Applying the theory of conceptual change to improve students’ understanding of science concepts with an educational recommender system

Ifeyinwa Okoye
University of Colorado
Boulder, Colorado, USA
okoye@colorado.edu

May 22, 2012
## Contents

1 Introduction ................................................. 4  
1.1 Motivation ............................................. 4  
1.2 Research Overview and Questions ....................... 4  
1.3 Anticipated Contributions ............................... 6  

2 Background and Related Work .......................... 8  
2.1 Conceptual Change as Learning Theory ................ 8  
2.2 Educational Recommender System as Computational Framework .......... 10  

3 Research Context and Educational Topic ............... 12  
3.1 CLICK version 1 ........................................ 12  
3.2 Seasons ................................................. 14  

4 My Conceptual Framework ................................ 16  

5 Research Design ............................................ 20  
5.1 Study 1 - Automatic Extraction of Core Learning Goals (Completed) .... 21  
5.2 Study 2 - Automatic Sequencing of Learning Goals (In progress) ......... 24  
5.3 Study 3 - Automatic Prioritization of student misconceptions (Proposed) ...... 29  
5.4 Study 4 - Design workshop for an educational recommender system that supports conceptual change (Proposed) .......................... 30  
5.5 Study 5 - Qualitative learning study to examine student processes and outcomes (Proposed) ...................................... 31  

6 Limitations, Risks and Mitigation ...................... 33  

7 Timeline ..................................................... 35  

A AAAS plate tectonics learning goals ..................... 36  

B Learning goal data 2 (LGD2) ............................. 37  

C Pedagogical sequencing of LGD2 by two subject experts ............... 40
List of Tables

5.1 Research questions and studies that will address each question ....................... 20
5.2 Coverage results for LGD1, LGD2 and LGD3 .................................................. 23
5.3 Example of a Precedence table ........................................................................ 24
5.4 Description of experts’ pedagogical sequence of LGD1 and 30 of the 32 concepts in LGD2 26
5.5 Accuracy results for Proxy Task ................................................................. 27
5.6 Accuracy results on pedagogical sequence ..................................................... 28
List of Figures

3.1 screen shot of CLICK .................................................. 13
3.2 Draft of AAAS idealized learning progression for seasons .......................... 15
4.1 My Conceptual Framework ............................................. 17
5.1 Distribution of coreness in LGD2(extraction) and LGD3(ranking) ............... 23
Chapter 1

Introduction

1.1 Motivation

Science misconceptions have proven to be persistent, deeply held and difficult to change (Vosniadou, 2008; Duit, 1999). Conceptual change learning theory applied in the classroom environment has proven effective in helping students remedy their misconceptions about science and in developing a more robust understanding (Vosniadou, 2008). The overarching goal of this research is to develop and study a new approach for creating educational recommender systems (ERS) based on conceptual change theory. Recommender systems have proven to be an effective technique for dealing with the information overload problem that confronts learners, as a result of the abundance of information available online (Adomavicius and Tuzhilin, 2005). An educational recommender system CLICK2 will be created as a result of this dissertation. The online system will be designed to improve student understanding of science concepts by recommending resources targeting their misconceptions.

This is a dual dissertation in computer science and cognitive science hence this work is expected to make contributions to both fields. The computational goal is to create a framework that extends educational recommender systems to facilitate science understanding. The learning science goal is to extend conceptual change learning theory into an online environment. In this research, I will draw subjects from students at CU-Boulder for my study population.

In the next section, I give an overview of the research questions and discuss the anticipated contributions of this dissertation in more detail. Chapter 2 examines related research in conceptual change theory and educational recommender systems. Chapter 3 explains the research context and educational topic that I will explore. Chapter 4 explains my conceptual framework; and chapter 5 goes into more detail about each of the five studies I am proposing to carry out as part of my dissertation. Chapter 6 reviews the limitations, risks and mitigations of the proposed research; and chapter 7 gives an overview of the time line for completing this dissertation.

1.2 Research Overview and Questions

The following scenario will motivate the need for this research. Mandy is a hardworking and conscientious female high school student in Boulder who is enrolled in an Earth science course. Mandy likes to be prepared for class and so, she usually reads a little bit about the topic for each upcoming class. The topic for her next Earth science class is seasons. Mandy assumes she is ready for this class because she has lived in Boulder all her life. She knows about the four seasons and when they occur.

The next day at school during the discussion on seasons, Mandy is surprised to realize that she was not ready for the class. She did not know that not everywhere on Earth had four distinct seasons like Boulder. Also, she did not know that winter in the southern hemisphere is harsher than winter...
in the northern hemisphere. At the end of the class session, seeing that the students did not fully understand the lesson, the teacher assigns the class an assignment, to research and write an essay report on seasons, specifically that answers the questions, *why do we have day and night?*, *what causes the seasons?* and *why is the winter in the northern hemisphere milder than the winter in the southern hemisphere?*.

Mandy goes home and starts trying to piece together all the information she needs to write a good essay. She decides to do more research about the seasons, so she goes online to a search engine and types in her search query *seasons on Earth* and gets back more than 94 million results. She starts browsing through the results, 3 hours later, she is still browsing and still does not have a clear idea of how the seasons come to be.

What Mandy needs is a system that can determine her current knowledge state and recommend resources to help her achieve the scientific understanding of the reason for the seasons. In order to create the automatic system that Mandy uses, I have identified five research questions that need to be addressed:

**(RQ1) How well can different computational methods identify the learning goals in a collection of documents?**

US schools have been criticized for not emphasizing the big ideas in science. Subsequently, there has been a push to target big ideas in depth rather than a long list of ideas in a superficial manner so as to promote robust understanding and produce citizens that are capable of reasoning correctly about scientific phenomena in the time constraints of the classroom (Vosniadou et al., 2001). Hence in order to guide a learner towards an efficient path of understanding, it is imperative that all learning systems work towards significant learning goals and not just assume that the student needs to see and learn every concept in a resource. So the question here is, given a collection of documents (chapters in a textbook or online resources), can we determine a good method for identifying the main ideas/learning goals in this collection?

**(RQ2) How well can machine learning classifiers model the pedagogical sequence of learning goals produced by human experts?**

Some concepts serve as building blocks for other concepts, and thus it is essential to learn the basic concepts before moving on to other concepts that depend on them i.e the order of acquisition of knowledge is important in promoting conceptual change (Vosniadou et al., 2001). Research has shown that instruction that follows particular learning progressions is important when trying to understand scientific topics (Plummer and Agan, 2010). For example, students must first understand the concept of the revolution of the Earth around the sun and the concept of the tilt of the Earth before they can understand the concept of seasons in the northern and southern hemisphere. There may exist several different but reasonable pedagogical sequences (also known as learning paths). We focus on generating a single sequence to provide students with a meaningful learning path through the resources. The sequence of learning goals that results from this module will serve as input for a later module that prioritizes a students misconceptions.

**(RQ3) How can we model expert strategies for prioritizing student misconceptions?**

As discussed earlier, it is essential that students address misconceptions associated with basic concepts before moving on to other dependent concepts. It is expected that correcting a basic misconception will induce the student to correct other misconceptions that were produced as a result of the basic misconception. However, the task here is not to merely follow the sequencing suggested by the pedagogical sequence because more than one misconception can be aligned to a specific learning
goal. Therefore we still need a strategy for prioritizing the misconceptions even after being given the pedagogical sequence of the learning goals. We chose to model the strategies used by expert Earth science teachers. When they pose a question in class and get several answers, they have to come up on the spot with how to structure their explanation to take care of the various misconceptions that the students have. And when they review a student’s work output such as an essay, they have to give constructive feedback that will help the student produce a better essay. But they do not just list all the problems the student has, they point out the major ones knowing that as those are fixed, the minor ones will be fixed too. Hence the goal of the resulting model is to prioritize the students misconceptions by ordering the misconceptions in a way that lets the learner learn more efficiently within a short period of time.

(RQ4) What are design options for creating an educational recommender system (ERS) with research-based support mechanisms for promoting conceptual change?

