(14):1695–1721.

Spitulnik, M. W., Krajcik, J., and Soloway, E. (1999). Construction of models to promote scientific understanding. In Feurzeig, W. and Roberts, N., editors, *Modeling and Simulation in Science and Mathematics Education*, pages 70–94. Springer-Verlag, New York.

Stone, C. A. (1998). The metaphor of scaffolding: Its utility for the field of learning disabilities. *Journal of Learning Disabilities*, 31(4):344–364.

Stratford, S. J. (1996). *Cognitive and other conceptual activities in dynamic modeling: case studies of cognitive modeling activities in precollege classrooms.* Unpublished doctoral dissertation, University of Michigan.

Tabak, I. (2004). Synergy: A complement to emerging patterns of distributed scaffolding. *Journal of the Learning Sciences*, 13(3):305–335.

Tiberghien, A. (2000). Designing teaching situations in the secondary school. In Milar, R., Leach, ., and Osborne, J., editors, *Improving Science Education: The contribution of research.* Open University Press, Buckingham, UK.

Vygotsky, L. S. (1986). *Thought and Language.* The MIT Press.

Weglinsky, H. (1998). Does it compute? the relationship between educational technology and student achievement in mathematics. Technical report, Educational Testing Service.

Wilensky, U. (1999). *NetLogo [computer softwarte].* Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Wilensky, U. and Reisman, K. (2006). Thinking like a wolf, a sheep, or a firefly: Learning biology through constructing and testing computational theories – an embodied modeling approach. *Cognition & Instruction*, 24(2):171–209.

Wood, D., Bruner, J., and Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17:89–100.

Yin, Y., Vanides, J., Ruiz-Primo, M., Ayala, C., and Shavelson, R. (2005). Comparison of two concept-mapping techniques: Implications for scoring, interpretations and use. *Journal for Research in Science Teaching*, 42(2):166–184.

Zhang, B. (2003). *Exploring Middle School Science Students' Computer-Based Modeling Practices and Their Changes Over Time.* PhD thesis, University of Michigan.