ADAPTING A FRAMEWORK FOR ASSESSING STUDENTS' APPROACHES TO MODELING

By

Steven Carl Bennett

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of requirements
for the degree of

Curriculum, Instruction, and Teacher Education—Doctor of Philosophy

2017

ABSTRACT

ADAPTING A FRAMEWORK FOR ASSESSING STUDENTS' APPROACHES TO MODELING

By

Steven Carl Bennett

We used an "approach to learning" theoretical framework to explicate the ways students engage in scientific modeling. Approach to learning theory suggests that when students approach learning deeply, they link science concepts with prior knowledge and experiences. Conversely, when students engage in a surface approach to learning, they memorize without understanding and therefore do not connect new content to prior knowledge and experiences. In this study we modified an approach to learning framework in order to investigate the extent to which undergraduate science majors use a deep or surface approach when modeling. Twenty students enrolled in an introductory biology course participated in two think-aloud interviews, one year apart. The students engaged in a total of five modeling tasks, with each task having different contextual and conceptual familiarity based on course content. We determined student approaches to modeling by observing their use of metacognition, generative thinking, and causal reasoning. We observed that within-student and between-student approach-to-modeling scores differed as the modeling tasks changed, and we attribute the differences to prompt construction; specifically, the presence of cueing words, the type and amount of background information provided, and familiarity with the contexts and concepts addressed in the prompts. Similarly, prompt construction also influences the depth of engagement needed to construct a correct model; some students were able to construct correct models without using deep modeling approaches, while other students were unable to generate correct models despite using deeper approaches. We concluded this study with recommendations for prompt construction that may

encourage a diverse group of learners to engage more deeply in modeling tasks. The findings of this study support a theme common in the science education – the importance of helping students connect new information to prior knowledge and experiences, which not only improves learning, but also modeling.

TABLE OF CONTENTS

LIST OF TABLES vii

LIST OF FIGURES ix

CHAPTER 1: INTRODUCTION TO THE STUDY 1
  Connecting Modeling to Approaches to Learning 2
  Research Questions 4
  Overview of the Study 4
  Organization of this Document 5

CHAPTER 2: REVIEW OF THE LITERATURE 7
  Modeling in Science 7
    Types of models 7
    Students and modeling 9
  Approach to Learning Theory 10
  Merging Modeling and Approaches-to-Learning Frameworks 12
    Capturing the ways students approach modeling 13
  Documenting how Students Approach Modeling 18
    Assessing the components of approaches to learning 18
      Generative thinking 19
      Metacognition 19
      Causal reasoning 20
    Assessing approaches to learning 21
    Capturing generative thinking, metacognition and causal reasoning
      with verbal reports 22
      Problem solving 22
      Problem types 23
  The Importance of Documenting How Students Approach Modeling 23
    Generative thinking 23
    Metacognition 24
    Causal reasoning and explanations 24

CHAPTER 3: RESEARCH DESIGN AND METHODS 26
  Theoretical Framework 26
    Problem-solving 27
  Research Context 27
    Course 27
    Student participants 28
  Interview Process and Protocols 30
    Conducting the interviews 31
      Think-aloud protocols 32
    Interview protocols 34

Consent and orientation (interview #1) 34
Model construction (interview #1) 36
Consent and orientation (interview #2) 38
Model construction (interview #2) 38
Prompt selection 40
Coding Protocols 41
Generative thinking 42
Prior knowledge 43
Analogical transfer 43
Scoring 43
Metacognition 45
Causal reasoning 46
Approach-to-modeling score 48
Model correctness 49
Data Analysis 50
Descriptive statistics 50
Inferential statistics 52
Chi-square test of independence 52
One-way ANOVA with post-hoc analysis 52
*Spearman's rank correlat*ion 52
Hermeneutic content analysis 53
Triangulating data 54

CHAPTER 4: FINDINGS 55
Modeling Tasks Influence Student Approaches to Modeling 55
Same student, different scores 55
Factors That Influence Approach-To-Modeling Scores 57
Student factors 57
Generative thinking 58
Causal reasoning 59
Modeling tasks 60
Attributes of the modeling tasks – contextual familiarity 61
Attributes of the modeling tasks – written cues 62
Capturing students' Approaches-to-Modeling 63
Differentiating between students 65
Differentiation between modeling tasks 66
Proximity of interview to exam date 66
The relationship between deep approaches to modeling and scientifically correct models 69
Components of modeling approaches 73
Indicators of weak cognitive structure 74
The role of novel prompts 78
Capturing metacognition 79

CHAPTER 5: DISCUSSION OF THE RESULTS 82
Variation in Approach-to-Modeling Scores 82

Prompt construction ··· 83
Cueing words ··· 83
*Eliciting schemata for sm*ell ··· 85
Eliciting schemata for wolf ··· 86
Background information ··· 86
Context ··· 89
Student characteristics ··· 90
Cognitive structure ··· 91
Correctness of Student Models ··· 93
Capturing Differences in Student Approaches to Modeling ··· 94
Approach-to-modeling framework: Strengths ··· 95
*Metacognitio*n ··· 95
Generative thinking ··· 95
Causal reasoning ··· 97
Approach-to-modeling framework: Challenges ··· 97
Planning ··· 98
Evaluation ··· 99
Monitoring ··· 100
Assessing progress ··· 101
CONCLUDING THOUGHTS ··· 102
Designing Modeling Prompts ··· 102
Context ··· 103
Cueing words ··· 103
Supporting Deep Approaches to Modeling ··· 104
Generative thinking ··· 104
Causal reasoning / explanations ··· 105
Metacognition ··· 105

REFERENCES ··· 106

LIST OF TABLES

Table 3.1 Demographics of Students Enrolled in the Introductory Biology Course ···· 29

Table 3.2 Class Tritile Groupings ···· 30

Table 3.3 Demographics for Student Participants in the Study ···· 30

Table 3.4 Characteristics of the Modeling Tasks ···· 41

Table 3.5 Definitions and Examples of the Coding Protocols ···· 44

Table 3.6 Coding Protocol for Causal Reasoning ···· 48

Table 3.7 Descriptive Statistics Used in Data Analysis ···· 50

Table 3.8 Inferential Statistics Used in Data Analysis ···· 53

Table 4.1 Heat Map Showing Range of Approach-to-Modeling Scores by Prompt. Students (in rows) are ordered by descending cumulative approach-to-modeling score. Prompts (in columns) are arranged in the order that they occurred within the interviews. Tritile reflects a measure of students' academic achievement with Tritile 3 being students with highest GPA ···· 56

Table 4.2 Top (50th percentile) Approaches-to-Modeling: 5 Prompt Total ···· 60

Table 4.3 Bottom (50th percentile) Approaches-to-Modeling: 5 Prompt Total ···· 60

Table 4.4 Contexts Students Used to Model the Zika Prompt ···· 63

Table 4.5 Contextual and Written Cues Students Used to Model the Wolf Prompt ···· 64

Table 4.6 Range of Students' Approach-to-Modeling Component Scores ···· 66

Table 4.7a Grouping Students' by Modeling Approach Scores (binned) and Model Correctness Scores (binned) for all 5 Prompts ···· 69

Table 4.7b Grouping Students' by Modeling Approach Scores and Model Correctness Scores for the Smell Prompt ···· 71

Table 4.7c Grouping Students' by Modeling Approach Scores and Model Correctness Scores for the Wolf Prompt ···· 71

Table 4.7d Grouping Students' by Modeling Approach Scores and Model Correctness Scores for the CFTR1 Prompt ························ 72

Table 4.7e Grouping Students' by Modeling Approach Scores and Model Correctness Scores for the Zika Prompt ························ 72

Table 4.7f Grouping Students' by Modeling Approach Scores and Model Correctness Scores for the CFTR2 Prompt ························ 73

Table 4.8 Lowest Approach-to-Modeling Scores by Prompt ························ 75

Table 5.1 Familiarity of Modeling Tasks ························ 90

LIST OF FIGURES

Figure 2.1 Student-constructed model showing structures in boxes and behaviors on arrows. The overall function of the model is representing the origin of genetic variation and effects of genetic variation on phenotypic variation and fitness of an organism 9

Figure 3.1 Timeline showing the dates of the interviews and, for 2015, course content that coincided with the interviews 31

Figure 3.2 Think aloud practice activity: Students share thoughts while connecting matching pairs of mittens 35

Figure 3.3 Visuals used with the wolf modeling task 36

Figure 3.4 Background information on the CFTR gene that was given to students on the second exam 37

Figure 3.5 Question given to students on the second exam that contains vocabulary to include in their models 38

Figure 3.6 Background information about where and when the Zika virus spread. This information was given to students during the second round of interviews 39

Figure 3.7 Background information about the Zika virus given to students during the second round of interviews 40

Figure 4.1 Approach-to-modeling scores by tritile and prompt 58

Figure 4.2 Generative thinking scores by prompt and tritile ranking 60

Figure 4.3 Mean approach-to-modeling scores by prompt, presented in the order in which they occurred during the interviews 61

Figure 4.4 Range of approach-to-modeling scores by student and approach-to-modeling Components 65

Figure 4.5 Mean attribute scores by prompt 67

Figure 4.6 Differences in CFTR1 scores by the number of days between the second exam and the interview. CFTR was a question on the second exam 68

Figure 4.7 Comparison of CFTR scores between the second exam and the Interview ····· 68

Figure 4.8 Comparison of model correctness scores and metacognition ····· 74

Figure 4.9 Comparison of model correctness and modeling approach scores ····· 74

Figure 4.10 Comparison of model correctness and generative thinking scores ····· 75

Figure 4.11 Sarah's Smell model: Wind blowing scent from a donut shop to a person ····· 77

Figure 4.12 Simba's Smell model: Wind blowing scent from a donut shop to a person ····· 78

Figure 4.13 Relationship between smell and overall approach-to-modeling Scores ····· 79

Figure 5.1 Model showing the interactions between student characteristics, characteristics of the modeling prompt, the ways students approach the modeling prompt, and the resulting explanatory model ····· 82

Figure 5.2 Elizabeth's model explaining how the malformed vertebrae became present in the wolf population ····· 87

CHAPTER 1: INTRODUCTION TO THE STUDY

Modeling is a foundational practice in science education (Brewer & Smith, 2011; Coll & Lajium, 2011; National Research Council [NRC], 2012; Svoboda and Passmore, 2011). The act of modeling leads to a greater understanding of its role as a core practice in science, and helps learners better understand key disciplinary concepts (Schwarz, et al., 2009). Modeling can also engage students in a science practice similar to what scientists do (Namdar & Shen, 2015), and, as such, the Framework for K-12 Science Education (NRC, 2012) recognizes developing and using models as one of the eight core science practices that should be included in science education.

One way to engage students in modeling is through model-based instruction (MBI) (Clement & Rea-Ramirez, 2008; Crawford & Cullin, 2004; Dauer et al., 2013). MBI actively engages students in their own learning by promoting student construction, evaluation and modification of their mental models, (Kahn, 2011; Namdar & Shen, 2015). Louca and Zacharia (2012) make a distinction between MBI, which includes instructors using models to teach, and modeling-based learning (MbL), which emphasizes the role of the students in learning through modeling. Whenever I refer to MBI in the pages that follow, I am including the ideas contained within MbL.

Missing from the modeling research are aspects of how students learn through modeling, partly because of the need to observe students' cognitive processes (Louca & Zacharia, 2012; Rea-Ramirez, Clement, & Núñez-Oviedo, 2008). Louca and Zacharia (2012) called for future research that would investigate the learning process associated with modeling. That is the focus of my study – to look at the ways in which students use cognitive processes when engaged in modeling tasks.

Connecting Modeling to Approaches to Learning

Modeling is a type of problem solving (Crawford & Cullin, 2004; Megowan-Romanowicz, 2011), and just as students can approach problems in different ways, they can approach both learning and modeling in different ways. Marton and Säljö (1976) observed the different ways in which students read an academic article when told that afterwards they would answer questions about it. The students who used surface-level processing when reading the article tried to learn the text in order to reproduce it when answering questions about the article. Students who used deep-level processing tried to understand the material by comprehending the point the author was trying to make. This and similar studies led to the theory of approaches to learning, which distinguished between those students who focused on understanding information and those who focused on reproducing information.

Approaches to learning theory posits that students who use a deep approach connect new ideas to previous knowledge and relate concepts to each other and to everyday experiences (Chin & Brown, 2000). In contrast, when students use a surface-level approach to learning, they memorize without understanding and therefore do not connect new content to prior knowledge and experiences (Chin & Brown, 2000; Postareff, Parpala & Lindblom-Ylänne, 2015).

This study explores the extent to which this type of framework may apply to how students approach scientific modeling. Ideally, students will use deep-level processing when modeling by linking science concepts with prior knowledge and experiences (Buckley, 2000; Stratford, Krajcik and Soloway, 1998). I have however observed students using surface-level processing when modeling by memorizing the structure or words within a model through rote learning and reproducing it on a test. Novak and Cañas (2006) extend a cautionary note that information learned by rote still resides in long-term memory, the difference between deep and surface approaches is that rote learning leads to little or no integration of new knowledge with existing knowledge. Because modeling as a way of learning relies on students' abilities to make connections among concepts, approaches to learning theory may provide a framework for capturing the differences in how students engage in modeling.

Modeling and approaches to learning are aligned in several ways. First, when students approach learning deeply they link new information, ideas, and experiences together in the form of a mental model (Chin & Brown, 2000, 2000a; Louca & Zacharia, 2012; Postareff et al., 2015); and modeling is the process of constructing and externalizing mental models as "expressed models" that communicate interconnected knowledge (Louca & Zacharia, 2012). Secondly, creating models engages students in "combining isolated, fragmented, inert knowledge about poorly-understood concepts and relationships into larger, more clearly-understood constructs" (Stratford, Krajcik and Soloway, 1998, p. 216), which are qualities of a deep approach to learning. Finally, causal reasoning, metacognition and knowledge construction are all cognitive skills associated with modeling (Jonassen, Strobel and Gottdenker, 2005; Louca & Zacharia, 2012); they also are components of a deep approach to learning (Löhner, et al., 2005; Postareff et al., 2015; Sins, Savelsbergh & van Joolingen, 2005; White & Frederiksen, 1998; Wittrock, 1992).

Chin and Brown (2000) investigated the observable differences between deep and surface approaches to learning science in a grade 8 class engaged in a 9-week long chemistry unit. They grouped the qualitative differences into five categories: (a) generative thinking, (b) nature of explanations, (c) asking questions, (d) metacognitive activity, and (e) approach to tasks. Many of the attributes they observed in a deep approach to learning also support effective modeling, such as generative thinking, engaging in causal reasoning, and metacognition (Buckley, 2000; Coll &

Lajium, 2011; Justi & Gilbert, 2002; Löhner et al., 2005; Sins, Savelsbergh, van Joolingen & van Hout-Wolters, 2009; Svoboda & Passmore, 2011; White & Frederiksen, 1998).

Because scientific modeling can be viewed as type of domain-specific problem solving (Crawford & Cullin, 2004; Megowan-Romanowicz, 2011), the purpose of this study is to investigate the extent to which an approach to learning framework can be used to capture differences in students' approaches to modeling.

Research Questions

Since many of the attributes of approaches-to-learning are used in modeling, I operationalized the Chin and Brown framework to explain differences in the way students approach modeling in order to pursue the following research questions:

1) To what extent do different modeling tasks elicit deep approaches to modeling?
    a. To what extent do the different modeling tasks elicit the different components of deep approaches to modeling (i.e. generative thinking, metacognition, causal reasoning)?
2) What elements of a modeling task may account for differences in student use of deep approaches to modeling?
3) Does student use of deep approaches to modeling result in scientifically correct models? If not, what may account for the discrepancy?

Overview of the Study

Participants in this study were undergraduate students majoring in the life sciences at a Midwest land grant university, and enrolled in the second semester of introductory biology for life science majors. The student volunteers were solicited from a section of the course that was taught using model-based instructional methods. Students from this section were taught concepts through models and modeling, and demonstrated learning by constructing models. Students were introduced to a wide variety of models, including mathematical (Hardy-Weinberg), explanatory (genotype to phenotype), and teaching (chromosomes, DNA, genes, alleles) models; concept models using a "box and arrow" format were used most widely for assessments.

Approaches to learning have been assessed quantitatively with self-report survey instruments and qualitatively with interviews (Dinsmore & Alexander, 2012; Ifenthaler, Masduki & Seel, 2011). The modeling research presented here relies heavily upon think-aloud (qualitative) protocols while students were actively engaged in problem solving, a research method that externalizes students' cognitive structures and approaches to learning (Ericsson & Simon, 1980; Ifenthaler, Masduki & Seel, 2011).

Organization of this Document

I organized this document in 5 chapters. In Chapter 2: Review of the Literature, I present research literature that supports the study. The review includes (1) descriptions of cognitive theories that connect approach to learning theory with model-based learning, (2) definitions of frequently used terms, (3) an overview of models and modeling in science, (4) a more in-depth explanation of approaches-to-learning theory, and (5) detailed descriptions of the components of approach-to-modeling and the different methods used to observe these components.

In Chapter 3: Methods, I walk the reader through the research project beginning with a description of the participants and how they were selected. Next, I present a timeline for the research followed by an explanation as to how and why I chose to use a qualitative method for collecting data. I then describe an interview, introducing the interviewer script and the visual aids shown to students. Finally, I explain the coding protocols used to convert the qualitative data into quantitative data, and how I analyzed the quantitative data.

In Chapter 4: Findings, I present summaries of all the data to answer the research questions, and in Chapter 5: Discussion of Findings, I explain the results by tying them back to the research literature. I conclude chapter 5 with concluding thoughts in which I summarize the main findings in the research and how they can inform modeling-based learning.

CHAPTER 2: REVIEW OF THE LITERATURE

In the introductory chapter I described how this research connects approaches to learning theory with modeling. In this chapter, I will define both constructs, describe different types of models, and explain how novel problems can be used to assess the overlapping components that comprise approaches to learning and modeling.

Modeling in Science

Models are tools used in science to help communicate conceptualizations of phenomena that are unobservable due to their size, time span, distance, or complexity (Crawford & Cullin, 2004; Lehrer & Schauble, 2012). They enhance investigations and understanding (Harrison & Treagust, 2000) and can help to generate predictions and explanations (Schwarz & White, 2005). While there are many definitions of models in science (e.g. Buckley, 2000; Gilbert, 2004; Mulder, Lazonder & de Jong, 2015) I use the following definition of models in this study: A scientific model is a representation of abstract ideas and causal mechanisms (observable and unobservable) that explains or predicts phenomena (Oh & Oh, 2011; Schwartz et al., 2009; Windschitl, Thompson & Braaten 2008). Furthermore, the expression of mental models is an integral part of the modeling process (Kahn, 2008; Nersessian, 2010).

Types of models. Models are constructed for different purposes and therefore can be used in different ways. Svoboda and Passmore (2011) refer to models as a “suite of tools” (p. 8) because many types of models exist (Harrison and Treagust, 2000) including:

Conceptual models – External, coherent representations of mental models (Greca & Moreira, 2000). Gobert and Buckley (2000) refer to these as *expressed models*.

Consensus models – Conceptual models that scientists or groups of learners have developed, tested, and agree upon (e.g., Krebs cycle) (Coll & Treagust, 2003; Gobert & Buckley, 2000). Clement (2000) refers to these as *expert consensus models.*

Mental models – Personal, internal, cognitive interpretations that people use in reasoning and understanding daily phenomena (Buckley, 2000; Coll & Treagust, 2003; Crawford & Cullin, 2004; Gobert & Buckley, 2000; Harrison & Treagust, 2000; Kahn, 2011).

Concept-process models – Complex and abstract models that address processes (e.g., phase changes, motion, photosynthesis) (Bamberger & Davis, 2011).

Explanatory Models – Theoretical, hypothesized, qualitative models of hidden, non-observable processes that researchers devise to explain how and why phenomena occur (e.g., the behavior of gas molecules) (Rea-Ramirez, Clement & Núñez-Oviedo, 2008). Rather than a condensed summary of empirical observations, an explanatory model contributes new theoretical ideas.

Teaching models –Used to promote understanding of a target system (Coll & Treagust, 2003; Gobert & Buckley, 2000), which are referred to as *simulation models* when used to build understanding of cause-and-effect relationships (Crawford & Cullin, 2004; Harrison & Treagust, 2000).

Additional types of models include *iconic models* such as chemical formulas and equations; *mathematical models* such as the equation showing the relationship between force, mass and acceleration; and *theoretical models* such as the representations of electromagnetic lines of force (Harrison & Treagust, 2000).

In this study, I asked students to construct models, leaving the type of model up to the participants. However, given the nature of the modeling prompts, most students drew a particular type of conceptual or explanatory model that they learned in class based on Goel and Stroulia's (1996) Structure-Behavior-Function (SBF) theory. These concept models were comprised of structures (components, commonly nouns) placed in boxes and connected by behaviors (relationships or mechanisms, commonly verbs) placed on arrows to collectively represent the overall function of a system (Dauer, Momsen, Speth, Makohon-Moore & Long, 2013). The models resembled concept maps (Novak & Cañas, 2006), but differed in that (1) the concept models were not hierarchical; (2) relationships in the concept models were causal (mechanistic relationships) rather than associative; and (3) the concept-models represented only the mechanistic relationships that were relevant to a function rather than representing students' total domain knowledge (Dauer et al., 2013; Novak & Cañas, 2006). Figure 2.1 illustrates a concept-model showing the origin of genetic variation and how genetic variation affects phenotypic variation and fitness of an organism. Note how the boxes contain structures and the arrows are labeled with behaviors.

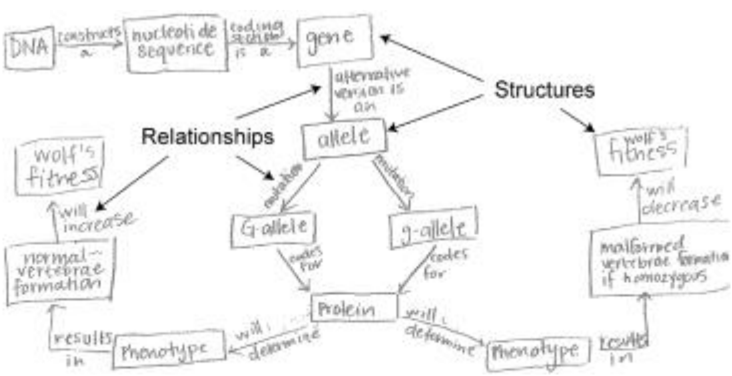


Figure 2.1 Student-constructed model showing structures in boxes and behaviors on arrows. The overall function of the model is representing the origin of genetic variation and effects of genetic variation on phenotypic variation and fitness of an organism (Dauer et al., 2013)

Students and modeling. Learning occurs when new information activates existing schema, resulting in the formation of a new schema, adjusting an existing schema, or reorganizing the schemata held in long-term memory (Ifenthaler, Masduki & Seel, 2011). In other words, learning involves an active search within long-term memory in which incoming information is reorganized and integrated with existing knowledge (Mayer, 1996).

Mental models are a type of schema (Derry, 1996), and learning is the process of adjusting mental models to accommodate new experiences (Louca & Zacharia, 2012). This is the theory behind model-based instruction (MBI) – students learn by constructing, evaluating and modifying their existing mental models (Kahn, 2011). Namdar and Shen (2015) articulate MBI as "students using, constructing and revising models to gain scientific knowledge and inquiry skills" (p. 993).

Model-based instruction has produced significant gains in student understanding of unobservable phenomena in science (Kahn, 2011) and its use is encouraged in education reform documents (e.g., NRC, 2012), partly because MBI allows students to actively participate in their own learning (Namdar & Shen, 2015). Additional benefits of MBI are helping students build subject matter expertise; and as a core practice in science, engaging students in modeling leads to more sophisticated understandings of key models in science (Schwarz et al., 2009). In the next section I explain the approaches-to-learning theory, and following that I will bring together ideas of modeling and approaches-to-learning.

Approach to Learning Theory

"Approach to learning" theory posits that students using a deep approach to learning seek to understand material by taking an interest in the subject matter, relating concepts to each other, connecting new ideas to previous knowledge, and relating new concepts to everyday experiences (Chin & Brown, 2000; Diseth & Martinsen, 2003; Lublin, 2003; Marton & Säljö, 1976). Conversely, students using a surface-level approach to learning memorize without understanding and therefore do not connect new content to prior knowledge and experiences. They are motivated by getting through the material expeditiously by memorizing information for assessments and using rote learning strategies (Chin & Brown, 2000; Diseth & Martinsen, 2003; Lublin, 2003; Marton & Säljö, 1976; Postareff, Parpala & Lindblom-Ylänne, 2015).

Marton & Säljö reported differences in students' "levels of processing" (pg. 7) and connected students' levels of processing (deep or surface level processing) to the way students approached learning (deep or surface), which may provide a framework for students in a MBI course. In other words, it may be possible to connect the way students process information to the way they approach modeling. Therefore, in this section I will provide a brief overview of approaches-to-learning theory and describe how components of the theory may apply to modeling.

Marton & Säljö (1976) categorize students' approaches to learning as operating at a deep-level or surface-level. With a deep approach to learning, students link ideas and build connections between new information and past experiences and understandings (Chin & Brown, 2000). In addition, students distill meaning from the content, pay attention to underlying meanings, access relevant prior knowledge, construct mental models, and generate analogies (Baeten, Kyndt, Struyven & Dochy, 2010; Warburton, 2003). The learner internalizes and personalizes the task (Chin & Brown, 2000 & 2000a) and constructs meaning from the content by relating ideas to each other (Postareff et al., 2015). In contrast, students who adopt a surface approach to learning tend to cope with course requirements; place more emphasis on rote-learning; perceive tasks as demands to be met; and memorize discrete facts, terms and procedures (Baeten et al., 2008; Dinsmore & Alexander, 2012; Warburton, 2003). Students who adopt a surface approach may correctly understand the material, but lack cognitive connections between the content and reasons for knowing the content beyond what is needed for a single use, usually an exam (Draper, 2009; Floyd, Harrington & Santiago, 2009).