In the ERS literature, the design of the feedback environment has been routinely ignored. In a review of the research on ERS, 19 in 20 of the recommenders used a standard e-commerce interface. Only 1 in 20 explored interface design to support education (Manouselis et al., 2011) and some ERS researchers clearly state that it is not an important aspect of the research (Buder and Schwind, 2011). Since the goal of educational recommender systems is to facilitate learning and not just recommend sources of information, the recommendation interface should be designed to facilitate learning. The research question here concerns how we can facilitate students understanding of science by designing a recommendation interface that incorporates into its feedback environment, research-based support mechanisms that have proven effective for promoting conceptual change in classroom environments.

(RQ5) How does the educational recommender system with its conceptual change support mechanisms affect users’ understanding of science content?

The end goal of the system is to help students improve their understanding of science. A research study to examine the effect of the system use on learners’ processes and outcomes will be conducted. The study will also explore if and how the conceptual change support mechanisms were used.

1.3 Anticipated Contributions

This is an interdisciplinary research in computer science and cognitive science. This research draws on conceptual change theory to extend the state of art in educational recommender systems. In doing so, this research generates contributions to both computer science and learning science. Below, I discuss the anticipated contributions of this dissertation to both fields.

Computer Science Contribution

Using conceptual change theory to extend the state of art in educational recommender systems requires significant advances in algorithms that assess learners’ information need. Conceptual change theory operationalizes learners’ information need as prioritized misconceptions that depend on an instructionally sound pedagogical sequence of core learning goals. This work will advance the state of art of algorithms that assess learners’ information need by applying machine learning and natural language processing techniques to create new algorithms that automate the instructional process of identifying core learning goals, sequencing the learning goals and prioritizing misconceptions.

Extending the state of art in educational recommender systems also requires advances in how the interface is designed, in order for educational recommender systems to support conceptual change in
learners. This work will advance the state of art in educational recommender systems by designing a new interface that supports conceptual change in users of educational recommender systems.

**Learning Science Contribution**

The learning science contribution of this work is studying conceptual change learning theory in the context of an online educational recommender system. This work will also extend and refine conceptual change learning theory for use in an informal, online, learner-driven learning context. In addition, this work will expand the role of educational recommender systems in the learning process from simply making recommendations to also serving as a formative assessment tool.
Chapter 2

Background and Related Work

The two main research areas that this work builds on are conceptual change theory and educational recommender systems. Below, I discuss the current state of these two research areas.

2.1 Conceptual Change as Learning Theory

Conceptual change is the process through which people’s initial understandings or beliefs are altered to more closely align with scientifically-held understandings (Vosniadou, 2008). There are competing theories on the nature of these initial understanding or beliefs. But for purposes of this dissertation, I ascribe to the framework theory because it is the most accepted theory by conceptual change practitioners. This view of conceptual change sees students naive ideas about science concepts as existing within a cogent framework with a distinct ontology and which can give rise to predictions and explanations for phenomena that it encounters. This naive framework has usually not been acquired through hypothesis testing and there is usually no metaconceptual awareness within its’ owner about its existence. Learning then is seen as a re-organization of the knowledge within the framework and not only its enrichment (Vosniadou et al., 2008).

Conceptual change can happen naturally in the process of cognitive development and can also be induced instructionally, the concept change can range from the mundane to the more radical (Vosniadou et al., 2008; Inagaki and Hatano, 2008). Conceptual change is a latent variable that cannot be directly observed or measured but is presumed to exert influence on other observable variables such as learning or achievement. Hence, conceptual change has been operationalized as a transformation in learners knowledge, belief and interest (Pintrich et al., 1993; Plummer et al., 2011; Clement and Vosniadou, 2008; Vosniadou et al., 2001; Vosniadou, 2008).

Conceptual change research has its root in constructivism and grew out from the observation that learners have persistent alternate conceptions that are very resistant to change, even when these learners have been taught the correct conceptions in schools. Conceptual change research has focused on understanding how these alternate conceptions develop, why are they difficult to change and how people can be induced to change them. One of the first theoretical frameworks for conceptual change was put forth by Posner et al. (1982), where they identified four successive conditions that can induce conceptual change in learners. First, the learners have to be dissatisfied with their current understanding, then the new knowledge has to be intelligible (understandable), plausible(believable) and fruitful (produce correct explanations about related phenomena). This classical approach to conceptual change persisted for a long time but has been criticized for its over-emphasis on logical and rational thinking, its sole focus on the learner’s cognition and not on the learner as a whole and for consequently, ignoring the affective (motivation, values, interests) and other social components of learning (Pintrich et al., 1993; Duit and Treagust, 2003). Newer theoretical frameworks such as
the cognitive-affective model of conceptual change (Sinatra, 2005) have incorporated the affective and social components. While the affective and social components of conceptual change are important, I am limiting the scope of this research to the cognitive aspect.

The basic tenets for how to induce conceptual change in students, is that students have to be made to confront their misconceptions and then provided with support mechanisms to encourage restructuring of their naive framework and cognitive accommodation of the scientific conceptions (Vosniadou, 2008). The most effective method for eliciting student alternative conceptions has been by having the students produce an explanatory model (Cartier and Center, 2000; Vosniadou, 2008). An explanatory model is a description of how and why a phenomena is the way it is.

Such a model is seen as the means by which a theory takes on meaning and, if used flexibly, it gives the theory the power to explain and make predictions for new cases that the subject has not yet seen. Significant changes in an explanatory model are one of the most important types of conceptual change (Clement and Vosniadou, 2008).

Explanatory models can be naive (mainly incorrect), synthetic (a mix of incorrect and correct) and scientific (correct current scientific understanding). An explanatory model of a phenomena can range from simple (containing few entities), such as the model of the circulatory system that a middle school student can produce to very complex, such as the model of the circulatory system that a physician can produce. The importance of explanatory models in promoting understanding is widely recognized thus explanatory models are being used in several projects aimed at improving students understanding of science. For example, the Modeling for Understanding in Science Education (MUSE) project (Cartier and Center, 2000; Cartier et al., 2001; Passmore and Stewart, 2002; Stewart et al., 2005) at the University of Wisconsin-Madison, helps students improve their understanding of science by having the students construct an explanatory model for a scientific phenomena and then through different support mechanisms, such as teaching and investigative activities; help the students to gradually produce a more scientific explanatory model.

The most effective method for inducing conceptual change in students has been by using constructive and dissonance strategies to repeatedly criticize students explanatory models. Students then continually revise their models until they produce a more scientific model, with multiple short cycles needed for complex models (Clement and Vosniadou, 2008; Frede, 2008; Vosniadou, 2008). This method has been effective in a cooperative and facilitative environment where the teacher is very knowledgeable about the topic, co-constructs the knowledge with the students and allows for reflection (Bruning et al., 1999; Scott et al., 1991).

Constructive strategies such as analogy, leverages students current understanding and makes use of students prior knowledge in a positive manner. However, sometimes the analog is not well understood, students might transfer all the characteristics including dissimilar ones from the analog to the target, the target might be too far from the analog for the students to make the connection and analogies do nothing to counteract the alternative conception. Dissonance strategies include direct contrast, discrepant events, scientific model presentation, contrastive teaching and the use of refutation text. Dissonance strategies have been criticized for having the ability to negatively affect students confidence and self-esteem skills. In addition, they do not encourage new models to be built, they only knock down the old model and students could very well develop another alternate conception rather than the scientific one. Although constructive and dissonance strategies have their individual failings, when paired together in the right environment, they are the most effective means of producing instructionally induced conceptual change. A meta-analysis has shown that they produce an effect size of up to 1.4999 in the knowledge of users that were taught using these strategies (Murphy and Alexander, 2008).

Conceptual change research has for the most part, been confined to the sciences, been targeted at middle school students, high school students, college students and pre-service elementary school teachers and it has mostly been implemented in a classroom setting.
2.2 Educational Recommender System as Computational Framework

A Recommender system is any system that produces personalized recommendations as output or has the effect of guiding the user in a personalized way to interesting or useful objects in a large space of possible options (Burke, 2002). Recommender systems are applicable to a wide variety of domain and task.

There are three key recommendation techniques: (a) collaborative filtering: recommending items that match items a user has rated before or recommending item that similar users have rated before (b) content-based filtering: recommending items based on the kinds of items a user has purchased or viewed before, and (c) knowledge-based: recommending items that are based on a user model or profile of interests that may have been captured as a result of explicit feedback or built from behavioral or interaction data. Knowledge-based recommender systems are context-aware recommender systems because they take the user’s current knowledge context into consideration rather than simply relying on things the user has done in the past as in collaborative filtering and content-based filtering.

A problem with using collaborative filtering is that it requires the user to use the system for a while before it can start making relevant recommendations to the user. This is known as the new user problem. However most users will want to start getting benefits and recommendations from the system without having to rate a lot of resources first. In addition collaborative filtering can run into the sparsity problem where the ratings for individual resources are sparse. Furthermore, the people doing the rating might not have contemporary science understandings and so the resources being recommended as a result of their ratings might not have good educational value.

Because both collaborative filtering and content-based filtering rely on ratings that have been assigned by users of the system, if a new resource is introduced into the system, it will be difficult for the resource to get recommended since it has no ratings associated with it. This is known as the new item problem. Educational recommender systems will always have new users that want good recommendations right away and it is expected that new items will continually be added to the knowledge repository. Thus, good educational recommendations cannot rely only on collaborative filtering or content. They also have to take the context and knowledge of the user into consideration (Adomavicius and Tuzhilin, 2011). Hybrid Systems built by combining the three recommendation concepts have enjoyed interest early in the development of recommendation technologies. They continue to grow in popularity as more profile information is extracted from user behavior the longer they interact and consume items from the same systems.

While recommendation systems have enjoyed a great deal of success in e-commerce systems such as Amazon and Netflix, their benefits are being proven in other areas. Recommendation systems are being developed and deployed in a number of diverse areas - personalized learning and education being one such area. Personalized learning has been recognized as an important advancement for learning in the digital era. It follows that educational recommender systems can form the backbone of personalized learning engines, particularly those that are built on top of the web. These web-based learning engines can provide ubiquitous, instant and continuous access to online learning opportunities that are adapted and customized to each learner’s individual needs.

There are several educational recommender systems available right now. Altered Vista uses collaborative filtering to recommend learning resources which have been rated highly by users (Recker and Walker, 2003). QSIA (Rafaeli et al., 2004) is a user-controlled collaborative filtering recommender system. The user can pick the people for example, friends or teaching assistants, whose profile should be used in collaborative filtering to recommend learning resources to him. Or he can decide to let the system decide which group of users to use in the collaborative filtering. Both of these systems rely only on the ratings that users have assigned to a resource. This can lead to the new item problem,
where a new item that is relevant does not get recommended because it does not have enough ratings. These systems also do not take the user’s profile into consideration. The resource being recommended might be too difficult for the user to comprehend (either maybe because of the wording or because the user needs more background information before tackling the information in the resource).

Shen and Shen (2004) propose a system that uses sequencing rules and an ontology of a domain to guide users through the domain. When non-scientific understandings are identified in the learner’s knowledge state, the rules are used to decide which resources to recommend to the learner. This system relies on a hand-crafted knowledge base consisting of rules and ontologies specific to a topic within a domain. Such a system will be very difficult to generalize to other topics and domains. CourseRank (Koutrika et al., 2009) is a hybrid system that recommends classes to take, using collaborative filtering based on a user’s profile such as knowledge state (prerequisites taken) and other attributes such as major and area of interest. Huang et al. (2009) use a Markov chain model to calculate transition probabilities between learning objects in a sequenced course of study. These two systems target sequencing at a higher level than I plan to do. Instead of sequencing the classes or the learning resources within a class, I will sequence concepts in order to create a personalized learning path through the content. I am targeting the learner’s misconceptions at a much lower and personal level.

The system I propose to build is similar to ISIS (Hummel et al., 2007), a hybrid ERS which uses ratings of other users and metadata from the learner’s profile and learning activity to recommend learning objects. However, I target the user’s specific information based on their misconceptions instead of simply recommending resources based on the user’s general learning activity. Another very similar system is a hybrid approach that was implemented in the Virtual University of Tunis (Khribi et al., 2008). It combines collaborative filtering with content-based filtering and also uses the knowledge of the user which it logs and mines from the user’s actions. Similar to ISIS, this system approximates the user’s specific information need by features which do not get to the user’s specific information need. For example, one of its features is to assume a user does not understand the content contained in a resource if the user spends a lot of time on the resource. But it is difficult to determine when a user is actually on the resource, or when the user has left it and just forgot to close the site.

Aside from the already discussed shortcomings of existing systems, a 2011 review of twenty reported educational recommender systems show that very few ERS have been evaluated on their impact on learners process and outcomes (Manouselis et al., 2011). In addition there is no readily available data set to empirically validate the recommendation algorithms. Thus it is difficult to compare and decide which ERS is state of art.

In summary most of the educational recommender systems have so far tried to use techniques and the types of user information found in e-commerce recommender systems to design and create educational recommender systems. It is generally accepted that the goals of educational recommender systems are different from commercial recommender systems. While the objective of e-commerce recommender systems is to provide customers with information to help them decide which products to purchase, the objective of educational recommender systems is to find good items that will address users’ knowledge needs. Because, the goal of educational recommender systems is very different from e-commerce recommender systems, educational recommender systems should be designed and evaluated in a different way than commercial recommender systems (Buder and Schwind, 2011).

This is what I plan to do for my dissertation. I will design an educational recommender system using the conceptual change learning theory in order to facilitate student understanding of science concepts. The research context for my work will be an educational recommender system called CLICK (de la Chica et al., 2008) that I helped to build and which is discussed in the next chapter.
Chapter 3

Research Context and Educational Topic

The research context for my dissertation will be an educational recommender system, the Customized Learning Service for Concept Knowledge (CLICK). CLICK has been under development for the past 7 years. In that time, it has become a robust computational infrastructure which can serve as a test bed for studying personalized educational recommendations because the core recommender system algorithms have been created. I will build on this existing educational recommender system so I can focus on algorithms to study conceptual change.

The educational topic I will explore during the course of my research is understanding the Earth system processes that give rise to the seasons. This topic has been well studied and research has shown that learners of all ages have persistent misconceptions about seasons (Trumper, 2000, 2001a,b,c; Atwood and Atwood, 1996; Sebastià and Torregrosa, 2005). In the classic video recorded in 1989 titled *A private universe* (Schneps et al., 1989), graduating Harvard students were asked to explain why we have seasons; many of their answers are classic misconceptions about seasons. Since this topic is relevant and applicable to everyone, I decided it was a good topic to explore during my dissertation.

3.1 CLICK version 1

The overarching goal of CLICK (de la Chica et al., 2008) was to create a scalable online service that recommends resources to users based on their conceptual understanding. Currently, CLICK uses the Digital Library for Earth Systems Education (DLESE) to support learners’ understanding of Earth science content. CLICK automatically constructs a domain knowledge base from digital library resources and evaluates users’ conceptual understandings against the domain knowledge through automatic essay analysis. CLICK detects flaws and gaps in users’ science knowledge of Earth system concepts and recommends digital library resources to address users’ misunderstandings and knowledge gaps. Users are encouraged to visit those resources and upon sufficient review, revise and re-write their essays for re-evaluation.

Prior work on CLICK investigated four components of a conceptual educational recommender system; the domain knowledge generator, misconception identifier, resource recommender and a preliminary recommender interface.

The domain knowledge generator is COGENT, a multi-document summarizer, optimized for the Earth science domain. Multi-document summarization is a computational technique for analyzing multiple documents and generating a summary of the information contained in the documents. COGENT extends a generalized multi-document summarizer, MEAD (Radev et al., 2004), by adding features such as educational standards, hypertext and content word density to determine which con-
cepts to extract from a collection of resources for use in building a knowledge base (de la Chica, 2009). Although COGENT is capable of identifying all the concepts needed for robust understanding of plate tectonics at the middle and high school levels (de la Chica, 2009), it does not identify the most important learning goals, the concepts that represent big ideas in science.

The initial misconception identifier was produced by graph comparison of a concept map of a student’s essay with a domain knowledge concept map. A concept map is a map where the nodes represent concepts and the links show the connections between concepts. The misconception identifier diagnosed three types of misconceptions - incorrect, incomplete (missing) and fragmented (Ahmad, 2009).

The links used in this initial misconception identifier were generated manually i.e this algorithm relied on links generated by human experts. Since one goal of CLICK is to be a fully automatic system, this original misconception identifier has been significantly modified over the past 2 years. The current version which I will use in this research uses entailment to determine misconceptions in students’ work. It does not make use of concept maps. Entailment is a text analytic technique for determining if the information in a text, \( H \), can be inferred from another text, \( T \). In this research, I will extend the misconception identifier to automatically prioritize the identified misconceptions.