Students using a deep approach to learning build multiple cognitive connections between the content and reasons for knowing the content, which allows them to use the concepts for multiple purposes. This is aligned with Wittrock's (1992) model of generative learning in which comprehension and understanding occur when the learner generates relations among concepts, and between new information and prior knowledge and experiences.

To perform well with a modeling task, Sins, Savelsbergh, Joolingen and van Hout-Wolters (2009) report that students need to engage in deep cognitive processing. They define deep cognitive processing as using prior knowledge for inductive reasoning and analysis during the modeling task, which is similar to the qualities associated with a deep approach to learning.

Merging Modeling and Approaches-to-Learning Frameworks

Modeling and approaches-to-learning theory are aligned in several ways. First, deep approaches-to-learning occur when learners link new information, ideas and experiences into an integrated cognitive structure (Louca & Zacharia, 2012; Postareff et al., 2015). In a similar fashion, students combine fragmented pieces of knowledge and draw upon their existing knowledge and prior experiences to construct models (Buckley, 2000; Stratford, Krajcik & Soloway, 1998). Secondly, in the process of modeling people formulate, test, revise, or reject mental models of phenomena and make use of analogies (Chin & Brown, 2000, 2000a; Louca & Zacharia, 2012). Similarly, when using a deep approach to learning students build mental imagery and generate analogies (Chin & Brown, 2000, 2000a). Thirdly, hypothesizing, inferring, conjecturing, causal reasoning and knowledge construction are all cognitive skills associated with modeling (Jonassen, Strobel and Gottdenker, 2005); they also are components of a deep approach to learning (Löhner et al., 2005; Louca & Zacharia, 2011; Sins et al., 2005; White & Frederiksen, 1998). Finally, modeling leads to learning and knowledge construction (Louca & Zacharia, 2012), which are frequently cited outcomes of a deep approach to learning (Postareff et al., 2015; Wittrock, 1992). Therefore, the high degree of overlap between the components of modeling and deep approaches to learning suggest potential for the approaches-to-learning framework to capture students' approaches to modeling.

Capturing the ways students approach modeling. Many of the characteristics of modeling listed previously (i.e., the use of analogies, mechanistic explanations, reflective thinking) are also indicators of deep approaches to learning. Chin and Brown (2000) empirically identified five elements that are indicators of one's approach to learning: (a) generative thinking, (b) nature of explanations, (c) asking questions, (d) metacognitive activity, and (e) approach to tasks. In this study, I use three of these components to define how students approach modeling tasks. I omit the "approach to task" category because that is what is being assessed (i.e., approach to modeling). "Asking questions" was similarly omitted because it was not as relevant in this study whereIwanted to capture students' thought processes with minimal interference or guidance from the researchers. A brief description of the remaining three elements follows, as reported by Chin and Brown (2000). In Chapter 3: Methods, I will provide more information about these three elements that make up whatIrefer to as an "approach-to-modeling."

***Generative thinking* –** The ability to generate an answer to an unfamiliar problem. It ecompasses creativity, lateral thinking and fluency in generating ideas.

***Nature of explanations* –** The depth and sophistication of student reasoning, ranging from a reformulation of a question to a microscopic causal explanation.

***Metacognitive activity* –** The degree to which students use comprehension-monitoring and evaluation strategies when engaged in a learning task.

These three elements contain attributes associated with effective modeling, including drawing on prior knowledge (generative thinking), reasoning by comparison or analogy (explanations and generative thinking), generating novel representations (generative thinking), constructing mechanistic explanations (explanations), and processing information (metacognition) (Louca & Zacharia, 2012; Mayer, 1998; Seel, 2003; Sins et al., 2009; Svoboda & Passmore, 2011). Therefore, these three elements will comprise an approach-to-modeling.

Löhner et al. (2005) found that students who participated in model-based inquiry spent most of their time engaged in orientation (which includes accessing prior knowledge and experiences), hypothesis formulation, and model evaluation, which are embedded in Chin & Brown's elements of deep approaches to learning. Similarly, Sins et al. (2005, 2009) analyzed student reasoning during a modeling activity in physics and reported that students who relied upon their physics knowledge and everyday experiences were the most successful at modeling. Finally, Stratford, Krajcik and Soloway (1998) reported that students' modeling practices included constructing causal relationships between components of their models and identifying the mechanisms underlying the causal relationships between the components. The modeling attributes reported in these three studies comprise many of the elements found in deep approaches-to-learning. In the following sections, I examine Chin and Brown's three approaches to learning categories through the lens of modeling.

(1) Generative thinking. Generative thinking refers to the ability to generate an answer when an immediate, ready-made solution to a problem is not available through simple recall or information learned by rote (Chin & Brown, 2000). Learners displaying generative thinking spontaneously construct plausible answers supported with specific examples, real life experiences, and self-generated analogies. They also possess well-integrated domain knowledge that allows them to generate hypotheses, elaborate upon and extrapolate ideas, and relate new and existing knowledge (Cavallo, 1996; Chin & Brown, 2000a; Duncan, 2007; Jonassen, 2000; Lee, Fradd, & Sutman 1995; Wong, 1993). Students displaying weak generative thinking extend little effort in solving novel problems, quickly discontinue their efforts, or give evasive responses that do not specifically answer the questions (Chin & Brown, 2000).

A person who can reason about novel and unfamiliar problems within a domain exhibits generative thinking (Duncan, 2007). Furthermore, a generative explanation will rarely be accurate because accuracy may require knowledge of intricate details, and if the details were known, the problem would not be unfamiliar and the explanation would be drawn from memory (Duncan, 2007). Duncan (2007) argues that for a generative explanation one must draw upon accepted theories and understandings in the domain to reason about unfamiliar problems.

Generating a model requires many of the same attributes. The process begins when learners encounter a problem that requires them to describe, predict or explain a phenomenon. Students then construct a model to analyze and solve the problem. In turn, the modeling process includes generating hypotheses about associative relationships, making predictions, using prior experiences, and drawing upon accepted theories and understandings in the domain (Duncan, 2007; Louca & Zacharia, 2011; Sins et al., 2005; White & Frederiksen, 1998).

(2) Metacognition. Metacognition, or the “knowledge of one’s own cognitive process” (Davidson, Deuser & Sternberg, 1996, p. 207), is composed of two closely associated activities – monitoring and regulation (i.e., control) (Anderson & Nashon, 2007; Anderson, Nashon & Thomas, 2009; Dinsmore, Alexander & Loughlin, 2008; Dunlosky & Metcalfe, 2009; Efklides, 2009; Kim, Park, Moore & Varma, 2013; Son, 2013). Monitoring is the awareness and evaluation of one's thinking (Kim et al., 2013). It is an assessment of one's knowledge, understanding, problem solving strategy, progress toward a solution, and accuracy of the solution. Regulation, on the other hand is a response to the monitoring process. For example, if monitoring reveals a lack of problem-solving progress, the regulatory function may indicate the need to try an alternative problem-solving strategy (Davidson, Deuser & Sternberg, 1996; Kim et al., 2013). Therefore, metacognitive monitoring and regulation are complimentary tasks (Son, 2013).

Goos (2002) distinguished between two types of monitoring; confirmatory monitoring and controlled monitoring. Confirmatory monitoring is passive, simply checking routine tasks such as alignment of digits in a math problem. Conversely, controlled monitoring is active. During controlled monitoring individuals actively look for "red flags" that will trigger regulatory processes. Examples of common red flags during problem solving include a lack of progress toward the goal, detecting errors in the implementation strategy, and detecting anomalous results in the final solution. During problem solving, individuals use metacognitive processes when identifying goals, identifying relationships among the elements in the problem and how they relate to the goal, developing a problem-solving strategy, and evaluating their solution (Davidson et al., 1996).

(3) Causal Reasoning. Modeling is a process of developing representations of underlying mechanisms that cause phenomena (Justi & Gilbert, 2002; Luca & Zacharia, 2011; White & Frederiksen, 1998); it attempts to describe how and why scientific phenomena happen. In other words, a model shows causality through mechanisms or processes (Bamberger & Davis, 2011).

Braaten & Windschitl (2011) list five philosophical models of scientific explanations that are relevant for science education. They are (1) the deductive–nomological (D–N) or "Covering Law" model, (2) statistical – probabilistic model, (3) causal model, (4) pragmatic model, and (5) explanatory unification model. The D-N model uses natural laws (regularities in the natural world) to account for phenomena, while the statistical – probabilistic model uses mathematical reasoning to account for events in the world. Unlike the D-N and statistical – probabilistic model, the central focus of the causal model is explanatory power, and the pragmatic model posits that the quality of an explanation is negotiated between those who are constructing the explanation. Finally, the explanatory unification model emphasizes an explanation that unifies seemingly disconnected events into a coherent relationship (Braaten & Windschitl, 2011).

While not always clearly defined, there is some consensus that an explanation is a causal account – grounded in scientific ideas and theories – that answers the question of "why" or "how" phenomena occur (Braaten & Windschitl, 2011; Lipton, 2004; NRC, 2012; Osborne & Patterson, 2011; Reiser, Berland & Kenyon, 2012; Russ, Scherr, Hammer & Mikeska, 2008; Treagust & Harrison, 2000; Trout, 2007).

Scientific reasoning can be considered sense-making in which students use causal accounts to connect evidence to claims in a progressive, logical fashion (Berland & Reiser, 2009; Braaten & Windschitl, 2011; Reiser, Berland, Kenyon, 2012; Russ et al., 2009). Some researchers refer to this as causal reasoning (Braaten & Windschitl, 2011), mechanistic reasoning (Russ et al., 2009), or causal mechanistic reasoning (Russ et al., 2008). In this study I use the Russ et al. (2008) causal/mechanistic conceptual framework to assess students' explanations, which is based on the work of Machamer, Darden and Carver (2000).

The causal/mechanistic framework connects a series of mechanisms from an initial starting point to produce observable phenomena (Machamer, Darden & Carver, 2000). The mechanisms consist of entities and activities, where entities engage in activities to produce change. One concern with this dual mechanistic view is reductionism. A reductionist approach takes causal accounts down to the atomic, sub-atomic, or energy levels (Machamer et al., 2000), but for any given domain, there is an accepted lower limit. The most commonly accepted lower limits within the life sciences are proteins, RNAs, DNA, and their activities (van Mil, Boerwinkel & Waarlo, 2013). Often activities such as DNA replication and protein synthesis are represented using mechanistic schema, or general descriptions of mechanisms that are comprised of known entities and activities, such as transcription and translation (Machamer et al., 2000).

Documenting how Students Approach Modeling

Approach-to-modeling, like approaches to learning, is a theoretical construct composed of several elements – generative thinking, metacognition and causal reasoning. In order to document changes in the way students approach modeling, I must first capture the three components that comprise a modeling approach. In this section I will present the different ways researchers have documented approaches-to-learning, including the pros and cons of the different assessments, and how they influenced our study design. I conclude this section with a review of the literature that informed our research methods and protocols for capturing generative thinking, metacognition and causal reasoning.

Assessing the components of approaches to learning. Researchers have developed many tools for capturing student reasoning, metacognition and generative thinking skills. Here I provide a brief overview of some of the more widely used assessment tools.

Generative thinking. Biggs and Collis' (1982) developed the Structures of the Observed Learning Outcomes (SOLO) taxonomy to assess the quality of student work. The taxonomy, which is closely tied to Piaget's stages of development, is based on four characteristics: capacity (the amount of working memory available for problem-solving); relating operation (the way a student response relates to the cue or question); consistency (forming conclusions supported by data); and structure (Biggs & Collis, 1982; Hattie & Brown, 2004). The SOLO taxonomy contains five hierarchical levels emerging from the four previous criteria. These levels, from lowest to highest, are: (1) prestructural, (2) unistructural, (3) multistructural, (4) relational and (5) extended abstract (Biggs & Collis, 1982; Hattie & Brown, 2004). The higher the taxonomic level, the more data and ideas are integrated to formulate the response (Hattie & Brown, 2004). Because SOLO is closely associated with problem solving and takes into account components of successful problem solving, such as structural knowledge, I used it as the basis for developing our generative thinking assessment protocol.

Metacognition. The tools used to measure metacognition fall into two broad categories – self-reports (Likert-type questionnaires and interviews), and objective behavior measurements (systematic observations and think-aloud protocols) that allow researchers to observe students' metacognition during problem solving (Akturk & Sahin, 2011). Unfamiliar and difficult questions that require higher-order thinking are most effective at eliciting metacognition when using objective behavior measurements (Kim et al., 2013).

Evidence of metacognition collected using think-aloud protocols can be analyzed quantitatively or qualitatively. Quantitatively researchers can count the number of times a participant demonstrates metacognition (Meijer, Veenman & van Hout-Wolters, 2006). For example, the participant may indicate monitoring or controlling using phrases such as "I've figured out what I wanted to say"; "I don't get it"; and "I didn't do that right, I'm getting confused" (Chin & Brown, 2000). The problem with counting metacognitive occurrences is the frequency of statements may bias results, giving higher ratings to participants who are verbose (Meijer et al., 2006). Qualitatively researchers can give each metacognitive activity a rating based on the depth of processing. For example, detecting an error would receive a lower rating than a detected error followed by a correction, or a detected error accompanied by an analysis of the cause of the error (Meijer et al., 2006).

Researchers can monitor both verbal and non-verbal communication for metacognition during think-aloud activities. Verbally, students may reveal comprehension monitoring using self-evaluative statements of their understanding; self-questioning when at an impasse; mentioning the problem and problem space; detecting and correcting errors; mentioning the limitations in their ideas; and noting anomalous data, counter-intuitive events and contradictory information (Chin & Brown, 2000a). Students may also verbalize the steps they are taking to solve the problem (Veenman, van Hout-Wolters & Afflerbach, 2006). Non-verbal indicators of metacognitive monitoring may be inferred from cognitive activities. For instance, rereading a portion of text, or pausing when looking over completed work. Although rereading is a cognitive activity, the decision to do so may indicate metacognitive monitoring or regulation (Meijer et al., 2006).

Causal reasoning. Russ et al. (2008) developed a causal/mechanical system for coding students' explanatory accounts of scientific phenomenon using the work of Machamer et al., (2000). The coding frame consists of nine categories and is described in greater detail in the methods section.

Assessing approaches to learning. Self-report questionnaires comprise the dominant method used to assess student approaches to learning *(*Dinsmore & Alexander, 2012*)*. Examples of self-report questionnaires include: The Study Process Questionnaire (SPQ) by Biggs (1987); the Revised Study Process Questionnaire (Biggs, Kember & Leung, 2001); the Course Experience Questionnaire (CEQ) – an indirect measure of approaches to learning based on student perception of content (Case & Marshall, 2009); Entwistle and Ramsden's (1982) Approaches to Studying Inventory (ASI); the Learning Approach Questionnaire (Chin & Brown, 2000); Tait, Entwistle, & McCune’s (1998) *Approaches and Study Skills Inventory for Students (ASSIST); Vermunt’s (1994) Inventory of Learning Styles in Higher Education (ILSHE); and the Lancaster Approaches to Studying Questionnaire, or LASQ (Ramsden, 1983).*

Dinsmore & Alexander (2012) argue that relying on self-report questionnaires presents two problems. The first being that students must be “meta-metacognitive” to report their own cognition, and the second being dichotomization of the data, that students are either deep or surface processers. Another limitation of inventory-generated data is “dissonance” in which students who report highly on a deep – surface inventory do not display any characteristics of the approach to learning identified on the survey (Case & Marshall, 2009).

Researchers who use in-depth interviews and other qualitative methods tend to conclude that approaches to learning are context-specific, while those who use survey questionnaires tend to view approaches to learning as consistent traits of the learner (Chin & Brown, 2000a). Postareff et al. (2015) noted that qualitative and mixed-methods research has been missing from research in the learning sciences since the early qualitative studies in the 1980’s, particularly studies using coded data collected from concurrent or retrospective interviews (Dinsmore & Alexander, 2012). Concurrent verbal reports or think-aloud protocols are collected as learners complete a task, while retrospective reports are collected following a task. Between 1983 and 2009, less than 15% of published approaches to learning studies used a coding scheme (Dinsmore & Alexander, 2012). Therefore, this study will rely upon concurrent verbal reports that are coded with a systematic and iteratively-designed rubric.

Capturing generative thinking, metacognition and causal reasoning with verbal reports. Generative thinking, metacognition, and causal reasoning are associated with structural knowledge, also known as cognitive structure, which consists of the integration of domain-specific concepts, the linking of pieces of knowledge and experiences, and the relationship among concepts in memory (Jonassen, 2000; Mayer, 2002; Newton, 2000; Wittrock, 1992). The more connections learners create between concepts, and the more they integrate new information with prior knowledge and experiences, the greater their structural knowledge (Newton, 2000; Wittrock, 1992). Structural knowledge can be assessed indirectly by looking for evidence of a coherent, mental structure as a learner engages in problem solving (Newton, 2000). This is often done using think-aloud protocols (Jonassen 2000; Newton, 2000).

Problem solving. Successful problem solving requires well organized and integrated content knowledge (Barnett & Ceci, 2002; Jonassen, Beissner, & Yacci, 1993; Mayer, 2002), and asking students to solve novel problems is a common method for distinguishing between students with well-integrated cognitive structures and students who are rote learners (Barnett & Ceci, 2002; Billing, 2007; Jonassen, 2000; National Research Council, 1999, 2000 & 2005). In addition, modeling is often used to solve problems; therefore, solving novel problems through modeling provides a means of assessing the way that students approach modeling (Cavallo, 1996; Chin & Brown, 2000; Davidson et al., 1996; Duncan, 2007; Jonassen, 2000).

Problem types. Problems may be classified according to the degree of transfer required to solve them. Rather than using the phrases "near" or "far" transfer, Bamberger and Davis (2011) speak of transfer-between-situations, and transfer-in-situations. The difference is contextual; when two problems occur in different contexts, they are considered transfer-between-situations, and it is up to the student to identify the differences and similarities between the challenges. Conversely, transfer-in-situations share the same context; connections between the problems are clear to the learner, and the learner will use similar knowledge, skills or practices to solve the problems. Using the language of Bransford in *How People Learn (*National Research Council, 2000), transfer-in-situations would be considered "near transfer" in which the knowledge and skills from one school task transfer to a highly similar task. Similarly, transfer-between-situations would be considered "far transfer" in which the knowledge and skills from one school task transfer to a non-school setting.

The Importance of Documenting How Students Approach Modeling

Generative thinking, metacognition and causal reasoning are factors that contribute to learning science. I will end this chapter with a brief discussion of how these components of modeling and deep processing of information support science learning by requiring the learner to draw upon prior knowledge and experiences and expand their cognitive structure.

Generative thinking. One desired outcome of science education is for students to use science knowledge generatively in order to solve problems and construct plausible explanations of phenomena (Duncan, 2007; NRC, 2000). When students engage in generative thinking they are able to construct mental models, meaningful relations among concepts, and meaningful relations between incoming information and prior knowledge and experiences (Wittrock, 1992), which expands their cognitive structure (Ifenthaler, Masduki & Seel, 2011). Leithwood, McAdie, Bascia & Rodrigue (2006) define “understanding” as a) understanding the part – whole relationships of content, b) making cognitive connections – establishing relationships and knowing how elements of content connect, and c) grasping concepts and principles well enough to apply the knowledge to new problems and situations. These attributes of understanding are aligned with generative thinking, which requires students to take in information, organize it, relate it to prior knowledge, and construct analogies between the new information and prior knowledge, which builds new cognitive connections and leads to deeper learning (Mayer, 2010).

Metacognition. Metacognition (i.e. self-regulation of learning) is an essential component of learning because it directs cognitive and affective activities, which improve student understanding of the subject matter and student transfer of learning to new situations (NRC, 1999; NRC, 2000; Hmelo-Silver, 2004; Vermunt & Vermetten, 2004). Metacognition is the active monitoring and regulation of the learning process and therefore involves students taking responsibility of their learning (Anderson & Nashon, 2007; Chin & Brown, 2000a; NRC, 2000; NRC, 2005; Son, 2013). In fact, metacognition is one of the distinguishing features between experts and novices, with experts monitoring their understanding and recognizing when additional information is needed to improve their understanding (NRC, 1999). Learning is the process of constructing and reconstructing cognitive structures in response to new experiences; metacognition facilitates the learning process through an awareness of comprehension difficulties, and an awareness of how new knowledge relates to or challenges prior knowledge (NRC, 2005; Williams & Lombrozo, 2010).

Causal reasoning and explanations. Engaging in causal reasoning and constructing explanations requires learners to draw upon prior knowledge and experiences and thereby facilitates the building of well-integrated declarative knowledge (Deans for Impact, 2015;

Hmelo-Silver, 2004; Newton, 2000; Nokes-Malach & Mestre, 2013; NRC, 1999). Causal reasoning also enhances learning when students use self-explanations (also a metacognitive process) to check their understanding (Deans for Impact, 2015; NRC, 2005; Williams & Lombrozo, 2010).

The process of reasoning and explaining may facilitate knowledge construction in two different ways. First, by having learners interpret what they are learning in terms of unifying patterns, they identify generalizations or "unifying regularities" (Williams & Lombrozo, 2010). Second, when constructing an explanation, the learner integrates new information with existing knowledge and adjusts their mental models accordingly (Williams & Lombrozo, 2010). Legare (2014) notes that relatively little is known about how explaining affects learning, but one possibility is that engaging in explanation allows the learner to selectively search for causal mechanisms within their cognitive structure and formulate generalizations. Lombrozo (2006) also notes that the benefits of explanations on learning are well established, but the mechanisms underlying the relationship are not well known. According to Lombrozo, constructing an explanation requires learners to draw upon prior knowledge and construct general patterns and then isolating and applying aspects of the general pattern to a new situation.

CHAPTER 3: RESEARCH DESIGN AND METHODS

This research is aligned with cognitive constructivism, so I will begin by juxtaposing the research design with the theoretical framework. Next, I will present the context in which this research is situated, including the sample population and sample selection procedures. After that I will discuss the interview process, which includes the protocols and reasoning behind the interview prompts selected. Finally, I will present and explain how I developed and employed coding protocols and analyzed the data.

Theoretical Framework

This study is situated in approaches to learning theory, which states that students may process information at a deep or surface level when engaged in a learning activity (Case & Marshall, 2009; Dinsmore & Alexander, 2012; Marton & Säljö, 1976). The main differences between a deep and surface approach to learning include:

Deep Approach:

- Relating ideas to previous knowledge and experience
- Looking for patterns and underlying principles
- Checking evidence and relating it to conclusions
- Examining logic and argument cautiously and critically
- Memorizing whatever is essential to understanding
- Monitoring understanding as learning progresses
- Personalizes learning tasks, making it meaningful to their own experiences and to the real world.

Surface (reproducing) Approach:

- Viewing the course as unrelated bits of knowledge
- Routinely memorizing facts (a.k.a mechanical memorization) and carrying out procedures – reproducing surface aspects of the task such as words used or a diagram.
- Seeing little value or meaning in the course or course tasks
- Studying without reflecting on either purpose or strategy.
- Views learning tasks as demands to be met
- Avoids personal or other meanings of a task

Problem-solving. When engaged in problem-solving, people construct mental models by mapping existing images and concepts held in long-term memory on to the current problem. In addition, the ability to solve problems is associated with the number of connections tied to concepts held in long-term memory – the more connections the greater the likelihood of accessing an existing schema that will aid in problem-solving.

Observing students' thoughts while solving novel problems can reveal the complexity of their cognitive structures, and having students draw models can reveal their mental models. Therefore, this study makes use of think-aloud protocols while students are engaged in modeling tasks and requires students to construct models of novel problems.

Research Context

In this section I describe the setting or context in which I conducted the research. I will first introduce the course from which I selected the student participants, and then describe the student participants. The research presented here was embedded within a larger project designed to document model-based performance for a cohort of students through 3 interviews spread over three years.

Course. We worked with students enrolled in the second semester of a two-semester introductory biology course (enrollment of 194) required for students majoring in the life sciences. In the 15-week long course, students were introduced to genetics, evolution, and ecology through model-based instruction. The conceptual theme of the course was genetic variation, specifically (1) the origin of genetic variation – that mutations are the basis for all variations within and among species; (2) the expression of genetic variation – that genetic variation interacts with the environment to produce traits; and (3) the consequences of genetic variation in terms of species – that the expressed genetic variation interacts with an organisms' environments and affects the fitness and persistence of the species.

Students were shown many examples of the big idea (genetic variation) through various cases (scenarios) such as bacterial resistance to antibiotics. To facilitate modeling-based learning (MbL) (Louca & Zacharia, 2013) students engaged in modeling exercises designed to strengthen their understanding of the relationships between genetic makeup (genotype) and protein production (phenotype). These modeling exercises were also case-based.

On tests, homework, and quizzes, students were given multiple transfer-between-situations (Bamberger & Davis, 2011) and asked to show, in a single model, (a) how variation arises in a genome; and (b) how variation at the molecular level affects traits expressed by organisms. (As a reminder, transfer-between-situations involves use of the same concept, but set in different contexts.) In their models, students were instructed to include 7 entities (components): allele, chromosome, DNA, gene, nucleotide sequence, phenotype and protein. Students received feedback through online rubrics, whole class discussions, and organized group activities.

Student participants. Course enrollment was predominantly students majoring in the life sciences: 91% in the life sciences (e.g., human biology, neuroscience, kinesiology), 5% in the social sciences (e.g., psychology, anthropology), and 4% in engineering. Student demographic data are shown in Table 3.1.I used an open solicitation process to select thirty student volunteers from one section of the introductory biology course (N = 194) to participate in the study. Volunteers were binned into tritiles determined by students' GPA upon enrollment in the course, and the first 10 students from each tritile were selected to participate (n=30). Tritile ranges are listed in Table 3.2. This study was embedded within a larger project working with a cohort of students engaged in 3 interviews over 3 years. Twenty of the original 30 students returned for the second interview, and results of the two interviews with these 20 students are reported here. Student volunteers were compensated for each interview.