Conceptual change theory highlights the need to focus on the most important concepts and the most important misconceptions first before moving on to other information about a topic.

The resource recommender uses graph analytic techniques to compare the resources and student misconceptions in order to recommend appropriate resources (Gu, 2009). The current recommender system relies solely on knowledge-based recommendations i.e it only uses information on students’ current misconception in its’ recommendation. In this research, I will extend this recommender system to incorporate user contributed information such as feedback from the students about resources they found most useful in their learning process.

Figure 3.1: screen shot of CLICK

The fourth component investigated in CLICK was the preliminary recommender interface shown in Figure 3.1. The text editor where students write their essay, is on the left. The feedback panel, where the recommendations are shown to students, is on the right. For each misconception a student has, the system displays the misconception, a manually constructed cognitive prompt that encourages the student to review the sentence and three recommended resources. For each recommended resource, the...
CHAPTER 3. RESEARCH CONTEXT AND EDUCATIONAL TOPIC

title, url and a description of the resource is also displayed. Although this educational recommender interface supports learning, it was not designed to facilitate conceptual change and has no mechanism for users to give feedback on the recommended resources. Although the current components work well on their own, they were not coupled together automatically and have room for improvement. My work will focus on creating this missing link and improving several components.

3.2 Seasons

The educational topic I will explore during my research is seasons. I chose this topic because I am interested in understanding why and how the four distinct seasons we experience in Boulder occur. Understanding seasons was not a topic of interest for me in Nigeria because we did not have really different seasons. But after living in Boulder for four years and experiencing the different seasons, I am very interested in exploring this topic.

Core high school Earth science concepts all students should know, include a robust understanding of seasons. However, it is difficult for many learners to attain a robust understanding of what causes the seasons, even though it is a phenomena we all experience (Trumper, 2000, 2001a,b,c; Atwood and Atwood, 1996; Sebastià and Torregrosa, 2005).

A survey of different research papers on this topic produced common misconceptions about seasons, some of which are enumerated below:

- Seasons are caused by the distance of the Earth from the sun. So in the summer the Earth is closer to the sun.

- The Earth’s orbit around the sun is a highly elongated (skinny) ellipse, making the distance between Earth and sun vary dramatically over the course of a year.

- The sun is pretty far off-center within the Earth’s orbit, making the distance between Earth and sun vary with time of year even more.

- The distance of different parts of the Earth from the sun, caused by the tilt of the Earth on its axis, causes the seasons

- Seasonal characteristics are the same everywhere on Earth.

- Seasons happen at the same time everywhere on Earth.

- The average temperature in the winter months in the Northern and Southern hemisphere are the same

Figure 3.2 shows the AAAS draft of an idealized learning progression for contemporary ideas about seasons (Willard et al., 2007). I hope that users of the system I will create will be able to address and replace their misconceptions with the conventional scientific understandings that are shown in Figure 3.2.
The seasonal variations in temperatures at different places on the surface of the Earth are explained by the differential heating of the Earth's surface as it rotates on an axis that is tilted relative to the plane of the Earth's orbit around the sun.

The intensity of sunlight striking a place on the surface of the Earth depends upon where the Earth is in its yearly orbit around the sun and how far the place is from the equator.

Because the Earth is a sphere, at any particular time, light from the sun strikes different parts of the Earth at different angles and therefore the intensity of light striking the surface of the Earth is different in different places.

The difference in how much of the day is daytime and how much is nighttime at a place on the surface of the Earth depends upon where the Earth is in its yearly orbit around the sun and how far the place is from the equator.

The axis of the Earth's rotation is tilted relative to the plane of the Earth's yearly orbit around the sun. As the Earth orbits the sun, the axis remains pointed to the same place in space.

The intensity of sunlight striking a place on the surface of the Earth varies depending on what time of day it is, what time of year it is, and on how far north or south of the equator the place is.

The temperature of a location on the Earth's surface depends upon the number of hours of sunlight and the intensity of that sunlight.

The rotation of the earth on its axis every 24 hours produces the night-and-day cycle. This turning of the planet makes it seem as though the sun, moon, and stars are orbiting around the earth once a day.

The earth is one of several planets that orbit the sun, and the moon orbits around the earth.

The earth is approximately spherical in shape. Like the earth, the sun and planets are spheres.

The temperature and amount of rain (or snow) tend to be high, low, or medium in the same months every year.

The seasonal variations in temperature patterns in variations of temperature

The intensity of sunlight striking a place on the surface of the Earth

The temperature of a location on the Earth's surface

The intensity of sunlight striking a place on the surface of the Earth

The temperature of a location on the Earth's surface

The number of hours of daylight or nighttime a location on the Earth's surface gets varies in a predictable pattern over the course of the year that depends upon how far north or south of the equator they are.

A number of planets of very different size, composition and surface features move around the sun in nearly circular orbits.

The rotation of the earth on its axis every 24 hours produces the night-and-day cycle. This turning of the planet makes it seem as though the sun, moon, and stars are orbiting around the earth once a day.

The earth is one of several planets that orbit the sun, and the moon orbits around the earth.

The earth is approximately spherical in shape. Like the earth, the sun and planets are spheres.

The temperature and amount of rain (or snow) tend to be high, low, or medium in the same months every year.

The sun warms the land, air, and water.

Light and other electromagnetic waves can warm objects. How much an object's temperature increases depends on how intense the light striking its surface is, how long it shines on the object, and how much of the light is absorbed.

Patterns in light warming objects

Patterns in variations of temperature

Patterns in the motions of the Earth
Chapter 4

My Conceptual Framework

The ultimate goal of the system that I will build is to improve learners’ understanding of science content using conceptual change learning theory. I will investigate how educational recommender systems can achieve this using the diagram shown in Figure 4.1. I will build on initial algorithms in CLICK and create a new system, called CLICK2. The shaded boxes reflect the new components that I will add to the existing CLICK system. CLICK2 will utilize COGENT, the updated misconception identifier and the resource recommender. However it will differ from CLICK in a number of ways. CLICK identified all the learning goals in a collection of resources while CLICK2 will investigate how algorithms for identifying learning goals can be optimized to identify core learning goals. Thus, CLICK2 will include a core learning goal (big science idea) identifier. It is necessary to identify the core learning goals in a collection of resources because conceptual change learning theory has highlighted the importance of focusing learners on core ideas rather than a plethora of all the ideas about a topic, in order to help learners develop more robust understanding.

Conceptual change learning theory has also stressed the importance of targeting student misconceptions in an order such that students deal with basic misconceptions before dependent misconceptions. Currently, CLICK generates a list of students’ misconceptions that are not prioritized. CLICK2 will study and create algorithms that can generate an instructionally sound sequence of core learning goals and algorithms that can prioritize students’ misconceptions. The pedagogical sequence generator is important to determine an ideal order in which the learning goals should be learned. The misconception prioritization module, which will depend on the pedagogical sequence of core learning goals, will ensure that learners have the correct conceptions about basic concepts before moving on to learn dependent concepts. Although CLICK was designed to target students’ misconceptions, it was not designed to help students deal with their persistent misconceptions. CLICK2 will be designed to include research-based support mechanisms that promote conceptual change. The instructional response that CLICK2 will generate will incorporate dissonance and analogy strategies from conceptual change theory. The CLICK2 interface will also be capable of receiving feedback from learners about the recommended resources.

My conceptual framework expands the role of educational recommender systems in the learning process from simply making recommendations to also serving as a formative assessment tool. It will serve as a formative assessment tool by providing learners with very targeted feedback on their current work i.e., their essays. Feedback is a very important aspect of effective formative assessments. Several meta analysis of its’ use in classrooms has shown it can induce an average effect size of 0.79 (Hattie and Timperley, 2007). And as can be seen in Figure 4.1, I will be giving feedback about their work output (essays) to the users.

Feedback is information provided by an agent regarding aspects of one’s performance or understanding (Hattie and Timperley, 2007).
According to the effective feedback model proposed by Hattie and Timperley (2007), feedback has to answer three questions; *where am I going?*, *How am I going?* and *Where to next?*, in order for it to be effective in reducing the gap between what the learner understands and what the learner needs to understand. The answers to these questions are built into my conceptual framework. Identifying the learning goals answers *where am I going?*, Sequencing the learning goals answers *how am I going?* and prioritizing the misconceptions answers *where to next?*. The feedback interface will ensure that the users can answer these questions for themselves as they use CLICK2.

The basic tenets in conceptual change research are (1) make learners aware of their misconceptions and (2) support learners to attain a more scientific understanding. Following these principles, in my conceptual framework, learners will be confronted with their misconceptions. Then I will use dissonance and analogy strategies from conceptual change learning theory to support the learners in addressing their misconceptions.

![Diagram of My Conceptual Framework](image-url)

Figure 4.1: My Conceptual Framework
How can CLICK2 help Mandy

The following scenario, which is a continuation of the scenario from chapter 1, and shows how CLICK2 can improve learners’ understanding of science content.

After unsuccessfully browsing through many websites, Mandy approaches her teacher in class the next day, explaining to her that she can tell her word for word what various online resources say about the reason for the seasons but she does not truly understand how it all works and hence is having difficulty writing an essay in her own words. Her teacher gives her the website of the CLICK2 recommender system and tells her that this will help her improve her understanding of seasons and enable her to write an essay that reflects the contemporary understanding of seasons. Mandy is doubtful about this because she has been to many websites and they have not been able to help her but she is resolved to try this website.

When she gets back home, Mandy goes to the CLICK2 website, where she is prompted to enter her search query, grade level and the number of learning goals about seasons that she would like to see. She types in why do we have night and day, what causes the seasons, why is winter in the northern hemisphere milder than winter in the southern hemisphere as the query, selects high school as the grade level, and requests to see six learning goals. The system searches DLESE using her query and retrieves 100 appropriate resources internally. Then it extracts six learning goals from the 100 resources and creates a learning path through the goals. Next the system prompts Mandy to write down all she knows about seasons in the text editor provided.

Mandy writes We have night and day because the Earth rotates on its axis every 24 hours and so the places facing the sun will have daylight and the ones facing away from the sun will have night. The reason why we have seasons is that the Earth is at different distances from the sun at different times of the year. When it is closer to the sun, we have summer and when it is farthest away from the sun, we have winter. The system processes her answers and annotates her answer for night and day as being correct and the answer for the cause of seasons as a misconception. It also infers that she needs to learn about factors affecting the temperature of any location on the surface of the Earth. So to the system, she has two problematic conceptions (an incorrect and a missing conception). The system indicates these misconceptions in the interface and Mandy is surprised that her understanding of why the seasons occur is wrong.

For her missing conception about winter in the northern and southern hemisphere, CLICK2 recommends 3 resources that discuss factors affecting the temperature of any place on Earth and characteristics of the northern and souther hemisphere. For her incorrect conception about what causes the seasons, CLICK2 recommends paragraphs from three resources that discuss what causes the seasons. These paragraphs were selected by the system to challenge Mandy’s explanation, thereby creating cognitive dissonance within Mandy’s mental model, which hopefully will lead her to generate a more scientific explanation. In order to help her address both her misconceptions, the system also displays an interactive correct simulated model of the Earth as it rotates on its axis and revolves around the sun. The tilt of the Earth is emphasized and the rays of the sun hitting any point on the Earth are illustrated. In addition, the seasons of the northern and southern hemisphere at each point in time is annotated on the model.

Mandy plays with the simulation, reads through the recommended resources and paragraphs and reflects on what she has observed and read. She uses this information to rewrite her essay and goes through several iterations of rewrites and recommendations before the system assures her that her essay now reflects the scientific understanding of seasons. The next week in class, after submitting their essays, the teacher asks what they learned from writing the essay. Mandy is able to say that prior to writing the essay, she thought that seasons were caused by the distance of the Earth from the sun. But from the research she did to write the essay, she learned that the seasons are caused by the tilt of the Earth as it revolves around the sun. The part of the Earth tilted towards the sun receives
more direct sunlight, hence it is warmer and experiences summer. The part of the Earth pointing away from the sun receives less, indirect sunlight, thus is colder and experiences winter.
Chapter 5

Research Design

My research design draws on methodologies from computer-human interaction, machine learning and natural language processing. The proposed research design is comprised of five studies. The five studies correspond to the five research questions discussed in Chapter 1. Table 5.1 shows the mapping between the research questions and the studies.

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>(RQ1) How well can different computational methods identify the learning goals in a collection of documents?</td>
<td>Study 1 - Automatic Extraction of Core Learning Goals ( Completed)</td>
</tr>
<tr>
<td>(RQ2) How well can machine learning classifiers model the pedagogical sequence of learning goals produced by human experts?</td>
<td>Study 2 - Automatic Sequencing of Learning Goals ( In progress)</td>
</tr>
<tr>
<td>(RQ3) How can we model expert strategies for prioritizing student misconceptions?</td>
<td>Study 3 - Automatic prioritization of student misconceptions (Proposed)</td>
</tr>
<tr>
<td>(RQ4) What are design options for creating an educational recommender system (ERS) with research-based support mechanisms for promoting conceptual change?</td>
<td>Study 4 - Design workshop for an educational recommender system that supports conceptual change (Proposed)</td>
</tr>
<tr>
<td>(RQ5) How does the educational recommender system, with its conceptual change support mechanisms, affect users’ understanding of science content?</td>
<td>Study 5 - Qualitative learning study to examine student processes and outcomes (Proposed)</td>
</tr>
</tbody>
</table>

The first three studies are concerned with comparing machine learning and natural language processing approaches to performing three critical tasks underpinning support for conceptual change theory: extracting learning goals, sequencing learning goals and prioritizing learners’ misconceptions. All three of these studies draw on analyses of human expert processes to inform the design and evaluation of the algorithms. In the fourth study, I will use participatory design methods to create a recommendation feedback interface. The fifth and final study will be a learning study that investigates how the CLICK2 system influences learners’ processes and outcomes.

Study 1 has already been completed and the results are described in this chapter. Study 2 is currently in progress; studies 3, 4 and 5 will be conducted as part of the research proposed here.
5.1 Study 1 - Automatic Extraction of Core Learning Goals (Completed)

Purpose
The purpose of this study was to compare two different multi-document summarization approaches for identifying core learning goals in a collection of documents. The outcomes of this study will be used to improve the core learning goal identifier algorithm in the CLICK2 system. The two methods to be compared are ranking of learning goals and reduction in the number of extracted learning goals. These two methods were selected because they are important variables to be considered when optimizing multi-document summarization techniques for different domains. In this experiment, the scientific topic was plate tectonics.

Materials
The first set of materials was twenty resources related to plate tectonics selected from DLESE by subject experts. The second set of materials is a set of extracted domain concepts, which were extracted from the twenty resources using the existing COGENT system with an extraction rate of 5%. When extracting concepts, COGENT is identifying the most promising sentences to represent the learning goals of the system. This extraction yielded a set of 97 concepts which I call the extracted domain concepts. This study will use these extracted domain concepts as the baseline learning goals from which to examine how well different methods for extracting core learning goals work.

Methodology
My methodology for this study was to compare and contrast the performance of two algorithms in identifying core learning goals. The measures I used to assess core learning goals was coverage and coreness. My measure of coverage is based on how well the resulting set of identified core learning goals corresponds to the AAAS benchmarks for plate tectonics. My measure for coreness is how well the learning goals identified by the two algorithms correspond to coreness rating of human subject matter experts. Thus, I used the standard machine learning technique of comparing the output of the algorithms to a gold standard set generated by human experts.

Producing the human expert evaluation set
I worked with two Earth science subject experts to create the evaluation set. The two Earth science subject experts were asked to annotate the 97 concepts in the extracted domain concepts on two dimensions: alignment and coreness to AAAS benchmark learning goals for plate tectonics. Alignment refers to similarity to the 12 AAAS benchmark learning goals for plate tectonics, shown in Appendix A. The experts assigned each concept in the extracted domain concepts to a benchmark learning goal, to which the concept was most related. Coreness in this context is defined as the centrality (degree of alignment) to the AAAS benchmark learning goals for plate tectonics. The experts assigned a coreness rating of 1 to 4 to each concept, with 4 being the most core. This resulted in an alignment and coreness rating for each concept in the extracted domain concepts.