Table 3.1 Demographics of Students Enrolled in the Introductory Biology Course

Ethnicity	**Course Enrollment (#)**	**Course Enrollment (%)**
American Indian/Alaskan Native (non-Hispanic)	1	0.5
Asian (non-Hispanic)	12	6.2
Black or African American (non-Hispanic)	9	4.6
Hispanic Ethnicity	5	2.6
International	12	6.2
Not Reported	2	1.0
Two or more races (non-Hispanic)	4	2.1
White (non-Hispanic)	149	76.8
Gender		
Male	83	43
Female	111	57
Class Rank		
Undergraduate - Freshman	8	4
Undergraduate - Sophomore	102	53
Undergraduate - Junior	65	33
Undergraduate - Senior	19	10

Table 3.2 Class Tritile Groupings

Range	Grade Point Avg.
Upper Third	4.00 – 3.576
Middle Third	3.575 – 3.155
Lower Third	3.135 – 1.667

The demographics of the students who participated in the study are listed in Table 3.3, which shows that 80% of the participants were sophomores, 85% white (non-Hispanic), and 65% female. Table 3.3 also shows that students came from diverse majors, and that 45% of the students were from the top tritile, 25% from the middle tritile, and 30% from the bottom tritile.

Interview Process and Protocols

We conducted two rounds of interviews with 20 volunteers to observe their thought process while constructing models. In this section I will explain the interview process, including the interview protocols and the equipment used to collect students' models and their thoughts while constructing models.

Table 3.3 Demographics for Student Participants in the Study

Pseudonym	GPA	Tritile	Gender	Ethnicity	Rank	Major
Rose	4.00	3	F	White	Sophomore	Human Biology
Leah	4.00	3	F	White	Sophomore	Environ Sci & Sust
Paul	3.95	3	M	White	Sophomore	Human Biology
Elizabeth	3.94	3	F	White	Sophomore	Pre-Veterinary
Billy	3.90	3	M	White	Sophomore	Biomed Lab Sci
Taylor	3.87	3	F	White	Sophomore	Human Biology
Omid	3.77	3	M	White	Junior	Neuroscience
Sandra	3.77	3	F	White	Junior	Neuroscience
Rick	3.68	3	M	White	Junior	Human Biology
Sarah	3.50	2	F	White	Sophomore	Environ Sci & Sust
Merle	3.40	2	M	White	Sophomore	Pre-Medical
Julia	3.34	2	F	White	Sophomore	Pre-Medical
Sydney	3.28	2	F	White	Sophomore	Anthropology
Brian	3.27	2	M	White	Sophomore	Neuroscience
Simba	3.12	1	F	White	Sophomore	Bioengineering

Table 3.3 (cont’d)

Anna	3.07	1	F	White	Sophomore	Pre-Medical
Sheri	2.94	1	F	White	Sophomore	Environ Sci & Sust
Rachel	2.81	1	F	Asian	Sophomore	Neuroscience
Eric	2.71	1	M	Asian	Junior	Human Biology
Samantha	2.48	1	F	African Amer.	Sophomore	Biochemistry / Molecular biology

Conducting the interviews. We conducted two sets of 60-minute think-aloud interviews. The first interviews took place following their second exam in October of 2015 while students were enrolled in the course. As I will discuss later, portions of the script for the first set of interviews were drawn from material on the second exam. The second round of interviews were conducted in October of 2016, one year following the first interview. Figure 3.1 is a timeline showing when the interviews took place, and the content covered in class preceding the first interview.

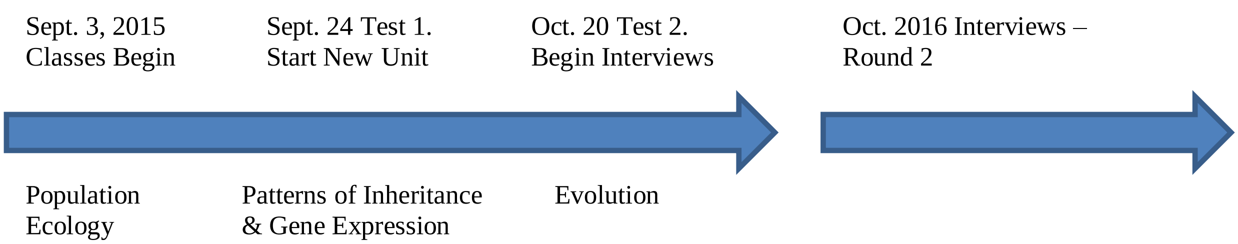


Figure 3.1 Timeline showing the dates of the interviews and, for 2015, course content that coincided with the interviews

Two researchers were present at each interview; one serving as the interviewer and the other as the scribe. Both jotted field notes during the interviews and discussed major observations and themes following each interview. A technician was nearby, but not in the room, to assist with technology, which included digital audio and video recording devices, microphones, and a Promethean board (a.k.a. Smart board or digital white board). During the interviews, I asked student participants to construct models on the Promethean board, which allowed us to capture their models electronically. I also asked students to share their thoughts while modeling, which I captured using the video and audio recorders. The use of the Promethean board along with video and audio recordings allowed us to synchronize student comments with their drawings, much like pairing a voice to the drawing activities on an Etch-A-Sketch.

Think-aloud protocols. Previous research assessed approaches-to-learning quantitatively using self-report survey instruments and qualitatively using interviews (Dinsmore & Alexander, 2012; Ifenthaler et al., 2011). The modeling research presented here relied upon think-aloud (qualitative) protocols, a research method that externalizes students' cognitive structures and approaches to learning (Ericsson & Simon, 1980; Ifenthaler et al., 2011).

Think-aloud protocols reveal students' cognitive processes while they perform a task (Goos & Galbraith, 1996; Ericsson & Simon, 1980). Cognitive psychologists use verbal data frequently, but also caution that there are limitations (Anderson, Nashon & Thomas, 2009; Goos & Galbraith, 1996), including:

- Students not fully articulating their cognitive processes (Goos & Galbraith, 1996).
- Environmental influences (stress, researcher intervention, task demands) that may affect cognitive processing (Goos & Galbraith, 1996).
- Students may verbalize cognitive processes that do not correspond to the observed behavior (Goos & Galbraith, 1996), although this is uncommon (Ericsson & Simon, 1980).

To minimize environmental influences and make students feel at ease, the interview room had minimal distractions and upon arrival the interviewers greeted the students warmly – engaging them in light, fun conversation, and offering refreshments and a comfortable seat (van Someren, Barnard & Sandberg, 1994). While it is possible for researchers to influence students' metacognition (Anderson, Nashon & Thomas, 2009), I minimized the risk by intervening as little as possible while students were modeling, and no personnel associated with the class participated in the interview process (Van Someren, Barnard & Sandberg, 1994). I also emphasized that there were no correct or incorrect answers; that I wanted to know the students' thoughts and therefore encouraged and reminded them to share their thoughts rather than narrate their modeling activities (Ericsson & Simon, 1993; Van Someren, Barnard & Sandberg, 1994). When students stopped talking I would remind them to "keep talking" or ask, "what are you thinking?"

Another concern with interviews is how the students frame, or view, the purpose of the interview. Russ, Lee and Sherin (2012) reported that students framed science interviews as inquiry, oral examination or expert, and switched frames during the interviews. When framing the interview as inquiry, students perceived their roles as constructing an explanation when they could not provide an immediate answer. When framed as an oral examination, students believed they were expected to clearly and concisely deliver a correct answer. Finally, when framed as experts, students viewed their job as discussing their own thinking. Interviewers can influence students' frames, beginning with the introductory comments about the purpose of the interview.

Following the examples that Russ, Lee and Sherin provided, I sent tacit and explicit messages to the students that their roles in the interview were that of experts and to assume a position of inquiry because I wished to capture their cognitive processes. Therefore, the introductory comments included phrases such as "please don't worry about right or wrong answers," "it's okay to not know the answer," and "we are interested in your thinking and therefore I expect you to talk a lot."

Interview protocols. In this section I will take you through an interview while also giving you a “behind-the-scene” look at the rationale for each component in the protocol.

Consent and orientation (interview #1). Upon each student's arrival, the interviewer provided a brief description of the study and its purpose. Students were asked to read an informed consent form, and to sign and date the form if they wished to participate in the study. After informing the students that the interview would be videotaped and shown where the video equipment was located, I provided an overview of how to use the touch screen (Promethean board) and software. Students were invited to 'play' with the technology to become comfortable with it by drawing a picture of their choice. Throughout the orientation students were allowed to ask questions about the interview and videotaping. Here is what the interviewer said:

> *Read the consent forms and I am glad to answer any questions you may have before you agree to participate. Once you have signed the consent form I will give you a copy for your records. During the interviews, we will audio and video record you with the devices over there. Do you have any questions about the interview or videotapes?*
>
> [After filling out consent form]
> *We are doing research in science education to study how students use models when they are learning biology. Today we are going to ask you to make some models of different scenarios. While some of these cases may be familiar to you, others may be unfamiliar. Try your best for all the scenarios. We are interested in how you think about and use models, so don’t worry if you are having trouble explaining some things during the interview.*
>
> *Before we begin the interview, I want to remind you that your participation is voluntary, so you can skip any question or leave the interview at any time without consequence. Do you have any questions for me now?*
>
> *Also, I’d like you to feel comfortable with the idea that we will not be using your name to identify you. Could you pick a fake name, which I can use to store the recordings? What name would you like me to use?*
>
> *Great, so [pseudonym] tell me a little bit about yourself…Where are you from?... What year are you?...What’s your major?*

Following the consent procedure and introductions, the interviewer proceeded to the orientation script:

Thank you, now we will move on to the next part of the interview. Have you had experience using a Smart board like this before?

That's okay, it's pretty simple. Have you ever used a tablet device for drawing? It is similar. What I would like you to do is grab one of the pens and I will have you practice drawing and writing on the board.

*Practice drawing some shapes and writing in them. Use the **pen tool** to create the shapes.*

*Okay, now **create a new page**. Note the **page numbers** in the top right corner. Practice drawing 3 lines of varying widths using the **width slider** on the right. Please, add a line that is a different color than what you used before.*

*Now, use the **Undo Tool** to remove the last line you drew. Then use the **Eraser Tool** to erase a small portion of the thickest line. Use the **width slider** to enlarge the Eraser area and completely remove the thickest line. Use the **select tool** to move the top line to the trash bin in the lower corner.*

*Now, use the **Undo Tool** to remove the last line you drew. Then use the **Eraser Tool** to erase a small portion of the thickest line. Use the **width slider** to enlarge the Eraser area and completely remove the thickest line. Use the **select tool** to move the top line to the trash bin in the lower corner. Now **Scroll** using the vertical slider on the right.*

To help the students get comfortable with thinking aloud while modeling, I had them complete the following activity, which you will see in Figure 3.2:

Go to the next page. Now we're going to practice thinking aloud. Going forward in this interview, we would like you to narrate your thoughts while you are performing activities. To practice thinking aloud, please try to find the matching pictures on the page. Again, we want you take us through your thought process.

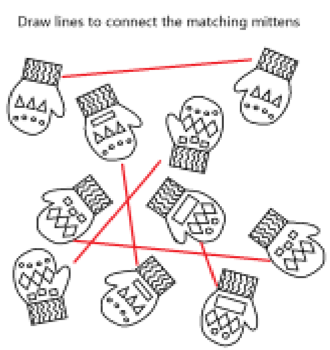


Figure 3.2 Think aloud practice activity: Students share thoughts while connecting matching pairs of mittens

http://www.makinglearningfun.com/Activities/snow/JacketIWearIntheSnow/JacketSnow-MittenMatchWorksheet.gif

Great job; Let's move on to the next portion of the interview. Again, I want to remind you that we are interested in the way that you build and think about models. We are not going to focus on the correctness of your explanations, so don't worry if you don't know some things, just do your best.

Model construction (interview #1). I began this portion of the interview by asking students to share their views of models. For example, the interviewer asked, "Tell me in your own words what you think a model is," "What purpose do you think models play in biology and science," and "In your opinion, what qualities make a scientific model effective?" These questions pertained to another set of research questions from the larger project, and I don't use the responses for this study. Following this discussion, we engaged students in three modeling tasks. First, I will present the modeling prompts in the order they appeared during the interview, and then I will explain why we selected these particular prompts.

Smell. *Using the smart board, I would like for you make a model to explain how a person can smell things from a distance. Again, I want to remind you to talk aloud as you construct your model.* (Shwartz, Weizman, Fortus, Krajcik, & Reiser, 2008).

Wolf. I*n class, you learned about Isle Royale located off the northern shore of Michigan's Upper Peninsula. The island, which was previously uninhabited by wolves, had its first introduction to wolves in 1950. Researchers tracked the wolf populations at Isle Royale, and after several years into the study, they noticed the appearance of malformed vertebrae (backbones) in some of the wolves.*

Visual aids that accompanied the wolf modeling tasks are shown in Figure 3.3.

Image A, on the top, shows a pack of wolves.
Image B, on the left, shows a normal wolf vertebra.
Image C, on the right, shows a malformed vertebra.

Figure 3.3 Visuals used with the wolf modeling task

Modeling Task: *Construct a model to explain how the malformed vertebrae became present among the wolf population on Isle Royale.*

CFTR1. *Thank you. Now we are going to move forward to the final part of the interview. We're going to construct another model using a prompt from the last test. Don't worry if you feel like your model doesn't match with what you were taught, we are interested in the process of how you build models.* [Share test prompt without the structures].

Modeling Task: *Using what you know about how genetic information is organized and expressed, construct a model that explains:*
a) how genetic variation originates at the CFTR gene, and,
b) how genetic variation ultimately results in expression of a trait (normal vs. cystic fibrosis)

"Smell" was designed to be a novel prompt – a prompt that would most effectively reveal students' cognitive structures. "Wolf" was a transfer-between scenario (same concept but different context) that embodied the genetic variation information from class, but in a different context. Finally, "CFTR" was a transfer-in (same concept and context) question because the content and context were familiar to the students – they had seen the same question on the current exam, and they had modeled a similar question for homework. Unlike the exam, however the students did not receive background information on CFTR, which is presented in Figure 3.4.

Case: Cystic Fibrosis [pts]
Cystic fibrosis is the most common lethal inherited disease in Caucasian populations and is caused by defects in the CFTR gene located on Chromosome 7. In healthy individuals, the wild-type allele (F) is dominant and contains the information necessary for producing normal CFTR protein.

Normal CFTR protein forms a channel that passes through cell membranes and regulates the movement of chloride ions out of cells. As chloride leaves cells, water follows and thins the mucus on cell surfaces, allowing it to flow freely. In individuals with cystic fibrosis, the CFTR proteins are defective and chloride ions and water build up inside cells. The inability to regulate chloride and water results in a drier mucus that is thick and sticky and accumulates on cell surfaces in the lungs, pancreas, digestive tract, and other internal organs. Individuals with cystic fibrosis experience frequent and serious bacterial infections, are unable to absorb adequate nutrients, and have chronic respiratory problems. If untreated, children with cystic fibrosis generally die before 5 years of age. However, daily chest pounding to clear mucus, along with heavy doses of antibiotics and other therapies have extended life expectancy for cystic fibrosis patients into their 20's and 30's.

Figure 3.4 Background information on the CFTR gene that was given to students on the second

In addition, students did not receive any vocabulary words to include in their models as they did on the exam (Figure 3.5).

> Using what you know about how genetic information is organized and expressed, construct a box-and-arrow model ***with drawings*** that explains:
>
> *a) how genetic variation originates at the CFTR gene*, and,
>
> *b) how genetic variation ultimately results in expression of a trait (normal vs. cystic fibrosis).*
>
> Include the following concepts in your response, but modify them to make them specific to this case: allele, chromosome, DNA, gene, nucleotide sequence, phenotype, protein.

Figure 3.5 Question given to students on the second exam that contains vocabulary to include in their models. Note: the interview did not include the vocabulary or the statement, "Include the following concepts in your response…"

Consent and orientation (interview #2). The consent and orientation phase was much shorter for the second round of interviews. After students read and signed the consent forms and were reminded of their fictitious research names, I initiated a short conversation to learn what types of biology or model-based courses they had taken since we last met. Some students discussed the biology they learned during internships and summer jobs. Here is the introductory portion of the interview script:

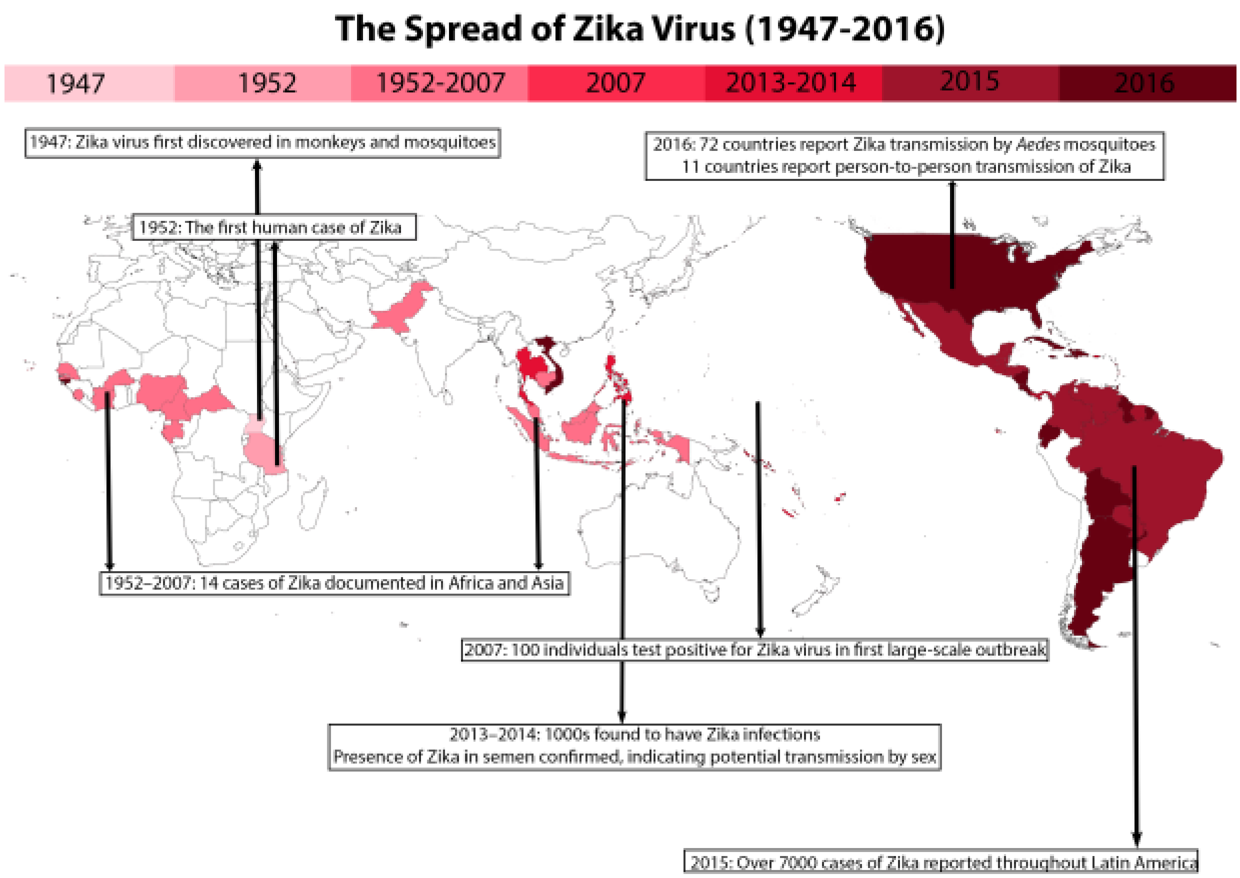
> *Welcome back, thank you for meeting with us again. How is your semester going? Have you taken any biology courses since Bio 162?*
>
> *As you know from the last interview, we are interested in how students learn about models in biology. Today, we have some additional tasks. Like last time you will use the smartboard. Take a minute to get familiar with it again by using the pen, cursor, and eraser tools.*

Model construction (interview #2). I engaged the students in two modeling tasks. The first, Zika, was a new prompt and the second, CFTR, the students had seen during the first interview. Because the students were most likely unfamiliar with Zika, we provided them with the background information shown in Figure 3.6 and Figure 3.7.

Zika virus. *I'm going to give you some information to read about Zika, a disease has been in the news, and then, once you are done reading, we will prompt you to make a model.* [hand student the prompt]
If you'll notice, you have a map of the world here - from Africa on the left to North and South America on the right. And there's a color code here [point to colored bar at top] *that corresponds to a timeline.*

The lightest color [point to map, location of Zika forest] *corresponds to 1947, when Zika was first discovered; the darkest color represents the current time (2016). Are you ready to proceed?*

[Handout #1]



Map of Timeline and Spread based on:
http://who.int/emergencies/zika-virus/zika_timeline.pdf?ua=1
http://who.int/emergencies/zika-virus/situation-report/1-september-2016/en/
http://who.int/emergencies/zika-virus/zika-historical-distribution.pdf?ua=1

Figure 3.6 Background information about where and when the Zika virus spread. This information was given to students during the second round of interviews

[Handout #2]

Virus
"Zika" refers to both a disease (Zika disease) and the virus responsible for causing it (Zika virus). Both are named for the Zika Forest in Uganda where the virus was first discovered. After its discovery, research revealed that the Zika virus consists of a single-stranded RNA genome contained within a protein shell. But, the virus depends on transmission primarily through the bite of mosquitoes belonging to the *Aedes* genus, which are unique because they are active during the day.

Symptoms
Until recently, risks associated with Zika infection were thought to be minimal. Individuals experienced symptoms similar to those of a mild cold or flu. These symptoms are so mild that infected individuals rarely seek medical care and feel better within a few days. However, evidence links Zika to microcephaly in babies born to women infected during pregnancy. "Microcephaly" refers to a rare birth defect in which incomplete brain development results in an abnormally small brain and head size.

Transmission
Although transmission is primarily due to bites from mosquitoes carrying the virus, recent clinical tests confirmed that Zika can also be transmitted person-to-person through childbirth, blood transfusions, and sexual contact.

Figure 3.7 Background information about the Zika virus given to students during the second round of interviews

Modeling Task: *Using what you know about the Zika virus, construct a model that explains how and why Zika has spread since it was first discovered.*

CFTR2. This is a repeat from the first interview.
Modeling Task: *Using what you know about how genetic information is organized and expressed, construct a model that explains:*
a) how genetic variation originates at the CFTR gene, and,
b) how genetic variation ultimately results in expression of a trait (normal vs. cystic fibrosis)

"Zika" was designed to be a transfer-between scenario, similar to the "Wolf" prompt in the first interview. "CFTR" was repeated to understand the extent to which students' models may change during a year-long interval.

Prompt selection. In addition to using prompts that represented novel, transfer-in, and transfer-between problems, I selected prompts that varied in terms of conceptual and contextual familiarity. The differences in these modeling tasks will allow us to answer the question, "Do students' approach-to-modeling scores change as the modeling tasks change?" The rationale for selecting these modeling tasks (prompts) is shown in Table 3.4, which lists the contexts in which students may have encountered the concepts addressed in the different modeling prompts.

We designed the Zika prompt to align with the CFTR and Wolf prompts keeping “Smell” as the novel modeling task. But as it turns out, Zika may be the only modeling prompt that students were unfamiliar with, not having seen it in class or daily activities. Furthermore, while genetic variation could be associated with Zika, students had not been exposed to the context in which it spread, thus making it a novel prompt.

Table 3.4 Characteristics of the Modeling Tasks

Modeling Task	**Context Familiarity**		**Conceptual Familiarity**		**Modeled as component of coursework?**
	Everyday occurrence?	**Covered in class?**	**Concept**	**Covered in class?**	
1- Smell *(Interview 1)*	Yes	No	Diffusion Olfactory	No	No
2- Cystic Fibrosis *(Interview 1&2)*	No	Yes	Gene Expression	Yes	Yes
3- Wolf Vertebrae *(Interview 1)*	No	Yes (context: ecology)	Gene Expression	Yes	No
4 – Zika Virus *(Interview 2)*	No / Yes	No	Gene Expression Population Growth	Yes	No

Coding Protocols

In this study, I collected qualitative data, and then quantified the data, making them measurable by “operationalizing” the qualitative constructs using coding protocols (Kim et al., 2013). The qualitative constructs I needed to define were “metacognition,” “generative thinking,” and “causal reasoning.”

We chose to not use self-report learning style inventories because researchers have raised concerns about the validity of self-report surveys, especially the need for participants to “think about their thinking of their own thinking” to complete the surveys (Dinsmore & Alexander, 2012, p. 502). In addition, Case & Marshall (2009) reported discrepancies between approaches to learning students identified on self-report surveys and the approaches to learning the researchers observed. Furthermore, researchers who used in-depth interviews and other qualitative methods tended to find that approaches to learning are context specific, while those who used quantitative survey questionnaires tended to view approaches to learning as consistent traits of the learner (Chin & Brown, 2000a). This is a concern because of the contextual nature of learning (Brown, Collins & Duguid, 1989).

The coding process began by reading through the transcripts three times looking for words and phrases indicating the presence of metacognition, then generative thinking, and then causal reasoning. I developed the coding protocols *a priori* and modified them as I engaged in the coding process. Next, I watched the videos, looking and listening for the context in which the highlighted phrases in the transcripts occurred. I watched each video three times, one viewing for each component (generative thinking, metacognition or causal reasoning) looking for non-verbal and verbal indications that did not emerge from the transcripts. All videos were dual coded.

Generative thinking. Using the research (e.g., Biggs and Collis, 1982; Cavallo, 1996; Chan, Tsui, Chan & Hong, 2002; Chin & Brown, 2000; Duncan, 2007; Lee, Fradd, & Sutman, 1995; Newton, 2000; and Wong (1993), I identified two components of generative thinking – student use of prior knowledge, and student use of analogical transfer. Drawing upon prior experiences is also a form of generative thinking (Chin & Brown, 2000; Chin & Brown, 2000a; Lee, Fradd, & Sutman, 1995), but I chose not to include it in the data analysis because there were only two instances of it out of the 100 events that I captured.