A gold standard set of core learning goals was created by putting all the concepts in the extracted domain concepts which had a rating of 4 into a set. This yielded learning goal data 1 (LGD1), a set of 29 core learning goals which at least one expert had rated as 4.
CHAPTER 5. RESEARCH DESIGN

Algorithm 1 : Reducing COGENT extraction rate

We used COGENT to extract core learning goals from the twenty DLESE resources. We chose to do
the extraction at 1% in order to produce a similar number of concepts to those in the gold standard.
This resulted in 32 concepts for learning goal data 2 (LGD2), shown in Appendix B. The decision
to reduce the extraction rate in COGENT to identify core learning goals came from a study which
showed that as the extraction rate for plate tectonics was decreased from 5% to 1% of words, the
average coreness of the extracted concepts increased steadily (Foster et al., 2012).

Algorithm 2 : Ranking in COGENT

We use ranking to extract core learning goals from the set of domain concepts. Ranking is a technique
used in information retrieval for identifying the most relevant resources. COGENT has a in-built
ranker, which it uses during the last stage of the multi-document summarization process, to decide
which concepts in a collection of documents to extract to create a summary. We use the COGENT
rankings to generate learning goal data 3 (LGD3). The top ranked 29 concepts in the set of domain
concepts generated by COGENT became learning goal data 3 (LGD3).

Results

We evaluate all automatically identified learning goal sets; i.e., LGD2 and LGD3 for coverage of and
coreness to the AAAS benchmark learning goals for plate tectonics. We used the human subject
expert annotations of alignment and coreness to score the output of the extraction and the ranking
algorithms. Table 5.2 shows the results for coverage for the automatically identified learning goals
sets, the core learning goals produced by reducing the extraction rate from 5% to 1% (LGD2) and the
core learning goals produced by ranking (LGD3). It also shows the coverage of LGD1 for the sake of
comparison. The first column in Table 5.2 refers to the 12 AAAS benchmark learning goals for plate
tectonics, 6 for middle school and 6 for high school. These benchmarks are available in Appendix A.
In Table 5.2, PT-BMK, refers to plate tectonics-benchmark. MS1 refers to middle school 1 and HS1
refers to middle school 1. Thus PT-BMK-HS1, refers to the first high school plate tectonics benchmark
learning goal. As Table 5.2 shows, LGD1, the gold standard set, had the best coverage. In the two
learning goal sets extracted by the algorithms, LGD2 covered more of the benchmark learning goals
than LGD3.

Because the subject expert annotations for alignment and coreness was done on the 97 domain
concept data and not on all the concepts in the 20 resources, we had 4 concepts in LGD2 that were not
part of the 97 domain concepts. We asked another domain expert to annotate those 4 concepts using
the same rubric. We know that the learning goals in the gold standard set, LGD1 are core learning
goals because they all have a coreness value of 4 (the highest coreness rating). LGD2 had an average
coreness of 3.25 while LGD3 had an average coreness of 2.3. As Figure 5.1 shows, LGD2 identified
more core learning goals than LGD3.

Findings and Discussion

Reducing the extraction rate significantly outperforms the rankings in COGENT. Table 5.2 shows that
LGD2 covered more of the AAAS benchmark learning goals for plate tectonics than LGD3. Figure
5.1 shows that LGD2 identified more core learning goals than LGD3. This demonstrates that it is
better to use an algorithm that reduces the extraction rate to generate core learning goals than using
the ranking in COGENT to generate core learning goals. By using an algorithm that reduces the
extraction rate, we can identify concepts with a higher coreness rating and better coverage and thus
Table 5.2: Coverage results for LGD1, LGD2 and LGD3

<table>
<thead>
<tr>
<th>#</th>
<th>AAAS plate tectonics learning goal</th>
<th>LGD1 human annotation</th>
<th>LGD2 extraction algorithm</th>
<th>LGD3 ranking algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>PT-BMK-MS1</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>(2)</td>
<td>PT-BMK-MS2</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>(3)</td>
<td>PT-BMK-MS3</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>(5)</td>
<td>PT-BMK-MS5</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>(6)</td>
<td>PT-BMK-MS6</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>(7)</td>
<td>PT-BMK-HS1</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>(8)</td>
<td>PT-BMK-HS2</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>(9)</td>
<td>PT-BMK-HS3</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>(10)</td>
<td>PT-BMK-HS4</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>(11)</td>
<td>PT-BMK-HS5</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>(12)</td>
<td>PT-BMK-HS6</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>83% (10/12)</td>
<td>75%(9/12)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>58%(7/12)</td>
</tr>
</tbody>
</table>

Figure 5.1: Distribution of coreness in LGD2(extraction) and LGD3(ranking)
can identify core learning goals in a collection of resources. Therefore, I will adjust the extraction rate in COGENT to 1% to identify the core learning goals in a collection of resources.

5.2 Study 2 - Automatic Sequencing of Learning Goals (In progress)

Purpose

The purpose of this study is to compare and contrast four different machine learning models for generating a pedagogical sequence. The four different machine learning classifiers I am comparing are naive Bayes, Maxent and two support vector machine algorithms, LibSVM and SMO. To break this down into a tractable task, I cast this as a pair-wise ordering problem; i.e, rather than focusing on trying to automatically generate entire pedagogical sequences, I am focusing on developing an algorithm capable of identifying when one learning goals precedes another.

The output of the trained models will be evaluated on pair-wise orderings of learning goals. To generate a pedagogical sequence from the resulting pair-wise judgments, first, we construct a precedence table from the pair-wise judgments and then generate a learning path from the precedence table. Table 5.3 is an example of a precedence table that contains four concepts. Three learning paths that can be generated from this table are: (1) $C \rightarrow A \rightarrow D \rightarrow B$, (2) $C \rightarrow D \rightarrow A \rightarrow B$ and (3) $C \rightarrow A \rightarrow B \rightarrow D$.

Table 5.3: Example of a Precedence table

<table>
<thead>
<tr>
<th>Concept</th>
<th>Preceded by</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>B</td>
<td>A C</td>
</tr>
<tr>
<td>C</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>C</td>
</tr>
</tbody>
</table>

We use two different pair-wise ordering tasks to compare and contrast the four machine learning models. One pair-wise ordering task uses the experts’ pedagogical sequences of the learning goals. However because that data set is so small - only 9 sequences - we needed to create a larger data set to help train these models. Therefore we created a proxy set using the pair-wise ordering of middle and high school sentences on plate tectonics.

Materials

This study uses three sets of materials. The first set of materials is the core learning goals identified by subject experts $LGD1$ and the core learning goals identified using the 1% extraction algorithm ($LGD2$) in Study 1. We chose to use LGD1 because it represents the best set of core learning goals. We chose to use LGD2 because it is the best set of core learning goals that we can create algorithmically.

The second set of materials are the pedagogical sequences produced by experts using the core learning goals. We collected 9 pedagogical sequences from 2 experts. As described later, each of these pedagogical sequences was then converted into pair-wise judgments between learning goals. These pair-wise judgments are used for both training and evaluating the machine learning models.

A third set of materials was created to help study pair-wise ordering. This third set of materials is a pair-wise ordering of middle and high school sentences. To construct this data set, we searched the DLESE website for text resources that contained the words earthquake or plate tectonics. We collected 10 such resources for each of the two grade cohorts: middle school (we allowed anything K-8) and high school (we allowed anything 9+). We downloaded the webpage for each resource, and used COGENT
to extract the 20 most important sentences from each. This resulted in 200 sentences for each of the
two grade cohorts. We divided the sentences in each grade cohort into three sets, one for training,
one for development and the third for testing. To create pairs of grade-ordered sentences, we paired
up middle and high school concepts both ways: middle school first (i.e. \( \text{sequence}(c_m, c_h) = 0 \)) and
high school first (i.e. \( \text{sequence}(c_h, c_m) = 1 \)). This resulted in 4356 grade-ordered sentence pairs for
each of the three sets (training, development and testing). These pair-wise orderings are only used for
training the models.

**Methodology**

We used two different approaches to train the four machine learning classifier models, naive Bayes,
Maxent and 2 support vector machine models, SMO and LibSVM in order to compare and contrast
their performance on experts’ pedagogical sequences. As discussed earlier, because we had a small
data set for the experts’ pedagogical sequences, we could not create dedicated training and testing sets
from them. Thus, we trained the models using two different data sets described under materials. We
trained the models using a 6-fold cross validation of the experts’ pedagogical sequence. In addition,
we also explored how the models perform when we train them on a larger proxy data set. Training
the 4 classifiers on two different data sets will produce 8 models. Each of the 8 models will then be
evaluated on the test set - expert’ pedagogical sequences of the learning goals. WEKA (Hall et al.,
2009) is a suite of machine learning algorithms implemented in JAVA and open sourced under GPL.
We used WEKA’s implementations of the four machine learning classifier algorithms, naive Bayes,
Maxent, LibSVM and SMO. Using WEKA enables us to concentrate on training the models for our
task rather than rewriting the algorithms.

**Producing the human expert evaluation set**

The human expert evaluation set for this study is the second set of materials, the pair-wise orderings
generated from experts’ sequencing of LGD1 and LGD2. We asked two subject experts to come up
with ideal learning paths i.e sequences for LGD1. Because the task of coming up with a learning path
for LGD1 was unconstrained, the first expert generated three different pedagogical sequences while
the second expert generated only one. Following the outcome of the unconstrained sequencing task
on LGD1, we asked the same two subject experts to come up with two learning paths each for LGD2;
but this time, we constrained the task. We requested that the first sequence follow an *evidence or research based* learning path while the second sequence follow a *traditional* learning path.

An *evidence or research based* learning path is a pedagogy where students are encouraged to
use the scientific method to learn about a phenomena, i.e they gather information by observing the
phenomena, forming a hypothesis, performing experiment, collecting and analyzing data and then
interpreting the data and drawing conclusions that hopefully align with the current understanding
about the phenomena. A teacher that uses this learning path acts as a *guide on the side*. A *traditional*
learning path, on the other hand, is the pedagogy where teachers are simply trying to pass on the
correct information to students rather than letting the students discover the information themselves.
In a classroom environment, a teacher using this learning path would be seen as the classical *sage on stage*. Both experts agreed that 2 of the 32 learning goals in LGD2 were not concepts but rather a
heading and a web trail, therefore they excluded them when generating pedagogical sequences from
LGD2. The first expert came up with one learning path for evidence based and two learning paths
for traditional while the second expert came up with one each, so we had five learning paths for 30 of
the 32 concepts in LGD2. The learning paths are shown in Appendix C.

The subject experts produced a partial ordering of the learning goals. So although there are
30 concepts to be sequenced in LGD2, both experts produced only 21 and 26 levels of ordering
respectively for the evidence based learning paths, with more than one learning goal occupying a level. If two learning goals are on the same level, it means that they do not have any precedence relationship between themselves and can be learned in any order. From these partial orderings, we generated a pair-wise ordering of all the concepts in the learning goal set and assigned each pair to a class \((C_1 < C_2)\) or \((C_1 \geq C_2)\). For example, if concept A is on level 1, concept B and C on level 2, then the orderings AB and AC will be assigned to class \((C_1 < C_2)\) i.e, concept A should be learned before concept B and concept A should be learned before concept C. The pair-wise orderings BC, BA, CA and CB will be assigned to class \((C_1 \geq C_2)\) because the partial ordering does not indicate that B and C should be learned in any specific order, and the partial ordering says they should be learned after concept A, not before. Table 5.4 describes this second set of materials, which we will use as the test data for the study. The Expert column identifies which of the two experts generated the associated data set. The Instances column is the number of pairs that were generated by pairing learning goals. The \(C_1 < C_2\) column is the percentage of instances that belong to this class and \(C_1 \geq C_2\) column is the percentage that belong to this class. Numlevels column is the number of levels into which the experts put the learning goals, since they did not produce a full ordering only a partial ordering. In the rows where we use a combination of the sequences produced by both subject experts, we are using the pairs of orderings for which they both agreed on the class label. This is why the number of instances in those rows is less than the numbers in rows where we only use sequences from one expert.

### Method 1: Training models from proxy task

The proxy task was ordering sentences by grade. In this task, the model is given two sentences \(s_1\) and \(s_2\), one written for middle school, \(s_1\) and another written for high school, \(s_2\), and asked to decide whether \(s_1 < s_2\) or \(s_2 < s_1\). We expect that a model for ordering sentences by grade should also be a reasonable model for ordering concepts for a pedagogical learning path. And importantly, getting
grade ordering data automatically is easy: the Digital Library for Earth System Education (DLESE) contains a variety of Earth science resources with metadata about the grade level they were written for.

We extracted 1702 unique non-stopwords from the training data, resulting in 3404 features per concept. For each word in each concept, we include the following two features:

- **local word count** - the number of times the word appeared in this concept
- **global word count** - the log of the ratio between the number of times the word occurred in the concept and the number of times it occurred in a background corpus, Gigaword (Graff, 2002)

These features are motivated by the work of Tanaka-Ishii et al. (2010) that showed that local and global word count features were sufficient to build a pair-wise readability classifier that achieved 90% accuracy. To generate a vector from a pair of concepts, we subtract the values of each feature from both concepts. This results in a vector containing 3404 features which are the differences between the feature values in both concepts. We used the development data set to tune the parameters of the classifiers. The best result for the Maxent model was having the ridge set to 0.1. For SMO, the best result was using a polynomial kernel of degree 1, setting the complexity parameter, C, to 0.01 and normalizing the data. For LibSVM, the best result was produced using a polynomial kernel of degree 1, setting the complexity parameter, C, to 0.1 and normalizing the data. Table 5.5 shows the accuracy results for evaluating the four classifiers on the testing data from the third set of materials, the pair-wise ordering of middle and high school sentences.

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy with Subtraction of Features and Optimized parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>naive Bayes</td>
<td>84.9%</td>
</tr>
<tr>
<td>Maxent</td>
<td>82.1%</td>
</tr>
<tr>
<td>LibSVM</td>
<td>83.5%</td>
</tr>
<tr>
<td>SMO</td>
<td>80.9%</td>
</tr>
</tbody>
</table>

**Method 2: Training models from 6-fold cross validation**

We used 6-fold cross validation on the second set of materials, the pair-wise orderings from the subject experts, to train the four models. Cross-validation is a standard technique used in machine learning for training a model and estimating its’ accuracy, especially when we do not have enough data to have dedicated training and evaluation sets. In K-fold cross validation, the data set is split into mutually exclusive subsets of approximately equal size for training and testing. The model is then trained and evaluated K times. The result is the average of the values produced by evaluating the model K times. We needed to use this method because we have a limited supply of expert data.

As a starting point, I opted to divide the data set into two equal parts 3 times, odd and even concepts, first fifteen and remaining concepts, middle fifteen and remaining concepts. I used each part as both as a training set and then as a testing set. This resulted in K being set to 6. Because we are doing a pair-wise ordering, we had to ensure that the two ways of ordering a pair of concepts belonged in the same training or testing set. We used the same word features that were used in the proxy task, local and global word count, because the domain is the same. But in addition, we included a baseline feature, which is the difference between the concept positions when they come out of COGENT. We did not tune the parameters as we did not have any development set on which to tune.
CHAPTER 5. RESEARCH DESIGN

Results

For evaluating our models, we had two baselines. The first baseline was the most frequent class baseline ($C1 \geq C2$). The second baseline was the order in which COGENT outputs the concepts, recast as pair-wise judgments. The gold standard pair-wise judgments were the pair-wise judgments inferred from the experts sequence of learning goals.

Table 5.6 shows the accuracy results to date. We calculated precision, recall and F-score for both classes but the results are not shown here. The Majority baseline column displays the majority class baseline accuracy result. The COGENT baseline column shows the COGENT baseline accuracy result. In the table, P refers to proxy task and NB to naive Bayes. Thus P-NB is the accuracy result from evaluating the naive Bayes model produced from the proxy task. While 6X refers to 6-fold cross validation, so 6X-NB is the accuracy result from evaluating the 6-fold cross validation naive Bayes model.