Prior knowledge. When observing prior knowledge, I listened for specific science principles students would have learned from the biology course in which they were currently enrolled, or from other secondary or post-secondary courses. Some students made this a simple task by stating where they learned the concept (such as diffusion in a chemistry course or the olfactory system in a neuroscience or psychology course). When students did not state where they learned a concept, I listened for sophisticated terms or concepts that would come from secondary or post-secondary science courses. Examples from their biology courses included channel proteins, frame-shift mutations, alleles, and the processes that comprise central dogma (i.e., genes coding for proteins through transcription and translation).

Analogical transfer. Analogical transfer is a problem-solving technique in which students notice analogous relationships between the current problem and previous problems and construct a parallel solution (Goldstein, 2008). This was fairly easy to detect because the two models (Wolf and CFTR1 for example) looked nearly identical. In that case the student would comment about the similarities or having just constructed a similar model. I include a summary of all the coding protocols in Table 3.5.

Scoring. Students could get a maximum score of 2 pts. for generative thinking: 1 point for accessing prior knowledge and 1 point for using analogical transfer. Therefore, a student could include both elements of generative thinking (score of 2), only one of the elements of generative thinking (score of 1), or none of the elements of generative thinking (score of 0). An alternative scoring system would give credit for each instance of a student using prior knowledge and

Table 3.5 Definitions and Examples of the Coding Protocols

Code	Definition	Indicators	Examples
Generative Thinking			
Prior Knowledge	Using information learned in secondary or post-secondary science classes	Students refer to the class where they learned the information; Student used concepts more advanced science concepts such as the process of diffusion, and how the olfactory system functions.	*I think that's the right word* [pause] *disseminate. And then they'll enter through, we'll say nose here. And after they're in the nose they'll go I think it's to the olfactory bulb. Which was high school psychology I learned about that. And they do that through neurotransmitters firing.*
Analogical Transfer	Students' notice similarities between the current problem and previous problems	Students make reference to a similar problem; Students construct similar models for between-transfer prompts.	*So that's our genotype. We gotta get to the phenotype. So now it's going to be, it's* [CFTR1] *kind of similar to the wolves…*
Metacognition			
Checking Prompt	Students compare their model and modeling progress to the prompt	Students rereading the prompt during modeling and then comparing what they read to their model	*This is the CFTR gene and the relationship between a gene and an allele would be contains, because it could have either one…* [Pause] *'How genetic variation ultimately results in expression of a trait'* [read prompt]. *The differences between the alleles is the nucleotide sequences…*
Checking for Errors	Students notice something is missing or wrong in their models	Students explicitly stated that they found a mistake or were looking for errors.	*So, I have chromosomes contain DNA which is composed of nucleotides.* [pause] *Okay, that it is correct.*
Causal Reasoning			
Explanation Absent	Explanations of the causal mechanisms are not grounded in science concepts; the simplicity of the explanation is not representative of a college sophomore or junior majoring in the sciences.	Mechanisms students refer to could have come from an elementary student. Science processes and concepts are not mentioned.	*...there is a doughnut shop. And it's windy, so they can smell it, so there's wind. And, that's all I got, I really don't know.*
Explanation Incomplete	Students explanations are generic and could be applied to many different models.	The use of covering laws; students would mention mechanisms but could not explain them, nor could they expand upon information in their models	*The smell gets picked up by the person's nose. Whatever happens with like the receptors, it's like received and, then the brain would tell you what they're smelling.*
Explanation Complete	Students provide in-depth explanations of their models using applicable science concepts	Elaborates upon information in the model rather than narrating what is being modeled. Provides details about the science mechanisms in the model.	*So, these chemicals will bind to the chemoreceptors on the nasal mucosa and it will send signals straight to the olfactory bulb. And you have sense of smell.*

analogical transfer. However, I found students reiterating phrases while thinking aloud; if I had counted each instance I would have over-represented generative thinking skills in the more verbose students. Meijer and colleagues (2006) reported a similar concern.

Metacognition. Evidence of metacognition collected using think-aloud protocols can be analyzed quantitatively or qualitatively. Quantitatively researchers can count the number of times a participant demonstrates metacognition (Meijer, et al., 2006). For example, count phrases such as, “I’ve figured out what I wanted to say”; “I don’t get it”; and “I didn’t do that right, I’m getting confused” (Chin & Brown, 2000). As with generative thinking, counting metacognitive occurrences as the frequency of statements may bias results, giving higher ratings to participants who are verbose (Meijer et al., 2006). Qualitatively each metacognitive activity is given a rating based on the depth of processing. For example, detecting an error (i.e. “monitoring”) would receive a lower rating than a detected error followed by a correction (i.e. “regulation”), or a detected error accompanied by an analysis of the cause of the error (Meijer et al., 2006). I chose to use a dichotomous coding scheme similar to that used for generative thinking, which I will describe shortly.

Researchers can monitor both verbal and non-verbal communication for metacognition during think-aloud activities. Verbally, students may (a) use self-evaluative statements of their understanding, (b) engage in self-questioning when at an impasse, (c) detect and correct errors, (d) mention the limitations in their ideas, and (e) note counter-intuitive ideas and contradictory information (Chin & Brown, 2000a). Students may also verbalize the steps they are taking to solve the problem (Veenman, van Hout-Wolters & Afflerbach, 2006). Non-verbal indicators of metacognitive monitoring may be inferred from cognitive activities. For instance, rereading a portion of text, or pausing when looking over completed work. Although rereading is a cognitive activity, the decision to do so may indicate metacognitive behavior (Meijer et al., 2006).

While I noticed many non-verbal indicators of cognition, I chose to count only those indicators that I could observe without inference. As a result, our metacognition coding protocol included the presence or absence of two verbal behaviors that may or may not have accompanied non-verbal behaviors: (1) Checking for errors (i.e. "monitoring"), and (2) checking the prompt (i.e. "regulation") while engaged in the modeling task. Similar to the coding protocol for generative thinking, students could earn a maximum score of 2 points: 1 point for checking the prompt while modeling, and 1 point for error checking. Therefore, a student could include both elements of metacognition (score of 2), only one of the elements of metacognition (score of 1), or none of the elements of metacognition (score of 0).

We included checking the prompt as an observable metacognitive activity because students would check the prompt in order to assess their modeling. For example, to make sure that their model was addressing the prompt. In most cases students would articulate why they rechecked the prompt, or the decision they reached after checking their model against the prompt. For example, after looking over the prompt a student might say, "I only answered the first part of the question." Similarly, when checking for errors students would typically stare at the board and then announce what they found such as, "I think that I should have my model branch to alleles from the nucleotide sequences rather than the chromosomes."

Causal reasoning. In our first set of coding protocols I attempted to capture causal reasoning using the Russ et al. (2008) causal/mechanical framework. I found, however, that interpreting students' words and models in terms of entities and activities was resulting in too much interpretation. For example, a student would mention "mutation" as a mechanism but then have it placed in a box in the model, which made it an entity.

In the next iteration of our protocols I coded structures and relationships and then subdivided the codes into micro or macro. Applying values to these codes however implied hierarchy, which I could not defend. Furthermore, the codes did not capture what I was seeing; there were clear differences in the way students engaged in causal reasoning, which I identified as demonstrating command of the content in such a way that students elaborated on the entities and mechanisms in their models. Conversely, there were students who would only state what their model represented. Furthermore, the science mechanisms that students cited were not sophisticated, reflecting science content learned in the primary grades. For example, the wolf population had malformed vertebrae because they fell through the ice.

Due to the limitations just described, I chose to use holistic codes based on the scientific sophistication of student model-based explanations (Dauer & Long, 2015), as shown in Table 3.6. Like Duncan (2007), I looked for explanations that were plausible and productive; and I used a constant comparison approach to construct reasoning codes that captured students' knowledge underlying plausible and domain-appropriate explanations. The resulting coding scheme resembled the one used by Dauer and Long (2015), which captured students who demonstrated complete understanding of the mechanisms in their models, could describe some aspects of their models, and those students who were unable to provide a coherent explanation of the mechanisms in their models.

Table 3.6 Coding Protocol for Causal Reasoning

Code	Score	Features of Explanation
Explanation Absent	0	Not grounded in science concepts, or science concepts addressed in secondary or post-secondary learning environments. For example, "we can smell because the wind blows odors into our noses."
Explanation Incomplete	1	The model and /or explanation is generic and could be used to explain multiple phenomena. For example, "a mutation changed the nucleotide sequence and that resulted in a different trait."
Explanation Complete	2	The student elaborates upon the mechanisms operating within the model, recognizing that each mechanism is a model or nominalization (Snow, DATE) with its own structures and relationships. For example, "this malformed protein is a channel protein that restricts the flow of chloride ions in and out of the cell, so not enough water flows out of the cell, which creates a sticky mucous within the bronchi."

Approach-to-modeling score. We calculated an approach-to-modeling score as the sum of the scores for generative thinking, metacognition, and causal reasoning. Each was worth two points, allowing for a maximum approach-to-learning score of 6 points for each modeling prompt. Mathematically approach-to-modeling scores appear as:

Generative Thinking (2 pts.) + Metacognition (2 pts.) + Causal Reasoning (2 pts.) = Approach-to-Modeling (6 pts.)

To summarize, the approach-to-modeling scores are based on observable behaviors, such as drawing and vocalizing (e.g., "I think I did this wrong"). Smell was designed to serve as the "novel" prompt; by definition, generative thinking should be assessed when students are faced with an unfamiliar task or "novel prompt, such as Smell. To maintain scoring consistency however, I assessed generative thinking in all of the tasks.

We captured the three approach-to-modeling components using StudioCode video coding software. Rather than counting each time a component occurred however, I counted it only once using a dichotomous system of being present or absence. The exception was the causal reasoning codes. All videos were transcribed verbatim, and I used the transcripts, videos and models to code the data.

Model correctness. A scientifically correct model is a mechanistic, plausible explanation containing appropriate science principles (Duncan, 2007). I captured these elements with causal reasoning scores, which included student verbal explanations and student models. I compare model correctness against the approach-to-modeling score as a way of identifying models that may have been constructed from memory from those constructed using metacognition and generative thinking. In making the comparison between model correctness and approach-to-modeling I removed the causal reasoning score from the approach-to-modeling score since it also captures model correctness.

I attempted to define model correctness using a correctness rubric for the CFTR1 models, but found that the assessment was incomplete because the illustrations (drawn models) did not include the verbal descriptions students provided. In other words, model correctness scores needed to include both verbal and illustrated explanations because some students left elements out of their models – choosing to explain these elements instead.

When using the CFTR1 correctness rubric to assess the drawn CFTR1 models, there was a significant correlation with the Causal Reasoning scores ($r = .470$, $n = 20$, $p = .037$), but neither Generative Thinking nor Metacognition scores. Furthermore, I found that that Causal Reasoning scores were significantly correlated with Generative Thinking scores ($r = .629$, $n = 20$, $p < .001$) and Metacognition scores ($r = .640$, $n = 20$, $p < .001$). Because model correctness scores obtained from the correctness rubric were correlated with Causal Reasoning scores, and Causal Reasoning scores were significantly correlated with Generative Thinking and Metacognition scores, I used the Causal Reasoning scores as indicators of model correctness.

Data Analysis

We used our coding protocols to convert qualitative data into quantitative data, which I analyzed statistically and visually to identify patterns. I then compared the patterns that I observed to the transcripts, models and videos that I collected in order to triangulate our data.

Descriptive statistics. We began data analysis by calculating standard deviations, standard errors, means, skewness and kurtosis. In addition, I looked at the differential statistics using multiple representations including histograms, stem-and-leaf plots, and box-and-whisker plots. I ran inferential statistics once the descriptive statistics indicated that conditions of normality had been met. Furthermore, I used descriptive statistics to bin (group) the continuous data (approach-to-modeling scores) in order to compute chi-square values. You will find a list of the descriptive statistics that I used in Table 3.7.

Table 3.7 Descriptive Statistics Used in Data Analysis

Data	Mean	Median	Mode	Std. Dev.	Percentile	Histogram	Stem-and-leaf	Box-and-whisker	Skewness	Kurtosis
Total A2M[1] Score – all prompts	X	X	X	X	X	X	X	X	X	X
Smell: A2M[1] GenThin[2] Metacog[3] Reason[4] Tritile	X	X	X	X	X	X	X	X	X	X

Table 3.7 (cont'd)

Wolf: A2M[1] GenThin[2] Metacog[3] Reason[4] Tritile	X	X	X	X	X	X	X	X	X	X
CFTR1: A2M[1] GenThin[2] Metacog[3] Reason[4] Tritile	X	X	X	X	X	X	X	X	X	X
Zika: A2M[1] GenThin[2] Metacog[3] Reason[4] Tritile	X	X	X	X	X	X	X	X	X	X
A2M by prompt	X	X	X	X	X	X	X	X	X	X
Total GenThin by prompt	X	X	X	X	X	X	X	X	X	X
Total Metacog by prompt	X	X	X	X	X	X	X	X	X	X
Total Reason by prompt	X	X	X	X	X	X	X	X	X	X
Model Correct-ness by prompt	X	X	X	X	X	X			X	X
Model Correct-ness (binned) by prompt	X	X	X	X	X	X			X	X

[1]A2M – Approach-to-Modeling scores
[2]GenThin – Generative thinking scores
[3]Metacog – Metacognition scores
[4]Reason – Causal Reasoning scores

Inferential statistics. I used the following inferential statistics:

Chi-square test of independence. I used a chi-square test to examine the relationships between two categorical variables. The variables that I investigated were: Approach-to-modeling scores (binned), metacognition scores (binned), generative thinking scores (binned), causal reasoning scores, prompts and tritiles. I list the comparisons in Table 3.8.

In addition to looking at the statistical significance of the relationships (i.e., p values), I also looked at the standardized residuals and flagged cells that had a residual value greater than 1.96 or less than a – 1.96 because standardized residuals are converted to z-scores (standard deviation units) in which 95% of the z-scores should lie between – 1.96 and + 1.96 (Field, 2013). I also constructed bar charts to graphically observe the relationships between the different categorical variables.

One-way ANOVA with post-hoc analysis. I used one-way Analysis of Variance (ANOVA) to examine relationships between a continuous variable and categorical variables. The variables that I investigated were: Approach-to-modeling scores and prompts. First, I used Levene's test to reject the assumption that the variances were equal. Once this assumption was nullified I ran the one-way ANOVA. If there was a significant relationship between the two variables, I ran a post-hoc analysis to determine which relationships were significant.

***Spearman's rank correlation*.** I used a Spearman's rank correlation to observe a relationship between ordinal data (those scored on a 0 – 2 scale). Spearman's rank correlation does not require a linear association like Pearson's r, and is often used to evaluate relationships involving ordinal variables (Field, 2013). I used this test to examine relationships between: Generative thinking, metacognition, causal reasoning, approach-to-modeling score for each student by prompt and cumulative approach-to-modeling score by students for all prompts.

Hermeneutic content analysis. I used a modified hermeneutic content analysis (HCA) procedure to group and analyze our qualitative and quantitative data. HCA is a mixed methods framework that is aligned with a constructivist framework (Tasjakkori & Teddlie, 2010). Using HCA, I first coded our data (transforming the qualitative data into quantitative data) and then analyzed the quantitative data to identify patterns. I then interpreted the patterns by associating them with what I saw in the interviews (going from quantitative back to qualitative data) (Tasjakkori & Teddlie, 2010).

Table 3.8 Inferential Statistics Used in Data Analysis

Chi-Square	
Data Comparisons	Additional Tests
Approach-to-modeling scores (binned) x metacognition scores (binned) x reasoning scores x generative thinking scores (binned) x tritile Combined model correctness scores for all prompts (binned) x combined modeling approach scores for all prompts – binned (subtracting causal reasoning scores from A2M scores). Model correctness score (binned) for each prompt x modeling approach score – binned for each prompt.	Pearson's R Spearman's rank correlation Fisher's exact test (for Chi-Square tests with smaller expected cell counts)
One-Way ANOVA w/ Post-Hoc Analysis	
Data Comparisons	*Additional Tests*
Approach-to-modeling scores for each prompt (multiple comparisons)	Homogeneity of variance Tukey HSD Kolmogorov-Smirnov and Shapiro-Wilk tests of normality
Total approach-to-modeling score and tritile rank	Homogeneity of variance Bonferroni Tukey HSD
Correlations	
Data Comparisons	Additional Tests
CFTR1 approach-to-modeling score x CFTR1 model score from an expert reviewer.	Pearson Spearman's rank
Combined scores (all prompts and all students) for metacognition x generative thinking x causal reasoning	Spearman's rank

Triangulating data. While coding, I typed comments as to why I applied a certain code. Three months later I went back through the coded video clips to triangulate our data – comparing the qualitative data with the quantitative results (Tasjakkori & Teddlie, 2010). To triangulate our data, I compared the codes to the video clips and to the transcripts to ensure I captured what I saw.

CHAPTER 4: FINDINGS

In this chapter I will answer the research questions as claims supported by evidence. I first will address the questions: (1) To what extent do the different modeling tasks elicit the different components of deep approaches to modeling and (2) What elements of a modeling task may account for differences in student use of deep approaches to modeling? I then will address the research question regarding the relationship between deep approaches to modeling and model correctness.

Modeling Tasks Influence Student Approaches to Modeling

In response to the research question: Do the ways that students approach modeling change as the modeling tasks change, my findings demonstrate that different modeling tasks elicit different approaches to modeling; and in this section I will present supporting evidence for this claim. Specifically, I will show that individual students' approach-to-modeling scores changed as the modeling tasks changed, and that modeling tasks had an effect on the approach-to-modeling scores for all students. I will begin by showing changes in individual students' approach-to-modeling scores as the modeling prompts changed.

Same student, different scores. Table 4.1 is a heat map showing patterns in students' approach-to-modeling scores, with different colors representing different approach-to-modeling scores. Shades of red represent lower scores, with the deepest red representing the lowest scores, and shades of green represent higher scores, with the deepest tone of green representing the highest scores. The rows are ordered from highest to lowest average scores for each prompt. Looking across a row illustrates within-student variation as the modeling tasks changed. For example, Rachel's scores (row 1) ranged from a low of 3 for the Zika prompt to a high of 4 for all other prompts giving a total variation of 1 point. In contrast, Omid and Merle had scores ranging from 0 to 5 resulting in total variation values of 5. No students had 0 variation in their approach-to-modeling scores, which suggests that students' approaches to modeling change as modeling tasks change.

Table 4.1 Heat Map Showing Range of Approach-to-Modeling Scores by Prompt. Students (in rows) are ordered by descending cumulative approach-to-modeling score. Prompts (in columns) are arranged in the order that they occurred within the interviews. Tritile reflects a measure of students' academic achievement with Tritile 3 being students with highest GPA

		Interview 1			Interview 2		
Pseudonym	Tritile	Smell	Wolf	CFTR1	Zika	CFTR2	**Total**
Billy	3	4	4	5	2	4	19
Rachel	1	4	4	4	3	4	19
Sydney	2	3	4	4	2	4	17
Taylor	3	2	4	5	2	3	16
Omid	3	3	4	5	0	2	14
Julia	2	3	4	3	1	3	14
Rose	3	1	3	5	3	2	14
Sandra	3	3	2	3	1	4	13
Brian	2	3	3	4	0	3	13
Paul	3	3	3	3	0	4	13
Samantha	1	2	0	2	3	3	10
Merle	2	0	3	2	0	5	10
Leia	3	2	1	4	0	2	9
Rick	3	2	2	2	0	2	8
Anna	1	1	1	3	1	2	8
Simba	1	0	1	4	0	1	6
Elizabeth	3	0	0	2	0	3	5
Eric	1	0	0	2	1	1	4
Sheri	1	0	0	3	0	1	4
Sarah	2	0	1	2	0	0	3

Looking down the columns in Table 4.1 reveals between-student variations, and two distinct patterns emerge. First, students had the highest approach-to-modeling scores for the CFTR1 modeling task, as seen by the prevalence of green hues. Second, students had the lowest approach-to-modeling scores for the Zika modeling task, as seen by the prevalence of red hues.

Factors That Influence Approach-To-Modeling Scores

I will now present data that reveal some of the factors that might address the second part of the previous research question: What elements of a modeling task may account for differences in student use of deep approaches to modeling? I will begin with factors related to student interactions with the modeling prompts and conclude with factors related to the wording of the prompts.

Student factors. Students' approach-to-modeling scores are not a measure of student academic achievement, although content knowledge does influence the way students approach modeling. As a reminder, tritiles are used to represent differences in student academic achievement, and tritile 3 is the highest. Students with the highest and lowest approach-to-modeling scores for all five modeling prompts combined are listed in Table 4.1 You will notice that the top and bottom scores shown in Table 4.1 include all three tritiles, which indicates that approach-to-modeling scores capture more than academic performance.

The distribution of students' approach-to-modeling scores by tritile for the five prompts are displayed in Figure 4.1. The error bars in Figure 4.1 indicate there is no statistical difference in approach-to-modeling scores between tritile 2 (mean = 2.28; standard deviation = 1.59) and tritile 3 (mean = 2.47; standard deviation = 1.50) students. Only for the Wolf prompt does it appear that there is a difference in approach-to-modeling scores between students in tritile 1 (mean = 1.00; standard deviation = 1.55) and tritile 2 (mean = 3.00; standard deviation = 1.225) and between students in tritile 1 (mean = 1.00; standard deviation = 1.55) and tritile 3 (mean = 2.56; standard deviation = 1.42); the differences however are not significant. Therefore, the differences in approach-to-modeling scores appear to be more than differences in academic achievement.

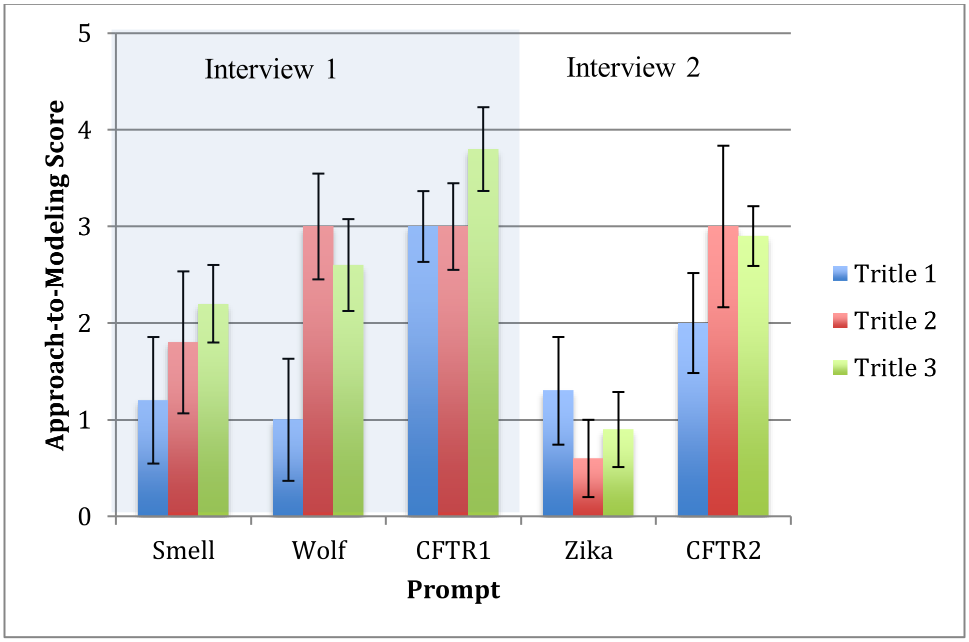


Figure 4.1 Approach-to-modeling scores by tritile and prompt

While not a measure of academic performance, approach-to-modeling scores do capture aspects of students' content knowledge. Recall that approach-to-modeling scores are the sum of the scores for generative thinking, causal reasoning and metacognition. These components of approach-to-modeling draw upon content knowledge and may be associated with tritile rankings. In the remainder of this section I will explore the association between tritile rankings, generative thinking and causal reasoning.

Generative thinking. A chi-square test of independence between generative thinking and student academic performance (represented by tritile ranking) was significant ($\chi^2 = 10.26$, df = 4, p = .036). This shows that generative thinking and student academic performance are associated, and students at higher tritiles tend to have higher generative thinking scores. Similarly, a Spearman's rank correlation indicates a moderate positive correlation between generative thinking scores and tritile ranking (ρ = .290, n = 100, p = .003).

Causal reasoning. A chi-square test of independence suggests that causal reasoning scores are associated with tritile rankings: (χ^2 = 10.83, df = 4, p = .029), with students at higher tritiles having higher causal reasoning scores. In addition, a Spearman's rank correlation indicates a moderate positive correlation between causal reasoning scores and tritile ranking (ρ = .289, n = 100, p = .004).

The relationship between tritile (academic performance) and generative thinking and causal reasoning scores may account for the prevalence of students from the higher tritiles in the top half of approach-to-modeling scores, as seen in Table 4.2. However, as shown in Table 4.3, a number of students from the highest tritile group are also found in the bottom 50^{th} percentile of approach-to-modeling scores. These findings further suggest that the approach-to-modeling framework is not merely capturing student academic performance.

The students in the second tritile were consistent in their approach-to-modeling scores (Figure 4.1) and generative thinking scores (Figure 4.2) across the CFTR1, CFTR2 and Wolf prompts. This consistency is noteworthy because the approach-to-modeling scores (Figure 4.1) and generative thinking scores (Figure 4.2) for the students in tritile 3 dropped slightly between CFTR1 and CFTR2 while the scores for students in tritile 2 remained the same.

Table 4.2 Top (50th percentile) Approaches-to-modeling: 5 Prompt Total

Student	Total Score	Tritile
Billy	19	3
Rachel	19	1
Sydney	17	2
Taylor	16	3
Omid	14	3
Julia	14	2
Rose	14	3
Sandra	13	3
Paul	13	3
Brian	13	2

Table 4.3 Bottom (50th percentile) Approaches-to-Modeling: 5 Prompt Total

Student	Total Score	Tritile
Merle	10	2
Samantha	10	1
Leia	9	3
Anna	8	1
Rick	8	3
Simba	6	1
Elizabeth	5	3
Sheri	4	1
Eric	4	1
Sarah	3	2

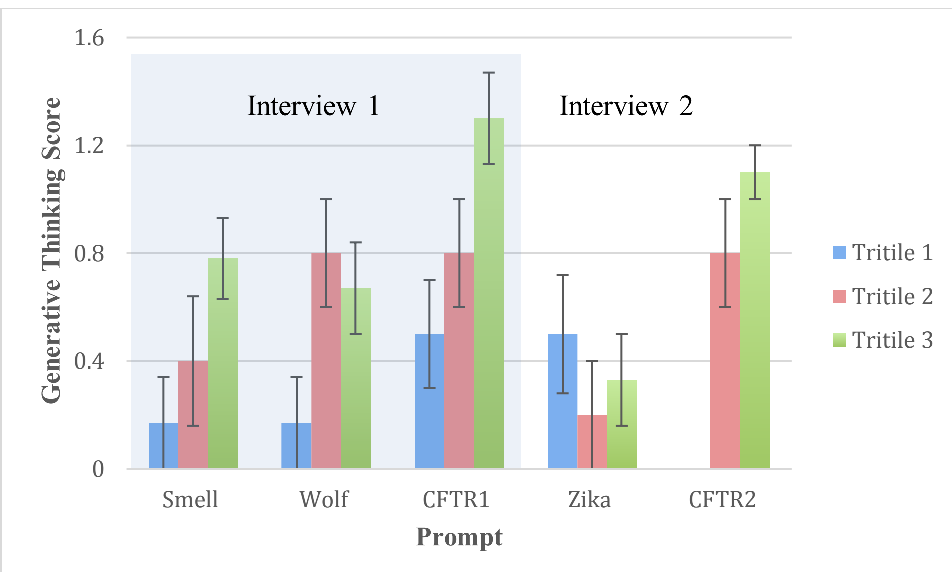


Figure 4.2 Generative thinking scores by prompt and tritile ranking

Modeling tasks. The approach-to-modeling scores in Figure 4.3 show that students had the highest approach-to-modeling scores for the CFTR1 modeling prompt, and the lowest approach-to-modeling scores for the Zika prompt. The scores are normally distributed with the assumption of normality accepted ($p < .001$) for both the Kolmogorov – Smirnov and Shapiro – Wilk tests of normality. Furthermore, there are statistically significant differences between the approach-to-modeling scores based on the prompt [F (4, 95) = 9.16, p < .001]. A Tukey post hoc test revealed that the mean approach-to-modeling scores for CFTR1 were significantly higher than those for the Smell prompt (1.80 ± 1.44 points, p = .003) and Zika prompt (.95 ± 1.15 points, p < .001). In addition, a Tukey post hoc test revealed that both the Wolf (2.20 ± 1.58 points, p = .031) and CFTR2 (2.65 ± 1.31 points, p = .031) were significantly higher than those for the Zika prompt.

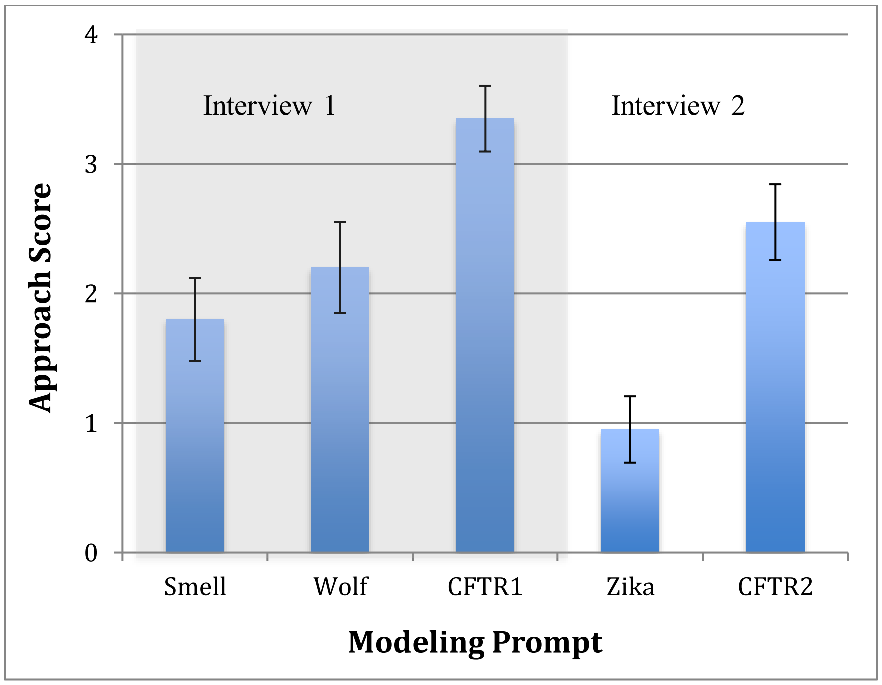


Figure 4.3 Mean approach-to-modeling scores by prompt, presented in the order in which they occurred during the interviews

Attributes of the modeling tasks – contextual familiarity. When learning, students compare new information to existing schemata (NRC, 2005). Similarly, when engaged in problem-solving, students search for schemata that may relate to the problem.

We designed the Zika prompt to be structurally similar to the CFTR and Wolf prompts, but students struggled with the Zika modeling task, as shown in Figure 4.3. Despite the overall low approach-to-modeling scores for Zika, six students had approach-to-modeling scores of ≥ 2 (out of 6) for the Zika prompt. These students were able to link information in the Zika prompt to a familiar context. In Table 4.4 you will see that these students linked the Zika modeling task to such familiar contexts as mosquito-borne illnesses and knowledge and experiences with viruses. For example, Samantha constructed an explanation for why the virus spread by thinking back to the different times she had a cold. Likewise, Rose, Sydney and Taylor drew upon their knowledge of viruses to explain how and why the disease spread. Finally, Billy relied upon what he had heard in the news and Rachel drew upon her experiences traveling to regions where mosquito-borne illnesses are common.

In a similar fashion, when presented with the Wolf modeling task seven students did not recognize the presence of the gene-to-protein mechanism they had modeled in class. Instead, they focused on the context which activated visual images held in their long-term memory. These images came from an introductory presentation about the wolves on Isle Royale they saw four to six weeks earlier during a population ecology unit. All of the familiar components that students recognized in the Wolf prompt are presented in Table 4.5 and will be discussed next.

Attributes of the modeling tasks – written cues. Table 4.5 shows the different cues that activated students' schemata when modeling the Wolf prompt. As mentioned in the previous paragraph, wolves on Isle Royale provided a context from which seven students constructed their models. The remaining students identified specific words that initiated their gene-to-protein schemata. Specifically, Table 4.5 shows that 4 students noticed the word "population" which lead to a reproduction schema. Six students focused on the word "malformed" which they associated with "mutation," which then activated their gene-to-protein schemata. Finally, 3 students recognized the phrase, "became present", which they learned in class indicated the presence of inheritance or reproduction.

Table 4.4 Contexts Students Used to Model the Zika Prompt

Student	Tritile	Zika Score	Contextual Anchor	Example
Samantha	1	3	Viral illnesses (common cold)	"I understand that, because if I got a small cold I probably wouldn't go to the doctor either…so the infected individual acts as a host because even though they don't feel it anymore, they are still spreading it."
Rose	3	3	Virology	"So, initiation would be…" (referring to the lytic cycle)
Rachel	1	3	International Travel & Mosquito-born illnesses	"It's like dengue fever, right?
Billy	3	2	News	"I know there's a lot of reports that the mosquito's now in Brazil because of the Olympics and the World Cup and a lot of people are freaking out about it..." there's talk that mosquitos would come up into the US, and they have as the summer went on...now there are mosquitos all over the world, at least in North and South America and I would imagine they're still in Africa and Asia. They can infect, the virus can infect even more populations of people."
Sydney	2	2	Virology	"I know back in class when we were learning about viruses…"
Taylor	3	2	Virology and Gene-to-Protein	I'm assuming it started with mutation in the RNA strand, which was in a mosquito or a monkey."

Capturing Students' Approaches-to-Modeling

In the methodology chapter I explained how we modified the Chin and Brown framework to capture approaches to modeling, and previously I provided evidence that the approach-to-modeling framework captured: (1) changes in students' approaches to modeling that were more than differences in their academic performance and (2) patterns in approach-to-modeling based on the prompt. In this section I will provide additional evidence that the approach-to-modeling framework was effective in detecting different modeling approaches by showing that the framework was able to differentiate between students' approaches to modeling tasks. As a reminder, when referring to a student's approach-to-modeling I am referring to their approach-to-modeling scores, which are the sum of their generative thinking, causal reasoning and metacognition scores. The higher the approach-to-modeling score, the more reflective the students were when drawing upon prior knowledge and reasoning through the modeling task.

Table 4.5 Contextual and Written Cues Students Used to Model the Wolf Prompt

Cues	# of Students	Example
Population: Reproduction	4	Prompt: Construct a model to explain how the malformed vertebrae became present among the **wolf population** on Isle Royale. *There must be some sort of reproduction within the wolf population. So, reproduction occurs and during that reproduction the baby wolf, at some point during translation of the DNA when it's replicating, a mutation occurred. So then this mutation of DNA from the normal sequence of nucleotides changed somewhere that indicated that malformed vertebrae should be produced, rather than a normal baby wolf without the vertebrae mutation.*
Malformed: Mutation	6	Prompt: Construct a model to explain how the **malformed vertebrae** became present among the wolf population on Isle Royale. *I'm just thinking about how the vertebrae would have become deformed and I'm thinking about the genetics of it. So, I'm thinking that there was a mutation in one of the genes that controls the shape of this vertebrae*
Became Present: Reproduction and Inheritance	3	Prompt: Construct a model to explain how the malformed vertebrae **became present among the wolf population** on Isle Royale. *Because it's such a small population, when they have the next generation there will be a greater likelihood that their offspring would have the same malformed vertebrae. So maybe now there would be twice as many that had it. In the next generation after that, it would be even more likely to be passed on. To the point that it would become common after many generations for them to have this type of vertebrae.*
Environment	7	*So, when the wolves are living together here, then they could interact with each other and fighting each other. [Also] some environmental effect, such as cold weather...* (Eric)

Differentiating between students. The approach-to-modeling framework characterizes differences in how students approach modeling by identifying the components that make up approach-to-modeling scores: generative thinking, metacognition and causal reasoning. Figure 4.4 shows the variation in students' approach-to-modeling scores, which ranged from 3 to 19. In addition to a wide range in approach-to-modeling scores, the component scores shown in Figure 4.4 are also quite variable, ranging from 1 to 9 for causal reasoning, 0 to 8 for metacognition, and from 0 to 7 for generative thinking. Furthermore, the highest and lowest scores for each of the three components all came from different students, and no two students had the same scores for each component. Table 4.6 is a summary of the results displayed in Figure 4.4, showing the range of scores for the three components of approach-to-modeling scores and the students who obtained those scores. Note that there is no consistent relationship between the three components. For example, Billy, Rachel, Simba and Sheri had nearly equal scores for all three components while Elizabeth, Leia, Merle, Paul, Rick and Sydney were strongest in reasoning and weakest in metacognition.

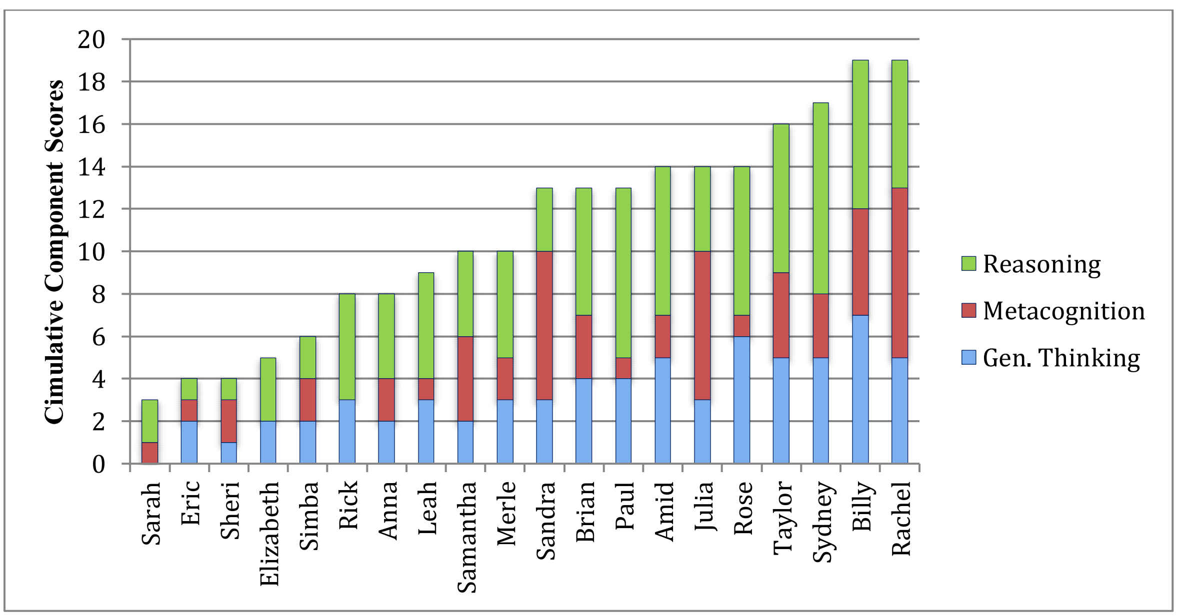


Figure 4.4 Range of approach-to-modeling scores by student and approach-to-modeling components

Table 4.6 Range of Students' Approach-to-Modeling Component Scores

	Causal Reasoning		Generative Thinking		Metacognition	
	Low	High	Low	High	Low	High
Score	1	9	0	7	0	8
Student	Eric Sheri	Sydney	Sarah	Billy	Elizabeth Rick	Rachel
Tritiles	1	2	2	3	3	1

Differentiation between modeling tasks. The approach-to-modeling framework captured differences between the modeling tasks and students' use of causal reasoning, generative thinking and metacognition. Figure 4.5 shows the extent to which different modeling tasks affected approach-to-modeling components (causal reasoning, generative thinking, metacognition). A clear pattern is evident in Figure 4.5 – the Zika prompt did not elicit deep modeling approaches. Figure 4.5 also shows that the CFTR prompts elicited the most generative thinking among students, and that all prompts except Zika encouraged most students to engage in causal reasoning.

Proximity of interview to exam date. Several students appeared to rely upon rote memorization when constructing models; this was most evident with the CFTR1 model where students had little difficulty recreating the model that they constructed on their second exam. As a reminder, the CFTR1 prompt during the first interview was the same prompt that students received on their second exam. Even though there was an interval of 1 – 24 days between the second exam and the students participating in the interviews, I found no significant difference in

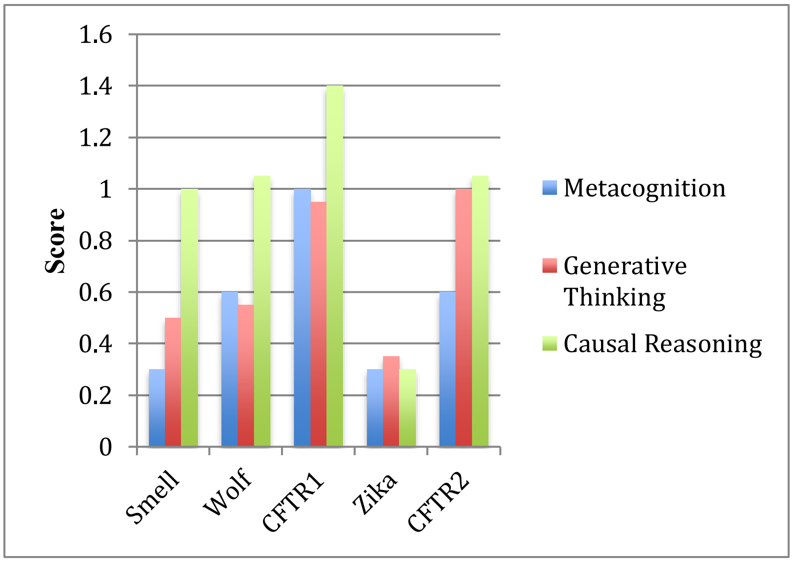


Figure 4.5 Mean attribute scores by prompt

the CFTR1 approach-to-modeling scores for students who participated in the interview 1 day after the test and up to 24 days after the test. As you will see in Figure 4.6, there is no predictive pattern in the relationship between CFTR1 approach-to-modeling scores and days from the second exam. Similarly, despite lag times between the test date and the interview date, CFTR exam scores and CFTR1 interview correctness scores are significantly correlated ($\rho = .626$, $n = 20$, $p = .003$). The relationship between the CFTR exam model scores (red bars) and CFTR1 model correctness scores (blue bars) are shown in Figure 4.7; six students had the same % correctness scores for both their interview models and exam models, and 5 students (Eric, Samantha, Billy, Rick, Anna) had similar scores.

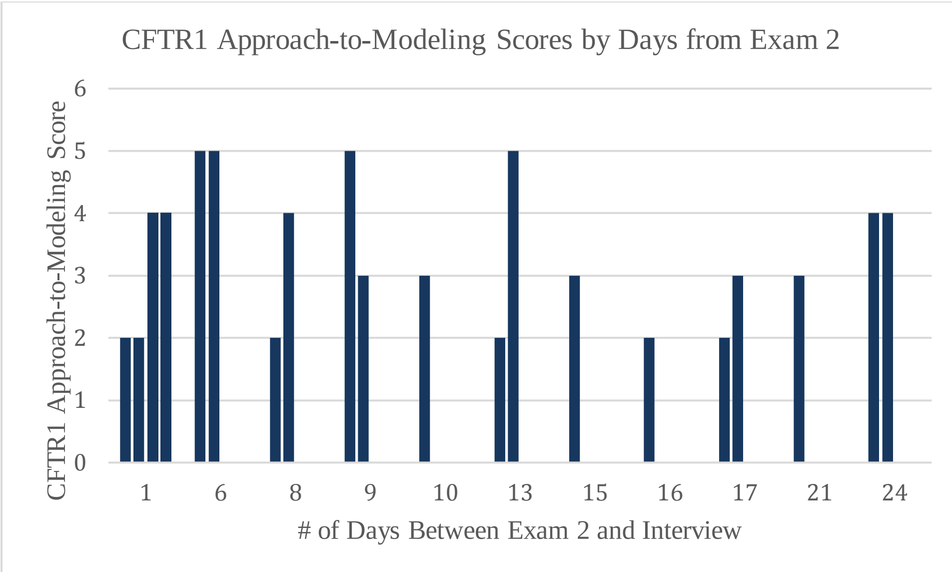


Figure 4.6 Differences in CFTR1 scores by the number of days between the second exam and the interview. CFTR was a question on the second exam

Because the CFTR model scores that students constructed for the exam were assessed using a detailed correctness rubric, I consider this strong correlation between CFTR model test scores and CFTR Causal Reasoning (a.k.a. model correctness scores) from the interviews as validating the use of causal reasoning scores as proxies for model correctness scores.

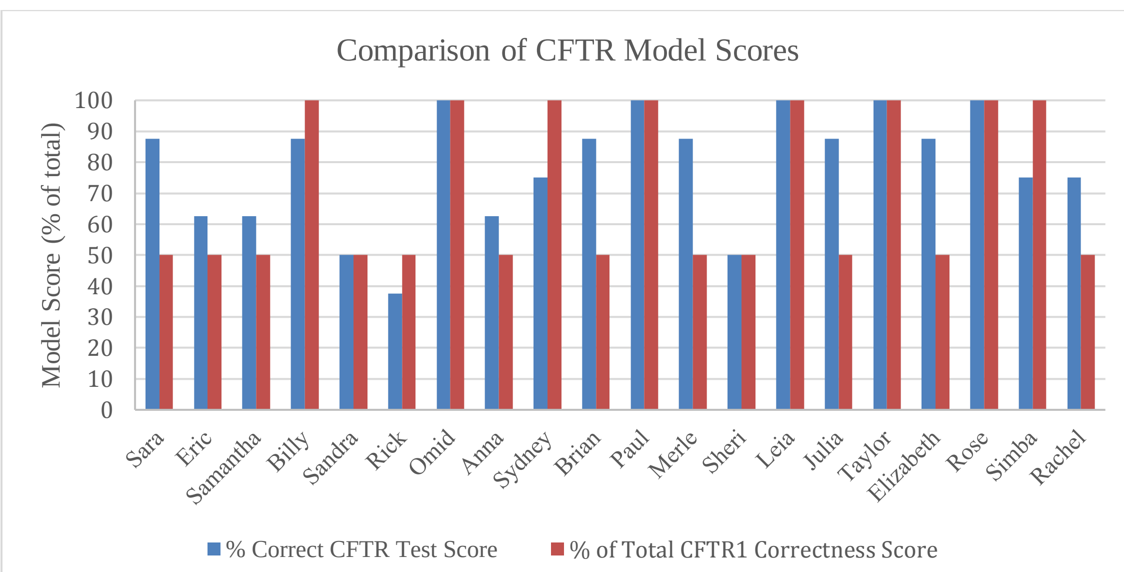


Figure 4.7 Comparison of CFTR scores between the second exam and the interview

The relationship between deep approaches to modeling and scientifically correct models. When comparing model correctness scores to modeling approach scores (Generative Thinking + Metacognition), the results vary by prompt and student, as shown in Tables 4.7a-f. Table 4.7a includes combined scores for all 5 models that the 20 students constructed. You will see a general trend of higher modeling approach scores being associated with higher model correctness scores, and students from the upper tritiles more likely to have higher modeling approach scores and model correctness scores. This overall pattern, however is not consistent across each prompt.

Table 4.7a Grouping Students' by Modeling Approach Scores (binned) and Model Correctness Scores (binned) for all 5 Prompts

Modeling Approach Scores[1] (binned)	**Correctness Scores** (binned)		
	Low (Score: 1-3)	Medium (Score: 4-6)	High (Score:7-9)
Low (Score: 1-4)	**5** Eric (1)[2] Sheri (1) Simba (1) Sarah (2) Elizabeth (3)	**3** Anna (1) Rick (3) Leia (3)	**0**
Medium (Score: 5-7)	**0**	**3** Samantha (1) Brian (2) Merle (2)	**3** Rose (3) Omid (3) Paul (3)
High (Score: 8-13)	**1** Sandra (3)	**2** Rachel (1) Julia (2)	**3** Sydney (2) Taylor (3) Billy (3)

[1] Modeling Approach Scores = Metacognition Score + Generative Thinking Score
[2] Student tritile ranking is shown in parentheses

When looking at Tables 4.7a-f, modeling approach scores comprise the rows, with the score indicating the presence of metacognition, generative thinking or both, as follows:

Modeling Approach Score	Interpretation
0	No evidence of metacognition or generative thinking.
1	One component of either metacognition or generative thinking observed.
2	Either one component of metacognition and generative thinking observed; or Two components of either metacognition or generative thinking observed.
3	One component of metacognition and generative thinking observed; An additional component of either metacognition or generative thinking observed.
4	Both components of metacognition and generative thinking observed.

Likewise, when students receive a correctness score of 0, it indicates that they did not refer to any of the science mechanisms that contribute to the phenomenon. For example, in Table 4.7b, seven students received correctness scores of 0, indicating that they did not refer to any of the science mechanisms involved in smelling, such as diffusion or the olfactory system.

In Table 4.7b (Smell), Table 4.7c (Wolf), and Table 4.7f (CFTR2), you will notice that high model correctness scores are more closely associated with medium (mid-level) modeling approach scores, which indicates that it is possible for some students to achieve high correctness scores without engaging in high levels of metacognition or generative thinking. For the CFTR1 (Table 4.7d) models, the pattern changes and increasing use of metacognition and generative thinking (higher modeling approach scores) are associated with higher model correctness scores, with several students approaching the maximum modeling approach score of 4, which would indicate equal use of metacognition and generative thinking. The pattern of mid-level modeling approach scores being associated with high model correctness scores also changed with the Zika prompt (Table 4.7e), where engaging in metacognitive behavior and generative thinking did not always produce correct models. In fact, only one student produced a scientifically correct model despite 10 students having mid-level modeling approach scores.

Table 4.7b Grouping Students' by Modeling Approach Scores and Model Correctness Scores for the Smell Prompt

Modeling Approach Scores[1]	**Correctness Scores** Low (Score: 0)	Medium (Score: 1)	High (Score: 2)
Low (Score: 0)	**6** Sheri (1)[2] Sarah (2) Simba (1) Merle (2) Eric (1) Elizabeth (3)	**1** Anna (1)	**1** Rick (3)
Medium (Score: 1)	**1** Rose (3)	**3** Samantha (1) Taylor (3) Leia (3)	**4** Sydney (2) Omid (3) Brian (2) Paul (3)
Medium High (Score: 2)	**0**	**2** Julia (2) Sandra (3)	**2** Rachel (1) Billy (3)

[1] Modeling Approach Scores = Metacognition Score + Generative Thinking Score
[2] Student tritile ranking is shown in parentheses

Table 4.7c Grouping Students' by Modeling Approach Scores and Model Correctness Scores for the Wolf Prompt

Modeling Approach Scores[1]	**Correctness Scores** Low (Score: 0)	Medium (Score: 1)	High (Score: 2)
Low (Score: 0)	**4** Eric (1)[2] Sheri (1) Samantha (1) Elizabeth (3)	**3** Anna (1) Sarah (2) Leia (3)	**0**
Medium (Score: 1)	**1** Simba (1)	**1** Rick (3)	**4** Merle (2) Rose (3) Brian (2) Paul (3)
Medium High (Score: 2)	**1** Sandra (3)	**0**	**3** Sydney (2) Billy (3) Omid (3)
High (Score: 3)	**0**	**3** Rachel (1) Julia (2) Taylor (3)	**0**

[1] Modeling Approach Scores = Metacognition Score + Generative Thinking Score
[2] Student tritile ranking is shown in parentheses

Table 4.7d Grouping Students' by Modeling Approach Scores and Model Correctness Scores for the CFTR1 Prompt

	Correctness Scores		
Modeling Approach Scores[1]	Low (Score: 0)	Medium (Score: 1)	High (Score: 2)
Low (Score: 0)	**0**	**0**	**0**
Medium (Score: 1)	**0**	**6** Samantha (1) Rick (3) Eric (1) Merle (2) Sarah (2) Elizabeth (3)	**1** Paul (3)
Medium High (Score: 2)	**0**	**4** Anna (1) Julia (2) Sheri (1) Sandra (3)	**3** Simba (1) Sydney (2) Leia (3)
High (Score: 3)	**0**	**2** Rachel (1) Brian (2)	**4** Rose (3) Billy (3) Taylor (3) Omid (3)

[1] Modeling Approach Scores = Metacognition Score + Generative Thinking Score
[2] Student tritile ranking is shown in parentheses

Table 4.7e Grouping Students' by Modeling Approach Scores and Model Correctness Scores for the Zika Prompt

	Correctness Scores		
Modeling Approach Scores[1]	Low (Score: 0)	Medium (Score: 1)	High (Score: 2)
Low (Score: 0)	**10** Sheri (1)[2] Elizabeth (3) Simba (1) Rick (3) Sarah (2) Omid (3) Merle (2) Paul (3) Brian (2) Leia (3)	**0**	**0**
Medium (Score: 1)	**4** Eric (1) Anna (1) Julia (2) Sandra (3)	**2** Sydney (2) Taylor (3)	**1** Rose (3)
Medium High (Score: 2)	**1** Billy (3)	**2** Rachel (1) Samantha (1)	**0**

[1] Modeling Approach Scores = Metacognition Score + Generative Thinking Score
[2] Student tritile ranking is shown in parentheses

Table 4.7f Grouping Students' by Modeling Approach Scores and Model Correctness Scores for the CFTR2 Prompt

	Correctness Scores		
Modeling Approach Scores[1]	Low (Score: 0)	Medium (Score: 1)	High (Score: 2)
Low (Score: 0)	**1** Sarah (2)[2]	**0**	**0**
Medium (Score: 1)	**3** Eric (1) Sheri (1) Simba (1)	**5** Anna (1) Leia (3) Rick (3) Rose (3) Omid (3)	**2** Taylor (3) Elizabeth (3)
Medium High (Score: 2)	**0**	**3** Samantha (1) Brian (2) Julia (2)	**2** Sydney (2) Paul (3)
High (Score: 3)	**0**	**3** Rachel (1) Sandra (3) Billy (3)	**1** Merle (2)

[1] Modeling Approach Scores = Metacognition Score + Generative Thinking Score
[2] Student tritile ranking is shown in parentheses

Components of modeling approaches. Overall, 14% of the time students with high modeling correctness scores did not demonstrate the use of any metacognitive skills. While not a statistically significant finding, it does raise questions about our ability to capture metacognition. The concerns about capturing metacognition are further seen in Figure 4.8 where the expected pattern of a direct, positive relationship between Metacognition and model correctness scores is not present; the expected pattern is clearly present however in Figure 4.9 between modeling approaches (excluding the reasoning component) and model correctness scores. The pattern also is seen, but to a lesser extent in Figure 4.10 between model correctness and Generative Thinking scores.

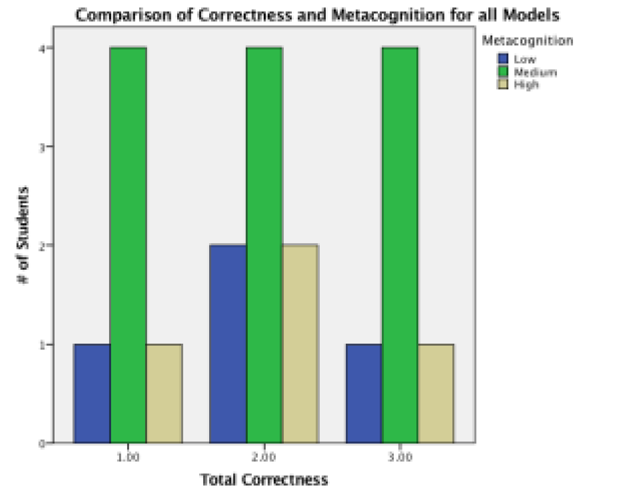


Figure 4.8 Comparison of model correctness scores and metacognition

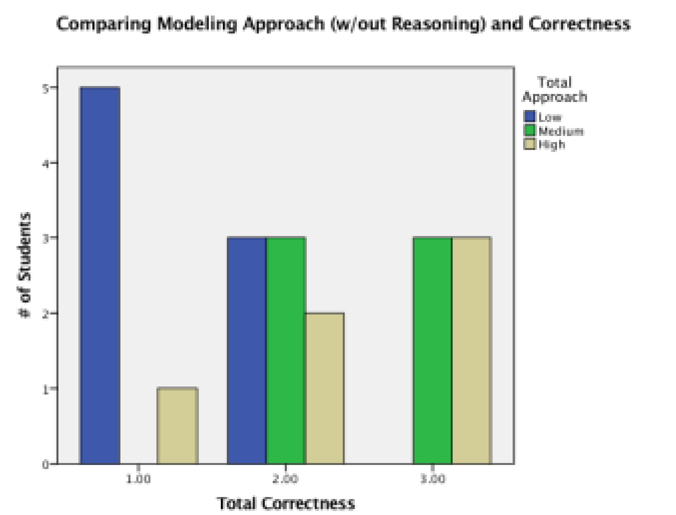


Figure 4.9 Comparison of model correctness and modeling approach scores

Indicators of weak cognitive structure. I cannot access students' cognitive structure directly, but students provide clues as to the interconnectedness of the items they hold in long-term memory. The approach-to-modeling scores provide a clue, which becomes stronger when paired with students' verbal and non-verbal communication. Simba and Sarah present two cases where this triangulation of data revealed weaker cognitive structures. Table 4.8 shows the CFTR1 and CFTR2 approach-to-modeling scores for Simba and Sarah had that dropped from

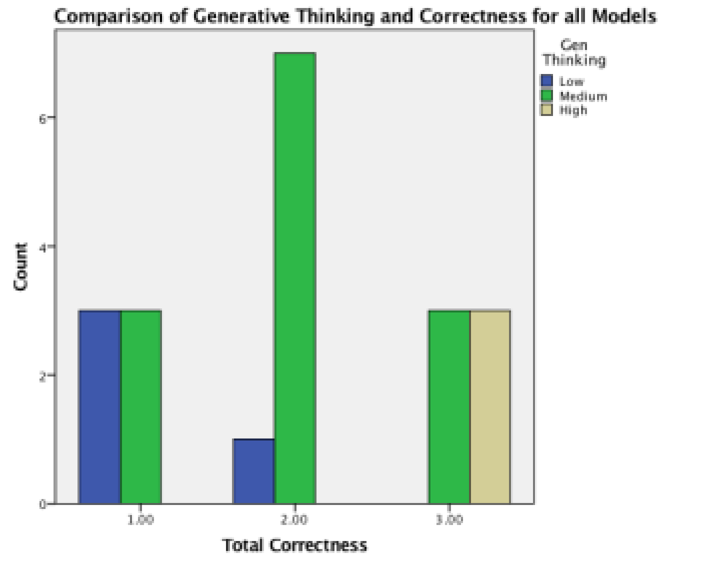


Figure 4.10 Comparison of model correctness and generative thinking scores

4 to 1, and 2 to 0, respectively. While the magnitude of difference in the approach-to-modeling scores is not striking, a more complete picture of their cognitive structure appears when looking at their comments and gestures while engaged in the modeling task. Both students stared at the Promethean Board for extended periods of time without talking. For example, after staring at the Promethean Board for nearly 1.5 minutes and only saying, “hmmm,” Simba commented, “Um, honestly I don’t remember [pause]. Honestly, I don’t remember like how it does originate, genetic variation, or how it developed.”

Table 4.8 Lowest Approach-to-Modeling Scores by Prompt

Student	CFTR1	Smell	Wolf	Zika	CFTR2	Total Score
Simba	4	0	1	0	1	6
Eric	2	0	0	1	1	4
Sarah	2	0	1	0	0	3
Elizabeth	2	0	0	0	3	5
Sheri	3	0	0	0	1	4

During the 4-minute modeling period, Simba mentioned genes, amino acids and proteins, knowing that they were somehow associated with the modeling task; she stated, “I can’t remember everything about genes [pause] and something with proteins and amino acids and the interlinkings of them, and the order makes up your genome and so that’s the genetic variation.” What the table and quotes cannot show are Simba’s facial expressions and discomfort in not being able to recall what she had done previously on the test, and not being able to draw upon what she learned in her bio class to generate an answer.

Sarah expressed similar frustration, both verbally and physically. She was very flustered during the 4.5-minute modeling activity stating several times that she had no idea and finally asking for a different question:

> *I'm really sorry. I like don't know how to do this one. I kind of think back to how these works. I don't know. Is there like another question I can answer? I'm really sorry, I don't know what else there is to do besides this to be honest.*

While Simba appeared to have memorized her CFTR1 model during the first round of interviews, Sarah openly shared that she memorized models and that is why she was having difficulties with the CFTR prompt in the second interview. When considering how to begin modeling the CFTR2 prompt, Sarah stated, “I mean, I like remember memorizing how this looked for an exam; and I remember, kind of remembering how I did it like a week later when I did this here. And so now I have no idea.” It appears that students who relied exclusively on memorization held concepts as isolated fragments, making it difficult to retrieve learned information from long-term memory.

The Smell prompt provided additional evidence that Simba and Sarah struggled to retrieve ideas from long-term memory and therefore could not engage in generative thinking or causal reasoning. Their models, shown in Figures 4.11 and 4.12, do not represent the type of causal explanations that were typical among their peers.

Figure 4.11 Sarah's Smell model: Wind blowing scent from a donut shop to a person

Sara shared the following while modeling how a person Smells things from a distance:

> *Well I'm going to draw the person first. And, oh this is really hard. So, okay so you'll be there like downtown somewhere. And there is a doughnut shop. And it's windy, so they can smell it, so there's wind. And, that's all I got, I really don't know.*

Like Sarah, Simba's explanation of how people Smell was quite generic:

> *I'm thinking how I don't know actually what smell is… I'm thinking how, I guess, in this sense I could say: goes through the air, the wind blows it. So…Let's say the wind is going to blow this to the person to smell. Not sure of the scientific way actually at all. Never thought of that. So, the wind is going to blow the smell towards the person* [pause while

drawing] *it's going to drift into his nose. And then the nose processes it somehow. That's all I got.*

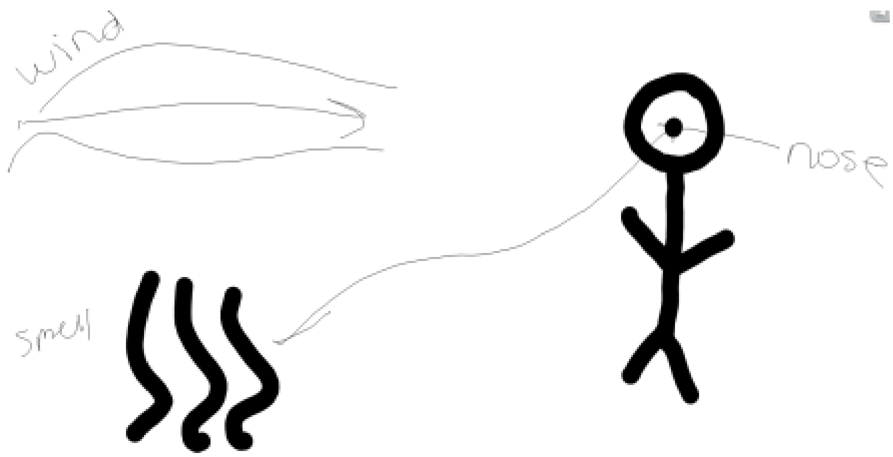


Figure 4.12 Simba's Smell model: Wind blowing scent from a donut shop to a person

The role of novel prompts. As discussed in Chapter 2 and again in Chapter 3, having students engage in a novel modeling task can reveal the depth of students' generative thinking, metacognition, and causal reasoning because it requires them to actively rely upon working memory to construct mental models from existing schema. Up to this point I have said very little about the Smell prompt, which was the novel prompt in the first interview. The way students approached the Smell prompt was a good indicator, both qualitatively and quantitatively, of how the students approached modeling overall. I computed a Spearman's rank correlation to determine the relationship between the approach-to-modeling score for Smell and the overall approach-to-modeling score for all five prompts. The correlation between overall approach scores and total approach scores for the Smell prompt ($\rho = .863$, $n = 20$, $p < .001$) is strong and positively correlated. This is also shown graphically in Figure 4.14 where there is a clear linear relationship between the combined approach-to-modeling scores and the approach-to-modeling scores for the Smell prompt.

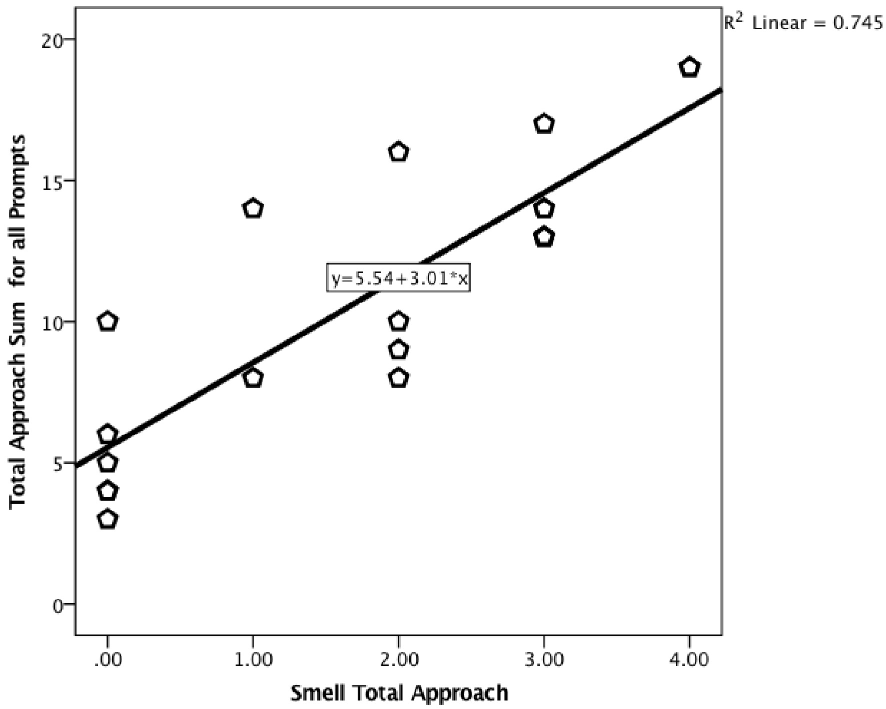


Figure 4.13 Relationship between smell and overall approach-to-modeling scores

The relationship between Smell and overall approach-to-modeling scores in Figure 4.13 is important because it indicates that the approach-to-modeling framework is able to capture students' modeling approaches when engaged in tasks not associated with contextual or conceptual information presented in the course.

Capturing metacognition. Earlier I showed that generative thinking and causal reasoning, components of approach-to-modeling, are associated with academic performance, as measured by tritile rankings. Metacognition, however, the third component that makes up approach-to-modeling scores, was not associated with student academic performance, i.e., tritile rankings ($\chi^2 = 1.66$, df = 4, p = .80). In addition, while generative thinking ($\rho = .290$, n = 100, p = .003) and causal reasoning (ρ =. 289, n = 100, p = .004) were moderately correlated with academic performance (i.e., tritile ranking), metacognition was not (ρ = -.113, n = 100, p = .264). Not having metacognition positively and significantly correlated with academic performance raises concerns about how well I captured metacognition. For example, when observing cross-tabulations I found 13 instances (13% of the samples) of students having low metacognition scores, but high causal reasoning scores. A plausible hypothesis is the students knowing the gene to protein concepts so well that they did not reveal their metacognition when constructing the CFTR and Wolf models.

Additional evidence suggesting that I may not have fully captured metacognition includes the associations and correlations between causal reasoning, metacognition, and generative thinking. Causal reasoning and generative thinking are associated (χ^2 = 45.02, df = 4, p < .001), as are the relationships between metacognition and generative thinking (χ^2 = 12.42, df = 4, p = .014), and metacognition and causal reasoning (χ^2 = 16.3, df = 4, p = .002). In addition, causal reasoning and generative thinking are strongly correlated (ρ =. 640, n = 100, p < .001). The Spearman's rank order correlations between metacognition and generative thinking (ρ =. 242, n = 100, p = .017) and metacognition and causal reasoning, while significant, were weaker (ρ = .259, n = 100, p = .009), which may suggest that I captured smaller elements of metacognition compared to metacognition and generative thinking.

To summarize, I observed within-student and between-student changes in approach-to-modeling scores for different prompts, with a significant decrease in approach-to-modeling scores for the Zika prompt. Approach-to-modeling scores were not significantly associated with student academic performance, as measured by tritile rankings. However, tritile rankings cannot be ignored since a large number of students in tritile 3 comprise the top half of approach-to-modeling scores, and a large number of students in tritile 1 make up the bottom half of approach-to-modeling scores. Furthermore, there is a significant association between tritile rankings and both generative thinking and causal reasoning. Finally, tritile rankings did not influence approach-to-modeling scores for the Zika prompt – all students struggled with the Zika modeling task. Therefore, something was quite different between Zika and the other modeling prompts, and these differences may account for differences in approach-to-modeling scores that are not tied to tritile rankings.

CHAPTER 5: DISCUSSION OF THE RESULTS

In the previous chapter I presented research results corresponding to my research questions. In this chapter I will explain the results in non-statistical terms and connect them to the literature. Similar to Chapter 4, the main sections in this chapter correspond to the research questions.

Variation in Approach-to-Modeling Scores

Research Question: *To what extent do different modeling tasks elicit the different components of deep approaches to modeling (i.e. generative thinking, metacognition, causal reasoning)?*

The approach-to-modeling framework captured variations in student approach-to-modeling scores as modeling tasks changed. I believe that the differences in modeling approaches are due to interactions between students and the prompts, as shown in Figure 5.1.

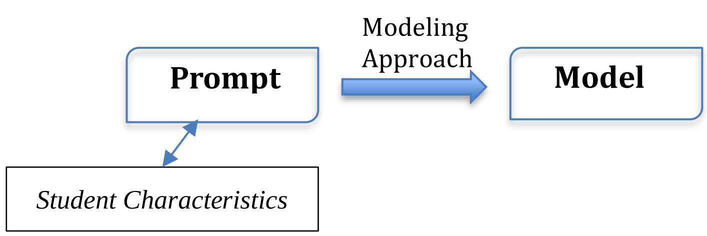


Figure 5.1 Model showing the interactions between student characteristics, characteristics of the modeling prompt, the ways students approach the modeling prompt, and the resulting explanatory model

Figure 5.1 shows that a student's modeling approach (measured with approach-to-modeling scores) is affected by aspects of the modeling prompt as well as by characteristics of that student. Figure 5.1 can be read like an equation in which the model is the output, and the approach to modeling is a variable that can change with the way the prompt is constructed and the ways students interact with the prompt.

In the previous chapter, Table 4.1 showed how each student's approach-to-modeling scores changed as the prompts changed, and that students' approach-to-modeling scores were significantly lower for the Zika prompt. In the remainder of this chapter I will present possible explanations for these findings by addressing the second research question:

Research Question: *What elements of a modeling task may account for differences in student use of deep approaches to modeling?*

In designing the modeling tasks, I focused on conceptual and contextual familiarity, which were discussed in Chapter 3. In this section I will explain how prompt construction may have affected the way students drew upon conceptual and contextual familiarity, and then explore possible interactions between prompt construction and student characteristics by looking at how students' cognitive structures may have influenced their approaches to modeling.

Prompt construction. The choice of words and the amount and type of background information provided in a prompt can affect the ways students engage in modeling tasks. In this section I will discuss each of these prompt components, beginning with word choice, which can activate items held in long-term memory.

Cueing words. Students rely upon cues to activate their thinking process (Wang, Li, Thummaphan & Ruiz-Primo, 2017); examples include the seven components (words) students learned to include in their gene-to-protein models: allele, chromosome, DNA, gene, nucleotide sequence, phenotype and protein. Any prompt containing these words, or words that directed students to these concepts may have initiated recall of information presented in class. When modeling the CFTR prompt for instance, students recognized the words and phrases "genetic information," "expressed," "genetic variation," and "gene" from class. At other times students developed their own associations, such as Sydney who told us, "When I hear the term 'originate', I think about the cellular level".

Key words can help students organize information for cognitive processing (Wang et al., 2017). In the Wolf prompt, for example, students focused on the words "malformed" and "population," which lead to thoughts of mutation, evolution, and breeding; and breeding, mutation, and evolution lead to thoughts about genes. Here are some quotes that illustrate how students constructed gene to protein models for the Wolf prompt based on words contained in the prompt:

- *It's a malformed vertebra;* [pause] *So I guess I would think of why it would appear would be some sort of mutation, and then that mutation gets passed on*.
- *So, I know that the malformed vertebrae first showed up because of a random genetic mutation in one of the wolves and when that happened it either helped him or hurt him and that would kind of decide…whether he could reproduce and pass that gene along, which obviously he could because it became present in the rest of the population, or in some of the population.*
- *It sounds like a gene if just some of the wolves have it.*

In class, students learned to associate "became present" (wording in the Wolf prompt) with inheritance and traits becoming present in a population. "Became present" is academic language that encompasses the processes of reproduction, inheritance, and gene expression (Snow, 2010). According to Snow (2010), academic language, especially in science, comprises nominalizations – words or phrases that represent specific, often complex concepts. For example, "became" refers to a transformation, which in this case is the genetic code contained in the nucleotide sequence; likewise, "present" means "seen" as in a genetic trait that can be observed phenotypically, which implies the role of proteins in the body, and proteins are the products of genes. Students who recognized and understood this academic language quickly recognized the gene-to-protein inheritance concept in the Wolf prompt; those who didn't recognize the academic language responded much like Samantha who stated, "I don't really understand how backs got into this. I don't know what could possibly, [pause] a predator maybe? I'm just going to link it to a predator because I couldn't understand it coming from any other direction."

Eliciting schemata for smell. For the open-ended Smell prompt, which did not include cueing words, 8 students demonstrated moderate levels of metacognition, generative thinking and causal reasoning (approach-to-modeling scores of 3 – 4), 4 students demonstrated moderately low levels of these components (approach-to-modeling scores of 2), and 8 students demonstrated low levels of metacognition, generative thinking and causal reasoning (approach-to-modeling scores of 0 – 1). The students with the higher scores constructed models that addressed diffusion, the olfactory system, or both. The students with the lowest approach-to-modeling scores demonstrated little causal reasoning and did not draw upon prior science knowledge when constructing their models. A lack of cueing words may partially explain why these students attributed smell to the wind blowing a scent in their direction, such as Sarah's and Simba's models shown in Chapter 4; however, a more likely explanation is that these students were unable to access applicable schemata.

All students identified a smell schema that ranged from stinky garbage to fresh-baked cookies to someone grilling hamburgers in the back yard. Those students who drew upon prior science knowledge however focused on the phrases "how a person <u>can smell,</u>" "how a person can <u>smell</u> things <u>from a distance,</u>" or both. Rick, Sydney, and Brian for example focused on "from a distance," which triggered in their minds the process of diffusion, which was linked to their smell schemata. Similarly, Billy, Omid, and Rachel focused on “how a person can smell,” which lead them to model the olfactory system, which also was linked to their smell schemata. Rachel and Billy, who had the highest overall approach-to-modeling scores demonstrated a more complex cognitive structure by connecting their smell schemata to diffusion as well as the olfactory system.

The 8 students with approach-to-modeling scores of 0 – 1 focused on the word “smell” and shared their smell schemata with us by including items such as perfume, donuts, and pasta in their models. Their smell schemata, however did not appear to be connected to any scientific mechanisms. Further support of this claim are their low approach-to-modeling scores across all modeling tasks, and the perfunctory way in which they constructed their CFTR1 models. Therefore, they may not have “learned how to learn” by connecting new ideas to existing ideas and experiences; as a result, they have difficulty problem-solving because they hold many concepts in isolation in long-term memory.

Eliciting schemata for wolf. During a previous unit on population ecology, students watched a presentation that introduced them to the Isle Royale ecosystem. Those students who did not recognize the gene-to-protein or reproduction and inheritance mechanisms in the Wolf prompt may have constructed models based on their Isle Royale schemata, which included images from the presentation. To these students it is possible that “wolf” and “Isle Royale” schemata include rugged terrain, harsh winters, and competition. As a result, their models attributed the malformed vertebrae to fighting between wolves, harsh winters, and rugged terrain, which was the mechanism shown in Elizabeth’s model (Figure 5.2).

Background information. Context refers to the supplemental information provided with a question, such as a description of a natural phenomenon (Wang et al., 2017). Large amounts of science information, especially set within a specific context, can increase problem-solving difficulty by putting too much demand on working memory (Ruiz-Primo & Li, 2015; Wang et al., 2017). Zika was the longest of the prompts, containing a large amount of background

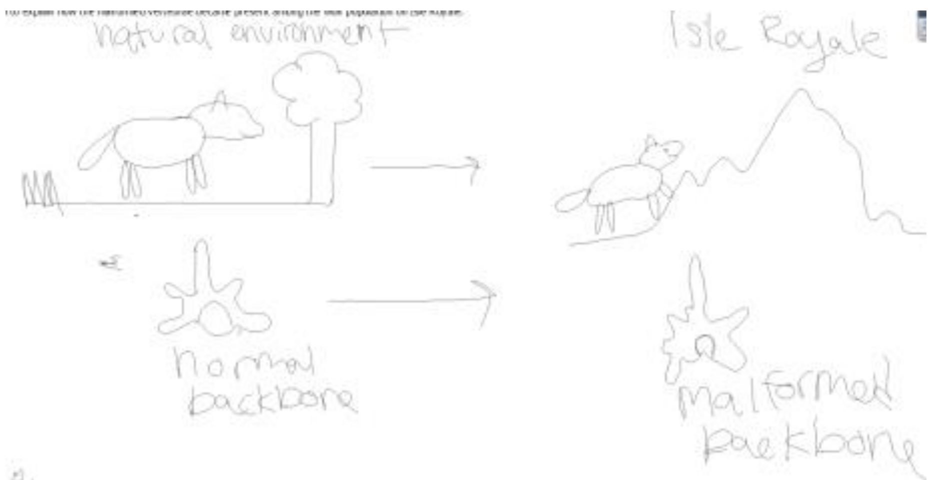


Figure 5.2 Elizabeth's model explaining how the malformed vertebrae became present in the wolf population

information, yet it did not contain any cueing words or phrases students would recognize from class. This may partially explain why Zika had such low approach-to-modeling scores.

Background information in the Zika prompt included (1) where it was first discovered (Zika Forest in Uganda), (2) that the virus consists of a single-stranded RNA genome contained within a protein shell, (3) that the virus is transmitted by mosquitoes belonging to the *Aedes* genus, (4) mosquitoes belonging to the *Aedes* genus are active during the day, (5) symptoms associated with a Zika infection resemble those of a mild cold or flu, (6) people rarely seek medical care because of the mild symptoms (7) people start feeling better in a few days, (8) Zika infections have been linked to microcephaly, (9) microcephaly occurs when women become infected during pregnancy, (10) microcephaly is a rare birth defect, (11) microcephaly refers to incomplete brain development, (12) children with microcephaly have abnormally small brain and head sizes, (13) there have been clinical tests conducted on transmission of the virus, (14) The disease is transmitted through mosquito bites, (15) the disease is transmitted through childbirth, (16) the disease is transmitted through blood transfusions, and (17) the disease is transmitted through sexual contact. In addition to making sense of a large amount of information, the only information that students may have been able to tie back to their introductory biology course was "single-stranded RNA genome," although three students focused on the word "virus" and recalled prior information they had acquired about viruses and viral infections.

The written background information in the Zika prompt was supplemented with a map and timeline, which may have compounded the complexity of the prompt by providing more information and requiring students to interpret the figures. Ruiz-Primo and Li (2015) reported a decrease in student problem-solving performance as the number of supplemental resources increased. In the Zika prompt, the additional resources included (adding to the previous list) (18) the disease was first detected in 1947, (19) the disease also infects monkeys, (20) the first human case of Zika infection was in 1952, (21) between 1952 and 2007 there were 14 cases of human Zika infections, (22) between 1952 and 2007 Zika was found in more than 15 countries in Africa, (23) between 1952 and 2007 Zika was found in more than 15 countries in Asia, and (24) details of how the cases of Zika infection in humans swelled to more than 8,000 spread over 4 continents.

The map in Chapter 3 showing the spread of Zika required students to orient themselves geographically because Africa was not located east of North America, to make sense of the color scheme representing different dates, and to compare text boxes containing information to both the countries shaded on the map and the dates. The presence of the complex diagrams and large amounts of background information, and the absence of cueing words all may have contributed to the low approach-to-modeling scores for Zika. In problem-solving, students take in and temporarily store information in working memory; they also pull related information from long-term memory into working memory to construct a response. The capacity of working memory is limited however, and can be overwhelmed by tasks that are cognitively too demanding as the result of too much information being provided at one time (Deans for Impact, 2015; Goldstein, 2008).

Context. Research by Nehm and Ha (2011) highlights the importance of context in helping students "perceive, use, internally represent, and solve problems" (p. 239) by showing how questions employing different contexts would result in different, yet predictable, student responses. When the context is familiar to what students experienced in class, their problem-solving performance increases (Wang et al., 2017). I saw this with the CFTR and Wolf modeling tasks. In class, students constructed many "between-transfer" task models in which the students constructed models showing how genetic variation originated at the gene level and how the variation was manifested phenotypically through the work of proteins. Each of these modeling activities, while including the same concept, were set in different contexts. Both the Wolf and CFTR prompts were similar "between-transfer" modeling tasks and the student approach-to-modeling scores indicate that the students were able to engage in these modeling tasks. As Nehm and Ha's (2011) research would predict, students had the most success modeling CFTR – the prompt they had seen before.

As stated previously in the discussion of schemata, some students did not recognize the gene-to-protein, genetic variation, or inheritance mechanisms in the Wolf prompt. Instead, they relied on a previous context they had seen in class during a population ecology unit – wolves and Isle Royale. Earlier I used this as an example of drawing upon schemata; those schemata were triggered by the presence of a familiar context. As a reminder, the prompts were based on conceptual and textual familiarity, as shown in Table 5.1.

Table 5.1 Familiarity of Modeling Tasks

Modeling Task	Concepts Covered in Class	Familiar Model	Between-Transfer
Smell	–	–	–
Wolf	+	–	+
CFTR	+	+	–
Zika	+	–	–

Students did better with the Smell prompt because it is a familiar context; they did not do as well with Zika because the context was unfamiliar to most students. As shown in Table 4.8, students who did do well on the Zika prompt found some aspect of the Zika context that they understood, including international travel and having viral infections.

While contextual familiarity is important in problem solving, structural knowledge may be a stronger predictor of problem-solving ability than familiarity (Jonassen, 2000), which I will discuss next.

Student characteristics. Academic performance, while not directly related to approach-to-modeling scores appears to be indirectly related. For example, Billy, Omid, and Rose, all in Tritile 3 were the only students to recognize similar patterns between the Wolf and CFTR prompts, demonstrating the expert practice of seeing organized sets of information rather than separate pieces of information (diSessa, 1993; NRC, 2005). While most students constructed models at a very detailed level by focusing on the CFTR gene and nucleotide sequences, Billy, Omid, and Rose focused on the larger pattern of genetic variation resulting in a new trait in the population and modeled the overall process on a larger scale than the other students. It is not surprising that the three students who used analogical thinking were from tritile 3 because both generative thinking and causal reasoning are strongly linked to content knowledge and each other (Duncan, 2007).

The students in the second tritile showed consistency in their approach-to-modeling scores and generative thinking scores across the CFTR1, CFTR2, and Wolf prompts. At the same time, the CFTR approach-to-modeling scores for the students in tritile 3 dropped between the first and second interviews. Dauer and Long (2015) reported a similar pattern where students in tritile 2 demonstrated stronger cognitive structures and relational knowledge than tritile 3 students 2.5 years after completing a model-based introductory biology course. In their study, Dauer and Long reported that the tritile 2 students demonstrated incomplete to complete understanding of all elements in their models, while half of all tritile 3 students demonstrated an absence of understanding.

Cognitive structure. Cognitive structure influences problem solving (Ifenthaler, Masduski & Seel, 2011) because well-organized cognitive structure enhances the learning and retention process (Ausubel, 1962) and allows students to quickly search for and locate relevant details and relationships held in long term memory (Dauer & Long, 2015). Simba, Eric, Sarah, and Sheri appeared to have memorized their CFTR1 models, but only Sarah confirmed that. Based on their modeling scores and difficulty with all but the CFTR1 modeling tasks, I suspected that they used a shallow approach to learning (Chin & Brown, 2000; Marton and Säljö, 1976) and therefore struggled to access information held in long-term memory, which lead to low approach-to-modeling scores.

We viewed students' models as external representations of their mental models (Louca & Zacharia, 2012), which students construct in their working memory by drawing upon schemata held in long-term memory (Al-Diban, 2008; Collins & Gentner, 1989; Derry, 1996; Gobert & Buckley, 2000; Nersessian, 2010; Seel, 2001). Unless information is learned deeply by connecting new concepts to existing concepts, students will struggle to access the necessary schemata held in long-term memory (Goldstein, 2008) because memorized information is held as discrete, somewhat isolated schemata in long term memory, as seen in Eric's comments while constructing the Wolf model:

> *Since it's an island the wolves would interact with themselves.* [Long pause]. *Some environmental effect, if some cold weather or if they are fighting each other, or if they fall, like those kinds of incidents that might twist the vertebrae to change the form itself, or I would say genetic stuff changed so like they are born with this kind of malformation. Because you know, that can happen. Also, those genes or the temperature can affect it also and also say the other vertebrae besides wolf, or the other animals that exist there defect them. Attacking or preying predators, predation stuff so that will defect from different animals, changing their vertebrae. I think that's about it.*

Robertson (1990) associated differences in problem-solving between low and high performing groups to the use of understanding versus memorization. Ideas acquired through memorization are vulnerable to being forgotten unless triggered by a similar task (Ausubel, 1962), which may account for the large difference in modeling performance between students with the lowest approach-to-modeling scores (ranging from 3 – 6) and those with the highest approach-to-modeling scores (16 – 19). Memorization may also explain why Sarah, Sheri, Simba, and Eric (students with the lowest approach-to-modeling scores) tackled the CFTR1 modeling task in a scripted, non-reflective manner, yet struggled to generate a model for the Smell, Wolf and CFTR2 prompts. How much someone knows about a domain is important to understanding the problem and generating solutions. However, that domain knowledge must be well integrated in order to support problem solving (Jonassen, 2000).

Correctness of Student Models

Research Question: Does student use of deep approaches to modeling result in scientifically correct models? If not, what may account for the discrepancy?

While the focus of this research is to understand the way students approach modeling tasks, I cannot ignore the role of model correctness, even though a strong modeling approach may not result in a correct model (Duncan, 2007). Neither can I ignore the occurrence of students producing correct models while displaying little to no metacognition or generative thinking. Two factors may account for this. First, I did not effectively capture metacognition, which I describe in detail in the next section. Second, the prompt was not challenging enough for the students and therefore they did not need to be metacognitive or to think generatively.

In both the Smell and Wolf modeling tasks I observed students who provided a correct answer without relying upon generative thinking or metacognition, such as Rick. Conversely, I observed students (e.g. Sandra) who did not produce a correct model even though they relied upon generative thinking and metacognition. Then again, there were students such as Sydney, Billy and Omid who produced correct models while also relying upon metacognition and generative thinking. The diversity in student approaches and corresponding range in model correctness scores may suggest that the modeling prompts were (a) too difficult, (b) too easy or (c) just right *depending on the student*. For example, while Merle produced a correct CFTR2 model and relied upon metacognition and generative thinking, Eric, Sheri and Simba used generative thinking to a limited extent and were unable to produce a correct model. This may be partly due to the modeling prompt. If a prompt is too difficult (e.g. includes too much background information or not enough cueing words), some students may not be able to produce a correct model or make use of generative thinking and metacognition. Conversely, if a prompt is too easy, some students may be able to produce a correct model through simple recall of information held in long term memory. For some students, the wolf prompt was too easy because they noticed the similarities to the CFTR1 prompt (e.g. Merle, Brian, Rose and Paul), while others struggled with the prompt (e.g. Eric, Sheri, Simba, Sandra). Interestingly, some students (e.g. Simba) produced correct CFTR1 models using little generative thinking or metacognition, but incorrect models for the other prompts where they appeared to use metacognition and generative thinking. It may be that these students learned about gene-to-protein mechanisms through memorization and therefore did not expand their cognitive structure by building connections to information and experiences held in long-term memory, or by building a new schema.

The Zika prompt (Table 4.7e) may be an example of a prompt that was too difficult, with half of the students not producing a correct model and not exhibiting generative thinking or metacognition. Therefore, the extent to which students produce correct models may be influenced by the modeling task and our ability to capture student use of generative thinking and metacognition. Furthermore, student use of generative thinking and metacognition does not necessarily result in a correct model, which may be an indication of the difficulty of the modeling task.

Capturing Differences in Student Approaches to Modeling

The approach-to-modeling framework captured differences within individual students as modeling tasks changed, as evidenced by no student having the same approach score for every prompt. Furthermore, the framework revealed differences between modeling tasks, with the Zika prompt resulting in significantly lower approach-to-modeling scores than the other prompts. In addition to capturing differences in the way students approach modeling, the approach-to-modeling framework has some strengths as a research tool, but also requires some modifications; I describe the strengths and challenges in this section.

Approach-to-modeling framework: Strengths. Rather than relying upon student self-reports, the approach-to-modeling framework is based on student behaviors while they are engaged in modeling tasks. In addition, the coding protocols were designed to minimize the need for inference on the part of the researchers. In the following paragraphs, I address these attributes as they pertain to the three components that comprise an approach-to-modeling score.

Metacognition. One of the difficulties in capturing metacognition in a learning environment is the need to infer what students are thinking based on their actions. The metacognition component of the approach-to-modeling framework, unlike self-report questionnaires and Likert-type tools (e.g. Anderson, Nashon & Thomas, 2009; Armour-Thomas & Haynes, 1988), relies upon what the researchers observe. For example, when students double-checked the prompt during modeling I counted it as an indication of metacognition only if the students provided evidence that they did so to assess their modeling approach, such as comparing their models to what was required in the prompt. Similarly, when students checked for errors during modeling I counted it as an indication of metacognition only when the students provided evidence of error checking by articulating or fixing errors (Chin & Brown, 2000a; Goos, 2002; Meier et al., 2006).

Generative thinking. Chin and Brown (2000) defined generative thinking as the ability to generate an answer when an immediate, ready-made solution to a problem was not available through simple recall. They observed that learners displaying generative thinking spontaneously constructed plausible answers supported with specific examples, real life experiences, and self-generated analogies. In addition, learners displaying generative thinking elaborated upon ideas and hypothesized, while students displaying weak generative capacity extended little effort in solving novel problems and gave evasive responses.

We were unable to reliably capture all of the behaviors that Chin and Brown identified, such as “effort.” Coding “effort” was highly subjective and imprecise so I did not include it in our final coding protocols. Similarly, I struggled to operationalize “evasive responses,” which I defined as students not answering the “how” or “why” questions in the prompts. In the end, I captured evasive responses with the causal reasoning codes because “how” and “why” questions are related to mechanistic explanations (Osborne & Patterson, 2011).

While I developed generative thinking codes that were readily observable, I also relied upon clues provided by the students. For instance, when tapping into prior experiences students would often use a personal pronoun and a tone of familiarity such as, “I used to breed dogs…” or “When I travel internationally…”. (As I mentioned in the methods section, I observed only two instances of students relying upon prior experiences when constructing models, so I did not include prior experiences in our data analysis.) Likewise, when drawing upon prior knowledge or using analogical mapping, I relied upon student clues such as, “My chemistry instructor demonstrated how odors move through a room…”, or “I’ve seen something like this before.”

When students did not reveal their sources of information, I had to infer that it came from another class or was covered in the current biology course based on the sophistication of the concept. Omid, for instance stated, “So these chemicals will bind to the chemoreceptors on the nasal mucosa and it will send signals straight to the olfactory bulb [pause] and you have sense of smell.” Likewise, Anna noted, “[At the] CFTR gene it could go either way if it's [pause] I think it was homozygous recessive. Then it would come out as, you would have cystic fibrosis. And then if you were heterozygous or homozygous dominant you would be normal." In both instances, I concluded that the students were drawing upon prior knowledge based on the sophistication and specificity of their answers. In other words, I made the assumption that most students would not be able to correctly state the inheritance patterns that would lead to cystic fibrosis without having formal education experiences. In summary, even though I developed codes that required little inference, I could not eliminate inference altogether.

Causal reasoning. In the Methods chapter I explained the iterations I went through to capture causal reasoning, beginning with an attempt to capture entities and activities (Russ et al., 2008). In the end, I settled upon a holistic rubric based upon the scientific sophistication of student explanations, which was significantly associated with generative thinking and student academic performance, as expected (Duncan, 2007; Trout, 2007). When giving causal accounts of scientific phenomena, I looked for students to explain "why" or "how" phenomena occurred by providing a chain of reasoning – grounded in scientific ideas, facts, and theories – that connected unseen mechanisms to the observed phenomena (Braaten & Windschitl, 2011; Cavallo, 1996; Lipton, 2004; NRC, 2012; Osborne & Patterson, 2011; Reiser, Berland & Kenyon, 2012). Using this definition of causal reasoning resulted in very few instances where a student did not clearly fit into one of the three categories: Explanation absent, Explanation incomplete, Explanation complete.

Approach-to-modeling framework: Challenges. Metacognition is difficult to capture. I measured what I could observe empirically without inference; therefore, other metacognitive activities may have been present, but were unobservable. As a result, the scores I reported likely underestimated metacognition. In the remainder of this section I will discuss aspects of metacognition that I may not have been able to detect.

According to the literature, metacognition consists of three phases: planning, implementation and evaluation (Anderson & Nashon, 2007; Davidson, Deuser & Sternberg, 1996; Goos, 2002; Meijer et al., 2006); and in each phase students monitor their work and make adjustments as needed (Efklides, 2009; Flavell, 1979; Goos, 2002; Son, 2013).

Planning. Planning may include the following behaviors that I did not count because they required inference: (1) Many of our subjects stared at the board for an extended period of time. (2) Many of our subjects reread the prompt multiple times before putting the pen to the board or talking about their ideas. In neither case did the students provide evidence that they were planning a modeling approach; while staring at the board they may have been rereading the prompt (which was located at the top of the board) or deciding what color line to use. Similarly, the students may have reread the prompt in order to orient themselves to the task. Therefore, while there may have been metacognitive activity taking place, I did not capture it.

Meijer et al. (2006) suggest that metacognitive activity can be inferred from cognitive activity. For example, rereading a paragraph, which is a cognitive activity, may represent metacognition if the decision to reread the passage was due to not fully comprehending it the first time. Students shared many thoughts during modeling that could have been interpreted as planning; however, they also could have been interpreted other ways. For example, Sarah stated:

> *I'm just thinking about how to make this longer than three boxes, because normally…when you have the population you'll have like the alleles that cause the change, and then they came from like genes, and then eventually you get your phenotype...*

This example may indicate planning because Sarah is planning how to include multiple components in her model. On the other hand, she was frustrated at not being able to recall enough information to build a model with more than three components, which would have been generative thinking.

Another example was this statement by Billy: “It’s a malformed vertebra; so, I guess I would think of why it would appear would be some sort of mutation, and then that mutation gets passed on.” In this example, Billy could have been planning a way to show how the mutation appeared, or how the trait was passed on through reproduction. Conversely, he may have been displaying analogical reasoning since the origin and expression of genetic variation were the big ideas woven through the biology course he was taking.

We piloted a coding method to capture metacognitive planning efforts by restricting planning to all thoughts and actions that took place prior to the students putting the pen on the board. Unfortunately, student thinking is not a linear process of planning-implementation-evaluation; in fact, some students appeared to be engaged in multiple cycles of planning-implementation-evaluation. In addition, coding “before inking” did not take into account students who planned by drawing (e.g. Punnett squares, genetic crosses, maps) or making lists of words. Finally, planning can be ongoing, similar to when playing chess or checkers; after drawing each component students may have engaged in planning their next move. In the end, I decided that it was more important to capture students’ thinking about their modeling approach than trying to decide if their thinking was part of planning, implementation or evaluation. What I initially tried to code as planning ended up fitting into “checking the prompt” or “checking for errors.”

Evaluation. I struggled to differentiate between monitoring and evaluation. As a reminder, “monitoring” refers to students continuously assessing their work, and “controlling” refers to actions taken to correct problems detected during monitoring (Dinsmore, Alexander & Loughlin, 2008; Flavell, 1979; Goos, 2002; Meijer et al., 2006). Our initial coding protocols had monitoring occurring during modeling and evaluation occurring at the end of modeling. Students, however rarely said when they were finished modeling. Students could stand back and observe their models for both monitoring their strategy and evaluating their finished model. This temporal strategy for differentiating between aspects of metacognition did not work because of the iterative nature of metacognitive activity. As with planning, I decided it was more important to capture students' thinking about their modeling approach than trying to decide if their thinking was evaluation or monitoring. What I initially tried to code as evaluation or monitoring ended up fitting into "checking the prompt" or "checking for errors."

Monitoring. Monitoring refers to checking over the modeling strategy and can take on many different forms. For example:

A. Erasing. Some students indicated that they did not like what they were doing, e.g., "I don't like this", erased parts of the model and then wrote different components or relationships on the board – rarely indicating what it was that they didn't like. This could have indicated metacognitive activity if they erased an element in their model that was out of sequence. However, it may not have indicated metacognitive activity because they may not have liked the spatial arrangement on the board, the quality of their writing or drawings, or the thickness of a line.

B. Pauses. Students often would pause or hesitate. This could have indicated metacognitive activity if they were reflecting on their work. However, they could have simply been resting their hands by taking a short break from writing.

Assessing progress. In our preliminary coding protocols, I attempted to identify the extent to which students assessed their modeling progress. Some students clearly shared their assessment activities by stating for example, "'Using what you know about zika' [pause] I think I got the 'how' pretty much done." Other students would simply stare at the board and not share their thoughts. In most cases students assessed their modeling progress by referring back to the prompt, as illustrated in the previous quote, so I captured modeling progress with the "checking the prompt"

To summarize, the approach-to-modeling framework was able to detect differences in the way students engaged in modeling tasks, although it most likely underestimated the metacognitive component. Metacognition is difficult to capture, so I attempted to assess it through direct observation. I also attempted to minimize the need to infer metacognitive behavior based on students' non-verbal behaviors, choosing to capture them with two over-arching codes: "checking the prompt" or "checking for errors." I may have tried to capture the nuances of metacognition if more students shared their thoughts as Taylor did when she said, "no, that's wrong, it should go…."

Students had the lowest approach-to-modeling scores for the Zika prompt. The large amount of background information, the use of complex graphics, and the lack of cueing words may partly explain why the students struggled with the Zika prompt. Similarly, I saw in the other prompts the importance of specific words or phrases that would elicit schemata that students could use to solve the problem. When students did not recognize cueing words, they relied on the context of the problem to engage schemata. I observed this with the Wolf prompt. Finally, I witnessed the difficulties in problem-solving (modeling) that arise when students memorize content. These students constructed their CFTR1 models in a machine-like fashion; when they got to the other prompts however, especially CFTR2, they struggled to put coherent thoughts on the board.

CONCLUDING THOUGHTS

In these concluding thoughts, I draw implications from my study for (1) designing modeling prompts to formatively and summatively assess student learning, and (2) enhancing instructional techniques that support deep approaches to modeling.

Designing Modeling Prompts

Students approach modeling differently depending on the construction of the prompt. This is important for several reasons. First, scientific modeling can support student understanding of the core ideas in science (Schwarz et al., 2009) as well as the nature of science (NRC, 2012). Therefore, knowing how to construct prompts that lead to deep approaches to modeling may ultimately support student learning. Secondly, modeling can reveal student thinking, allowing instructors to formatively and summatively assess student learning, but only if the prompts are designed to reveal student thinking and modeling practices.

With modeling serving multiple functions in a learning environment, instructors need guidance in designing modeling tasks. My findings suggest that prompts should include enough background information to facilitate student construction of mental models, but not so much information that it makes problem-solving more difficult (Ruiz-Primo & Li, 2015). I used the Zika prompt to test how students would approach a problem that they were unfamiliar with. Because it was an emerging news story and unlikely that students would have much knowledge about it, I provided a lot of background information. The results indicated that even the strongest modelers struggled with the prompt, possibly due to the large amount of information. Therefore, it is important to consider the how much information is important to scaffold students as the model and how much overwhelms them.

Context. Another possible explanation for the low modeling scores for the Zika prompt was a lack of personal connection. With the exception of those who engaged in travel to locations known to have illnesses related to mosquitoes, or those students with knowledge of viruses, the context for the Zika prompt was unfamiliar to most students. I observed the importance of context with the Wolf prompt where 7 students drew upon contextual information they had seen in class to construct their models. Even though the contextual information came from a different instructional unit and the students' models did not adequately address the prompt, it was fascinating to observe how the students gravitated toward the familiar context. When designing prompts, it is important to consider what students may draw on to ground their understandings. This is especially true for prompts that are more distal from models that were studied in class.

Cueing words. A third possible explanation for the low Zika modeling scores was the absence of cueing words. Prompts should contain cueing words used in class and related to the modeling task in order to activate students' schemata. Instruction must also be clear regarding cueing words since cueing words, like contexts, may elicit unintended schemata (Ruiz-Primo & Li, 2015). In other words, instructors should explain the many concepts contained in cueing words (Snow, 2010) – concepts that students will incorporate in their models when they see a familiar cueing word. Unless modeling tasks used for assessments include familiar contexts and cueing words, the instructor may be assessing something other than student content knowledge and modeling skills.

Constructing prompts that contain familiar contexts, familiar cueing words, and the right amount of background information will help students construct models that address the prompt.

As I saw with the Zika prompt, attempting to assess student knowledge through the use of unfamiliar prompts, prompts containing large amounts of background information, or prompts lacking cueing words and a familiar context, may not accurately reflect students' mental models, conceptual understanding, or modeling skills. Even the best modelers and 4.00 students struggled with the Zika prompt.

Supporting Deep Approaches to Modeling

My findings echo a common theme in the science education literature – the importance of helping students connect new information to prior knowledge and experiences. Although my study did not focus on model-based instruction, the findings may provide some guidance for how to structure teaching to support deep approaches to modeling. This is especially important given my findings that students from different tritiles demonstrated both strong and weak approaches to modeling, showing that modeling may provide opportunities for students from all academic performance levels to be successful. By encouraging the use of generative thinking, teaching metacognitive skills, and providing guidelines for causal reasoning, instructors may help all students assemble and use schemata to engage successfully in problem-solving, regardless of their [prior achievement/prior performance/academic standing].

Generative thinking. During the interviews, I observed students who seemed to have memorized information given in class. They had not integrated the new information with prior knowledge or experiences and as a result, they possessed fragmented, disconnected pieces of information that inhibited problem solving. These students performed well constructing their CFTR1 models, but struggled with the other models. Students will need to be supported as they make sense of new information by connecting it to existing knowledge and experiences.

Causal reasoning /explanations. While there is no widely-accepted definition of a scientific explanation, there is some consensus that it is causal – consisting of a sequence of activities (microscopic and/or macroscopic) that produce observed phenomena. Getting students to engage in causal reasoning may require direct instruction and constant reminders, through scaffolds, that models should be explanatory; in other words, show the mechanisms that lead to a phenomenon. Scaffolds are temporary supports that fade over time (Krajcik, Slotta, McNeill & Reiser, 2008), and may be verbal or written, presented as questions or prompts, and used to support one learner or multiple learners (Quintana et al., 2004). One scaffolding suggestion is to include written reminders in the prompts, such as "remember to…" or "does your model…?"

Metacognition. Students engage in metacognition often unconsciously (Veenman, 2012), but instruction can bring the use of metacognitive skills to the conscious level. Using scaffolds, students can be taught to monitor their work (which most do unconsciously) for "red flags" (Goos, 2002), and instructors can provide written supports in the form of questions students should ask themselves while modeling: (1) Am I addressing the "how" or "why" question in the prompt? (2) Does the model I've drawn match my understanding of the problem/system (i.e., match the 'image in my head')? (3) Did I check the overall model for clarity and accuracy? (4) Am I checking the smaller components of the model for clarity and accuracy?

All in all, these findings suggest the importance of supporting students in modeling as well as struggles that students may face. The framework (i.e., approaches-to-modeling) used in this study provides a way of better understanding how students approach modeling tasks. The findings provide insights into how to best structure modeling prompts as well as how to scaffold students as they engage in modeling in the classroom.

REFERENCES

REFERENCES

Akturk, A. O., & Sahin, I. (2011). Literature review on metacognition and its measurement. Procedia-Social and Behavioral Sciences, 15, 3731-3736.

Al-Diban, S. (2008). Progress in the diagnostics of mental models. In *Understanding Models for Learning and Instruction: Essays in Honor of Norbert M. Seel*. D. Ifenthaler, P. Pirnay-Dummer, & M. J. Spector (eds) (pp. 81-101). New York: Springer.

Anderson, D., & Nashon, S. (2007). Predators of knowledge construction: Interpreting students' metacognition in an amusement park physics program. *Science Education, 91*(2), 298-320. doi:10.1002/sce.20176

Anderson, D., Nashon, S. M., & Thomas, G. P. (2009). Evolution of research methods for probing and understanding metacognition. *Research in Science Education, 39*(2), 181-195. doi:10.1007/s11165-007-9078-1

Armour-Thomas, E., & Haynes, N. M. (1988). Assessment of metacognition in problem solving. *Journal of Instructional Psychology, 15*, 87-93.

Ausubel, D. P. (1962). A subsumption theory of meaningful verbal learning and retention. *The Journal of General Psychology*, *66*(2), 213-224.

Baeten, M., Kyndt, E., Struyven, K., & Dochy, F. (2010). Using student-centered learning environments to stimulate deep approaches to learning: Factors encouraging or discouraging their effectiveness. *Educational Research Review*, *5*(3), 243-260. doi:10.1016/j.edurev.2010.06.001

Bamberger, Y. M., & Davis, E. A. (2011). Middle-school science students’ scientific modelling performances across content areas and within a learning progression. *International Journal of Science Education, 35*(2), 1-26. doi:10.1080/09500693.2011.624133

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological bulletin*, *128*(4), 612. doi:10.1037//0033-2909.128.4.612

Berland, L. K., & Reiser, B. J. (2009). Making sense of argumentation and explanation. *Science Education*, *93*(1), 26-55. doi:10.1002/sce.20286

Biggs, J. B. (1987). *Student approaches to learning and studying.* [Research Monograph]. Australian Council for Educational Research Ltd. Distributed by ERIC Clearinghouse.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (structure of the observed learning outcome)*. New York: Academic Press.

Biggs, J., Kember, D., & Leung, D. Y. (2001). The revised two-factor study process questionnaire: R-SPQ-2F. *British Journal of Educational Psychology*, *71*(1), 133-149. doi:10.1348/000709901158433

Billing, D. (2007). Teaching for transfer of core/key skills in higher education: Cognitive skills. *Higher education*, *53*(4), 483-516. doi:10.1007/s10734-005-5628-5

Braaten, M. & Windschitl, M. (2011). Working toward a stronger conceptualization of scientific explanation for science education. *Science Education*, *95*(4), 639-669. doi:10.1002/sce.20449

Brewer, C. A., and Smith, D. (Eds). (2011). *Vision and change in undergraduate biology education: A call to action.* Washington, DC: American Association for the Advancement of Science.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-42. doi:10.2307/1176008

Buckley, B. C. (2000). Interactive multimedia and model-based learning in biology. *International Journal of Science Education, 22*(9), 895-935. doi:10.1080/095006900416848

Case, J. M., & Marshall, D. (2009). Approaches to learning. In M. Tight, K. H. Mok, J. Huisman, and C. Morphew (Eds.) *The Routledge International Handbook of Higher Education* (pp. 9-22). New York: Routledge. doi:10.4324/9780203882221

Cavallo, A. M. (1996). Meaningful learning, reasoning ability, and students' understanding and problem solving of topics in genetics. *Journal of Research in Science Teaching, 33*(6), 625-656.

Chan, C. C., Tsui, M. S., Chan, M. Y. C., & Hong, J. H. (2002). Applying the structure of the observed learning outcomes (SOLO) taxonomy on student's learning outcomes: An empirical study. *Assessment & Evaluation in Higher Education, 27*(6), 511-527. doi:10.1080/0260293022000020282

Chin, C., & Brown, D. E. (2000). Learning in science: A comparison of deep and surface approaches. *Journal of research in science teaching*, *37*(2), 109-138. doi:10.1002/(SICI)1098-2736(200002)37:2<109::AID-TEA3>3.0.CO;2-7

Chin, C., & Brown, D. E. (2000a). Learning deeply in science: An analysis and reintegration of deep approaches in two case studies of grade 8 students. *Research in Science Education, 30*(2), 173-197. doi:10.1007/BF02461627

Clement, J. (2000). Model based learning as a key research area for science education. *International Journal of Science Education, 22*(9), 1041-1053. doi:10.1080/095006900416901

Clement, J., & Rea-Ramirez, M. A. (2008). *Model based learning and instruction in science*. New York: Springer.

Coll, R.K. and Lajium, D. (2011). Modeling and the future of science learning. In M. S. Khine & I. M. Saleh (Eds). *Models and modeling: Cognitive Tools for Scientific Enquiry* (pp. 3 - 19). Springer: Netherlands.

Coll, R. K., & Treagust, D. F. (2003). Investigation of secondary school, undergraduate, and graduate learners' mental models of ionic bonding. *Journal of Research in Science Teaching, 40*(5), 464-486. doi:10.1002/tea.10085

Collins, A., & Gentner, D. (1987). How people construct mental models. In D. C Holland & N. Quinn (Eds), *Cultural models in language and thought* (pp. 243-265). Cambridge [Cambridgeshire]; New York: Cambridge University Press.

Crawford, B. A., & Cullin, M. J. (2004). Supporting prospective teachers' conceptions of modelling in science. *International Journal of Science Education, 26*(11), 1379-1401. doi:10.1080/09500690410001673775

Dauer, J. T., & Long, T. M. (2015). Long-term conceptual retrieval by college biology majors following model-based instruction. *Journal of Research in Science Teaching, 52*(8), 1188-1206. doi:10.1002/tea.21258

Dauer, J. T., Momsen, J. L., Speth, E. B., Makohon-Moore, S. C., & Long, T. M. (2013). Analyzing change in students' gene-to-evolution models in college-level introductory biology. *Journal of Research in Science Teaching, 50*(6), 639-659. doi:10.1002/tea.21094

Davidson, J. E., Deuser, R., & Sternberg, R. J. (1996). The role of metacognition in problem solving. In J. Metcalfe and A. P. Shimamura (Ed.), *Metacognition: Knowing about Knowing* (pp. 207–226). Cambridge: MIT Press.

Deans for Impact (2015). *The Science of Learning.* Austin, TX: Deans for Impact.

Derry, S. J. (1996). Cognitive schema theory in the constructivist debate. *Educational Psychologist, 31*(3-4), 163-174. doi:10.1080/00461520.1996.9653264

Dinsmore, D. L., & Alexander, P. A. (2012). A critical discussion of deep and surface processing: What it means, how it is measured, the role of context, and model specification. *Educational Psychology Review, 24*(4), 499-567. doi:10.1007/s10648-012-9198-7

Dinsmore, D. L., Alexander, P. A., & Loughlin, S. M. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review, 20*(4), 391-409. doi:10.1007/s10648-008-9083-6

Diseth, Å., & Martinsen, Ø. (2003). Approaches to learning, cognitive style, and motives as predictors of academic achievement. *Educational Psychology, 23*(2), 195-207. doi:10.1080/01443410303225

diSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction, 10*(2-3), 105-225. doi:10.1080/07370008.1985.9649008

Draper, S. W. (2009). Catalytic assessment: Understanding how MCQs and EVS can foster deep learning. *British Journal of Educational Technology, 40*(2), 285-293. doi:10.1111/j.1467-8535.2008.00920.x

Duncan, R. G. (2007). The role of domain-specific knowledge in generative reasoning about complicated multileveled phenomena. *Cognition and Instruction, 25*(4), 271-336. doi:10.1080/07370000701632355

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Los Angeles: SAGE.

Efklides, A. (2009). The new look in metacognition: From individual to social, from cognitive to affective. In C. B. Larson (Ed.) *Metacogntion: New Research Development*. New York: Nova Science Publishers.

Entwistle, N. J., & Ramsden, P. (1982). *Understanding student learning.* New York: Nichols Publishing. Distributed by ERIC Clearinghouse.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*(3), 215-251. doi:10.1037/0033-295X.87.3.215

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, Mass: MIT.

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. London: Sage.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906-911. doi:10.1037//0003-066X.34.10.906

Floyd, K. S., Harrington, S. J., & Santiago, J. (2009). The effect of engagement and perceived course value on deep and surface learning strategies. *Informing Science: The International Journal of an Emerging Transdiscipline, 12*, 181-190.

Gilbert, J. K. (2004). Models and modelling: Routes to more authentic science education. *International Journal of Science and Mathematics Education, 2*(2), 115-130. doi:10.1007/s10763-004-3186-4

Gobert, J. D., & Buckley, B. C. (2000). Introduction to model-based teaching and learning in science education. *International Journal of Science Education, 22*(9), 891-894. doi:10.1080/095006900416839

Goel, A., & Stroulia, E. (1996). Functional device models and model-based diagnosis in adaptive design. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 10*(04), 355–370. doi: 10.1017/S0890060400001670

Goldstein, E. B. (2008). *Cognitive psychology: Connecting mind, research and everyday experience* (2nd ed.). Belmont, CA: Thomson/Wadsworth.

Goos, M. (2002). Understanding metacognitive failure. *The Journal of Mathematical Behavior, 21*(3), 283-302. doi:10.1016/S0732-3123(02)00130-X

Goos, M., & Galbraith, P. (1996). Do it this way! metacognitive strategies in collaborative mathematical problem solving. *Educational Studies in Mathematics, 30*(3), 229-260. doi:10.1007/BF00304567

Greca, I. M., & Moreira, M. A. (2000). Mental models, conceptual models, and modelling. *International Journal of Science Education, 22*(1), 1-11. doi:10.1080/095006900289976

Harrison, A. G., & Treagust, D. F. (2000). A typology of school science models. *International Journal of Science Education, 22*(9), 1011-1026. doi:10.1080/095006900416884

Hattie, J.A.C., & Brown, G.T.L. (2004). *Cognitive processes in asTTle: The SOLO taxonomy*. asTTle Technical Report #43, University of Auckland/Ministry of Education.

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review, 16*(3), 235-266. doi:10.1023/B:EDPR.0000034022.16470.f3

Ifenthaler, D., Masduki, I., & Seel, N. M. (2011). The mystery of cognitive structure and how we can detect it: Tracking the development of cognitive structures over time. *Instructional Science, 39*(1), 41-61. doi:10.1007/s11251-009-9097-6

Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development, 48*(4), 63-85. doi:10.1007/BF02300500

Jonassen, D. H., Beissner, K., & Yacci, M. (1993). *Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge*. Hillsdale, N.J: L. Erlbaum Associates.

Jonassen, D., Strobel, J., & Gottdenker. J. (2005). Model building for conceptual change, *Interactive Learning Environments, 13*(1-2), 15-37, DOI: 10.1080/10494820500173292

Justi, R. S., & Gilbert, J. K. (2002). Modelling, teachers' views on the nature of modelling, and implications for the education of modellers. *International Journal of Science Education, 24*(4), 369-387. doi:10.1080/09500690110110142

Kahn, S. (2008). Co-construction and model evolution in chemistry. In Clement, J. J., & Rea-Ramirez, M. A. (Eds.) *Model based learning and instruction in science*. Netherlands: Springer.

Khan, S. (2011). What's missing in model-based teaching. *Journal of Science Teacher Education, 22*(6), 535-560. doi:10.1007/s10972-011-9248-x

Kim, Y. R., Park, M. S., Moore, T. J., & Varma, S. (2013). Multiple levels of metacognition and their elicitation through complex problem-solving tasks. *The Journal of Mathematical Behavior*, *32*(3), 377-396. doi:10.1016/j.jmathb.2013.04.002

Krajcik, J., Slotta, J. D., McNeill, K. L., & Reiser, B. J. (2008). Designing learning environments to support students' integrated understanding. In Y. Kali, M. C. Linn & J. E. Roseman (Eds), *Designing coherent science education: Implications for curriculum, instruction, and policy* (pp. 39-64). New York: Teachers College, Columbia University.

Lee, O., Fradd, S. H., & Sutman, F. X. (1995). Science knowledge and cognitive strategy use among culturally and linguistically diverse students. *Journal of Research in Science Teaching, 32*(8), 797-816. doi:10.1002/tea.3660320804

Legare, C. H. (2014). The contributions of explanation and exploration to children's scientific reasoning. *Child Development Perspectives, 8*(2), 101-106. doi:10.1111/cdep.12070

Lehrer, R., & Schauble, L. (2012). Seeding evolutionary thinking by engaging children in modeling its foundations. *Science Education, 96*(4), 701-724. doi:10.1002/sce.20475

Leithwood, K., McAdie, P., Bascia, N., & Rodrigue, A. (2006). Teaching for deep understanding: What every educator should know. Corwin Press, Thousand Oaks, CA.

Lipton, P. (2004). What good is an explanation. In J. Cornwell (Ed.), *Explanations: Styles of explanation in science (pp. 1 – 21)*. Oxford;New York: Oxford University Press.

Löhner, S., van Joolingen, W. R., Savelsbergh, E. R., & van Hout-Wolters, B. (2005). Students' reasoning during modeling in an inquiry learning environment. *Computers in Human Behavior*, *21*(3), 441-461. doi:10.1016/j.chb.2004.10.037

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences, 10*(10), 464-470. doi:10.1016/j.tics.2006.08.004:

Louca, L. T., & Zacharia, Z. C. (2012). Modeling-based learning in science education: Cognitive, metacognitive, social, material and epistemological contributions. *Educational Review, 64*(4), 471-22. doi:10.1080/00131911.2011.628748

Lublin, J. (2003). Deep, surface and strategic approaches to learning. Centre for Teaching and Learning. UCD Dublin, p. 1-11.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*(1), 1-25. doi:10.1086/392759

Marton, F., & Säljö, R. (1976). On Qualitative Differences in Learning: I—Outcome and process. *British journal of educational psychology*, *46*(1), 4-11. doi:10.1111/j.2044-8279.1976.tb02980.x

Mayer, R. (1996). Learners as information processors: Legacies and limitations of educational psychology's second. *Educational Psychologist, 31*(3), 151-161. doi:10.1207/s15326985ep3103&4_1

Mayer, R. E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science, 26*(1-2), 49-63. doi:10.1023/A:1003088013286

Mayer, R. E. (2002). Rote versus meaningful learning. *Theory into Practice, 41*(4), 226-232. doi:10.1207/s15430421tip4104_4

Mayer, R. E. (2010). Merlin C. Wittrock's enduring contributions to the science of learning. *Educational Psychologist, 45*(1), 46-50. doi:10.1080/00461520903433547

Megowan-Romanowicz, C. 2011. Helping students construct robust conceptual models. In M. S. Khine, & I. M. Saleh (Eds.), *Models and modeling: Cognitive tools for scientific enquiry*. P. 99-120. Dordrecht: Springer.

Meijer, J., Veenman, M. V. J., & van Hout-Wolters, Bernadette H. A. M. (2006). Metacognitive activities in text-studying and problem-solving: Development of a taxonomy. *Educational Research and Evaluation, 12*(3), 209-237. doi:10.1080/13803610500479991

Mulder, Y. G., Lazonder, A. W., & Jong, d., A.J.M. (2015;2014;). Key characteristics of successful science learning: The promise of learning by modelling. *Journal of Science Education and Technology, 24*, 168-177. doi:10.1007/s10956-014-9537-1

Namdar, B., & Shen, J. (2015). Modeling-oriented assessment in K-12 science education: A synthesis of research from 1980 to 2013 and new directions. *International Journal of Science Education, 37*(7), 993-1023. doi:10.1080/09500693.2015.1012185

National Research Council, 1999. *How People Learn: Bridging Research and Practice.* S. M. Donovan, J. D. Bransford, & J. W. Pellegrino (Eds.), Washington, DC: The National Academies Press.

National Research Council, 2000. *How People Learn: Brain, Mind, Experience, and School.* J. D. Bransford, L. Brown, & R. Cocking (Eds.), Washington, DC: The National Academies Press.

National Research Council. (2005). *How Students Learn: Science in the Classroom.* Committee on *How People Learn*, a targeted report for teachers, M.S. Donovan and J.D. Bransford, (Eds.), Washington, DC: The National Academies Press.

National Research Council (2012). A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. Washington, DC: National Academies Press.

Nehm, R. H., & Ha, M. (2011). Item feature effects in evolution assessment. *Journal of Research in Science Teaching, 48*(3), 237-256. doi:10.1002/tea.20400

Nersessian, N. J. (2010;2014;). *Creating scientific concepts*. Cambridge: Bradford Books, Publishers, Incorporated.

Newton, D. P. (2000). *Teaching for Understanding: What it is and How to do it*. New York: Routledge Falmer.

Nokes-Malach, T. J., & Mestre, J. P. (2013). Toward a model of transfer as sense-making. *Educational Psychologist, 48*(3), 184-207. doi:10.1080/00461520.2013.807556

Novak, J. D., & Cañas, A. J. (2006). The origins of the concept mapping tool and the continuing evolution of the tool. *Information Visualization, 5*(3), 175-184. doi:10.1057/palgrave.ivs.9500126

Oh, P. S., & Oh, S. J. (2011). What teachers of science need to know about models: An overview. *International Journal of Science Education, 33*(8), 1109-1130. doi:10.1080/09500693.2010.502191

Osborne, J. F., & Patterson, A. (2011). Scientific argument and explanation: A necessary distinction? *Science Education, 95*(4), 627-638. doi:10.1002/sce.20438

Postareff, L., Parpala, A., & Lindblom-Ylänne, S. (2015). Factors contributing to changes in a deep approach to learning in different learning environments. *Learning Environments Research, 18*(3), 315-333. doi:10.1007/s10984-015-9186-1

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., . . . Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences, 13*(3), 337-386. doi:10.1207/s15327809jls1303_4

Ramsden, P. (1983). The Lancaster Approaches to Studying and Course Perceptions Questionnaire: Lecturer’s handbook. Educational Methods Unit, Oxford Polytechnic, Oxford, OX3 0BP.

Rea-Ramirez, M.A., Clement, J.J., & Núñez-Oviedo, M.C. (2008). An instructional model derived from model construction and criticism theory. In J.J. Clement & M. A. Rea-Ramirez (Eds.) *Model Based Learning and Instruction in Science* (vol. 2.) (pp. 23-43). Springer.

Reiser, B. J., Berland, L. K., & Kenyon, L. (2012). Engaging students in the scientific practices of explanation and argumentation. *Science and Children, 49*(8), 8-13.

Robertson, W. C. (1990). Detection of cognitive structure with protocol data: Predicting performance on physics transfer problems. *Cognitive Science, 14*(2), 253-280. doi:10.1016/0364-0213(90)90004-G

Ruiz-Primo, M. A., & Li, M. (2015). The relationship between item context characteristics and student performance: The case of the 2006 and 2009 PISA science items. *Teachers College Record, 117*(1), 1-36.

Russ, R. S., Coffey, J. E., Hammer, D., & Hutchison, P. (2009). Making classroom assessment more accountable to scientific reasoning: A case for attending to mechanistic thinking. *Science Education, 93*(5), 875-891. doi:10.1002/sce.20320

Russ, R. S., Lee, V. R., & Sherin, B. L. (2012). Framing in cognitive clinical interviews about intuitive science knowledge: Dynamic student understandings of the discourse interaction. *Science Education, 96*(4), 573-599. doi:10.1002/sce.21014

Russ, R. S., Scherr, R. E., Hammer, D., & Mikeska, J. (2008). Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education, 92*(3), 499-525. doi:10.1002/sce.20264

Schwarz, C. V., & White, B. Y. (2005). Metamodeling knowledge: Developing students' understanding of scientific modeling. *Cognition and Instruction, 23*(2), 165-205. doi:10.1207/s1532690xci2302_1

Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., . . . Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching, 46*(6), 632-654. doi:10.1002/tea.20311

Seel, N. M. (2001). Epistemology, situated cognition, and mental models: 'like a bridge over troubled water'. *Instructional Science, 29*(4/5), 403-427. doi:10.1023/A:1011952010705

Seel, N. M. (2003). Model-centered learning and instruction. *Technology, Instruction, Cognition and Learning, 1*(1), 59-85.

Shwartz, Y., Weizman, A., Fortus, D., Krajcik, J., & Reiser, B. (2008). The IQWST experience: Using coherence as a design principle for a middle school science curriculum. *The Elementary School Journal, 109*(2), 199-219. doi:10.1086/590526.

Sins, P. H. M., Savelsbergh, E. R., van Joolingen, W. R., & van Hout-Wolters, Bernadette H. A. M. (2009). The relation between students' epistemological understanding of computer models and their cognitive processing on a modelling task. *International Journal of Science Education, 31*(9), 1205-1229. doi:10.1080/09500690802192181

Sins, P. H. M., Savelsbergh, E. R., & van Joolingen, W. R. (2005). The difficult process of scientific modelling: An analysis of novices' reasoning during computer-based modelling. *International Journal of Science Education, 27*(14), 1695. doi:10.1080/09500690500206408

Snow, C. E. (2010). Academic language and the challenge of reading for learning about science. *Science, 328*(5977), 450-452. doi:10.1126/science.1182597

Son, L. K. (2013). *Metacognition*. New York: Oxford University Press.

Stratford, S. J., Krajcik, J., & Soloway, E. (1998). Secondary students' dynamic modeling processes: Analyzing, reasoning about, synthesizing, and testing models of stream ecosystems. *Journal of Science Education and Technology, 7*(3), 215-234. doi:10.1023/A:1021840407112

Svoboda, J., & Passmore, C. (2011). The Strategies of Modeling in Biology Education. *Science & Education*, 1-24. doi: 10.1007/s11191-011-9425-5

Tait, H., Entwistle, N. J., & McCune, V. (1998). ASSIST: A reconceptualisation of the approaches to studying inventory. *Improving Student Learning: Improving Students as Learners*, 262-271.

Tashakkori, A., & Teddlie, C. (2010). *Sage Handbook of Mixed Methods in Social & Behavioral Research* (2nd ed.). Los Angeles: SAGE Publications.

Treagust, D. F., & Harrison, A. G. (2000). In search of explanatory frameworks: An analysis of Richard Feynman's lecture 'atoms in motion'. *International Journal of Science Education, 22*(11), 1157-1170. doi:10.1080/09500690050166733

Trout, J. D. (2007). The psychology of scientific explanation. *Philosophy Compass, 2*(3), 564-591. doi:10.1111/j.1747-9991.2007.00081.x

van Mil, M. H. W., Boerwinkel, D. J., & Waarlo, A. J. (2013). Modelling molecular mechanisms: A framework of scientific reasoning to construct molecular-level explanations for cellular behaviour. *Science and Education, 22*(1), 93-118. doi:10.1007/s11191-011-9379-7

Van Someren, M.W., Barnard, Y. F., & Sandberg, J.A.C. (1994). The think aloud method: A practical guide to modelling cognitive processes: (Knowledge Based Systems Series). Academic Press, San Diego (1994). xii+208 pp., ISBN 0-12-714270-3

Veeman, M. V. J. (2012). Metacognition in science education: Definitions, constituents, and their intricate relation with cognition. In A. Zohar & Y. J. Dori (Eds.), *Metacognition in Science Education: Trends in Current Research* (pp. 21-36). New York: Springer.

Veenman, M. V. J., Hout-Wolters, v., B.H.A.M, & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning, 1*(1), 3-14. doi:10.1007/s11409-006-6893-0

Vermunt, J. D. H. M. (1994). *Inventory of Learning Styles in Higher Education: Scoring Key for the Inventory of Learning Styles in Higher Education*. Tilburg, Netherlands: Tilburg University Department of Educational Psychology.

Vermunt, J. D., & Vermetten, Y. J. (2004). Patterns in student learning: Relationships between learning strategies, conceptions of learning, and learning orientations. *Educational Psychology Review, 16*(4), 359-384. doi:10.1007/s10648-004-0005-y

Wang, T., Li, M., Thummaphan, P., & Ruiz-Primo, M. A. (2017). The Effect of sequential cues of item contexts in science assessment. *International Journal of Testing*, 1-29.

Warburton, K. (2003). Deep learning and education for sustainability. *International Journal of Sustainability in Higher Education, 4*(1), 44-56. doi:10.1108/14676370310455332

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*(1), 3-118. doi:10.1207/s1532690xci1601_2

Williams, J. J., & Lombrozo, T. (2010). The role of explanation in discovery and generalization: Evidence from category learning. *Cognitive Science, 34*(5), 776-806. doi:10.1111/j.1551-6709.2010.01113.x

Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations. *Science Education, 92*(5), 941-967. doi:10.1002/sce.20259

Wittrock, M. C. (1992). Generative learning processes of the brain. *Educational Psychologist, 27*(4), 531-541. doi:10.1207/s15326985ep2704_8

Wong, D. E. (1993). Understanding the generative capacity of analogies as a tool for explanation. *Journal of Research in Science Teaching, 30*(10), 1259-1272. doi:10.1002/tea.3660301008