<table>
<thead>
<tr>
<th>Id</th>
<th>Majority baseline</th>
<th>COGENT baseline</th>
<th>P-NB</th>
<th>P-SMO</th>
<th>P-Lib SVM</th>
<th>P-Maxent</th>
<th>6X-NB</th>
<th>6X-SMO</th>
<th>6X-Lib SVM</th>
<th>6X-Maxent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>55</td>
<td>44</td>
<td>39.5</td>
<td>41.5</td>
<td>41.5</td>
<td>54</td>
<td>55.9</td>
<td>51.4</td>
<td>55</td>
</tr>
<tr>
<td>3</td>
<td>50</td>
<td>62</td>
<td>51.8</td>
<td>49.9</td>
<td>45</td>
<td>50.8</td>
<td>53</td>
<td>53.6</td>
<td>53.5</td>
<td>53.6</td>
</tr>
<tr>
<td>5</td>
<td>50</td>
<td>62</td>
<td>59.2</td>
<td>54.8</td>
<td>55.8</td>
<td>56.3</td>
<td>49</td>
<td>50.7</td>
<td>52.6</td>
<td>50</td>
</tr>
<tr>
<td>7</td>
<td>50</td>
<td>50</td>
<td>40.6</td>
<td>40.1</td>
<td>33.5</td>
<td>44.1</td>
<td>53.5</td>
<td>53.9</td>
<td>55.2</td>
<td>53.9</td>
</tr>
<tr>
<td>1,3,5,7</td>
<td>51.5</td>
<td>66</td>
<td>51.9</td>
<td>43.7</td>
<td>43.4</td>
<td>48.4</td>
<td>44.3</td>
<td>43.3</td>
<td>51.8</td>
<td>46.6</td>
</tr>
<tr>
<td>8</td>
<td>50</td>
<td>62</td>
<td>46.7</td>
<td>45.6</td>
<td>47.3</td>
<td>45</td>
<td>50.3</td>
<td>53.9</td>
<td>48.7</td>
<td>51.6</td>
</tr>
<tr>
<td>9</td>
<td>50</td>
<td>59</td>
<td>46.9</td>
<td>44.9</td>
<td>49.2</td>
<td>43.6</td>
<td>41.9</td>
<td>44.6</td>
<td>46.9</td>
<td>45.9</td>
</tr>
<tr>
<td>10</td>
<td>50</td>
<td>56</td>
<td>57.8</td>
<td>50.8</td>
<td>56.4</td>
<td>51.7</td>
<td>41.6</td>
<td>39.9</td>
<td>38.9</td>
<td>42.8</td>
</tr>
<tr>
<td>11</td>
<td>50</td>
<td>51</td>
<td>51.4</td>
<td>52.4</td>
<td>50</td>
<td>54.3</td>
<td>44.7</td>
<td>49.1</td>
<td>46.5</td>
<td>47.8</td>
</tr>
<tr>
<td>12</td>
<td>50</td>
<td>53</td>
<td>51.4</td>
<td>53.1</td>
<td>51.6</td>
<td>54</td>
<td>52.9</td>
<td>52.9</td>
<td>49.2</td>
<td>53.3</td>
</tr>
<tr>
<td>8,10</td>
<td>51.5</td>
<td>63</td>
<td>53.3</td>
<td>47.4</td>
<td>52.7</td>
<td>47.7</td>
<td>42.3</td>
<td>47</td>
<td>47.1</td>
<td>40.5</td>
</tr>
<tr>
<td>9,11,12</td>
<td>51.5</td>
<td>57</td>
<td>52.2</td>
<td>56</td>
<td>50.6</td>
<td>57.3</td>
<td>54.1</td>
<td>53.4</td>
<td>51.6</td>
<td>57.7</td>
</tr>
</tbody>
</table>

Findings and Discussion

As can be seen from Table 5.6, to date, no single model has been able to perform consistently better than the majority class baseline. The COGENT baseline outperforms all the models and the majority class baseline. Generally when you get results like these, it means your features are not descriptive enough to train a good discriminating model. Therefore I am going to explore other features and also look at other mechanisms for combining features. I might also re-cast the task as a regression problem. In addition, I am also exploring creating two types of prioritization models. A coarse model that classifies core learning goals into two levels, middle or high school and a finer-grained model that produces an ordering among all the core learning goals. Should at the end, none of these algorithms match or exceed the performance of the COGENT baseline, I will use that.
5.3 Study 3 - Automatic Prioritization of student misconceptions (Proposed)

Purpose

The purpose of this study is to create an algorithm that automatically prioritizes student misconceptions. As in the previous studies, we hope to emulate the prioritization that a subject expert would generate when asked to prioritize misconceptions.

Materials

The data for this study will come from a learning study that is being run on CLICK by our affiliates at the University of Utah this summer, commencing in June. The study will be run with 100 students, who will be asked to write two essays about their understanding of weather and climate at different points in time. We will use the misconceptions detected by CLICK from the students two essays as the data for this study.

Methodology

Our approach to automatically prioritizing the students misconceptions is to understand and model expert Earth science teacher processes for prioritizing student misconceptions. We will do a small-scale annotation study in which two expert Earth science teachers will prioritize the misconceptions that have been identified in the essays written by 100 students. The annotators will be given the essays with potential misconceptions identified by CLICK highlighted within the essay and repeated on the bottom of the essay as a list. In addition to prioritizing the misconceptions in the essays, they will also be asked to explain how they came up with each prioritization sequence. We will analyze this data and determine features and heuristics for coming up with an effective strategy for prioritizing student misconceptions. To date, we have identified some heuristics that could be used for prioritizing students misconceptions.

1. Alignment to learning goals that are in a pedagogical sequence. For example if a student has two misconceptions, a and b and they are aligned accordingly to two of the learning goals, L1 and L2 where L2 comes before L1 in the pedagogical sequence, then, b should have a higher priority than a.

2. Position where the misconceptions occurred in the students’ work. It might be judicious to prioritize misconceptions that occur earlier in the student’s work than later. The reason being that the earlier misconception might have given rise to the later misconception. For coherency sake, it could also be better to prioritize earlier misconceptions so that students can remedy the misconceptions in their essays starting from the top of the essay and going down, rather than jumping around in the essay.

3. Alignment to Bloom’s taxonomy (Bloom et al., 1971) of educational objectives. We could identify what type of sentence the misconception is using the codes - analogy, explanation, definition and example - from another study (Foster et al., 2012). The intuition here is that if an example aligns to Bloom’s lowest level, definition to the level above that, explanation to the next level after that and then analogy to the next level. Then we would have to determine if it is better to address higher or lower level misconceptions first.

We will divide the data from each expert into two, use the first one for developing the heuristics and use the second set for testing the heuristics. We expect to put the misconceptions into a fully
ordered list. We will evaluate our results using the accuracy, precision, recall and f-measure constructs. For each pairing \( a < b \) and \( b < a \), the task would be to either assign a value of true or false to the equation. With true meaning that the first misconception precedes the second misconception and false meaning that the first misconception does not precede the second misconception. Accuracy will be calculated as the percentage of instances that were assigned to the right class. Precision for each class would be the ratio of the number of instances that were correctly assigned to the class to the total number of instances that were assigned to the class. Recall for each class would be the ratio of the number of instances that were correctly assigned to each class to the number of instances that belong to the class. F-measure will be the harmonic mean of the precision and recall.

**Anticipated Results**

The anticipated result for this study is an algorithm for prioritizing student misconceptions. This algorithm will be embedded in the CLICK2 system.

### 5.4 Study 4 - Design workshop for an educational recommender system that supports conceptual change (Proposed)

**Purpose**

In this study, we will be studying how educational recommender system interface can be enhanced with research-based support mechanisms from conceptual change, specifically, analogies and dissonance strategies. These mechanisms have been shown to promote conceptual change in the classroom environment when used together (Murphy and Alexander, 2008).

According to conceptual change theory, each of these strategies can be supported through presentation of textual information or through presentation of model based information. Model based information means presenting the learner with a visual or graphic depiction of the scientific phenomena.

In these design workshops, we will gather feedback from learners on different ways for representing analogy and dissonance strategies in the interface. And gather feedback on how they want to interact with the support mechanisms in the interface.

**Materials**

The data for this study will be common instruments for design workshops such as scenarios, mock-ups, video tapes and detailed notes from the think-aloud and debrief sessions.

**Methodology**

This study will be a series of two design workshops and a final debrief session. We hope to recruit about 3 to 4 students for each of the design workshops and then bring all the students back together during the debrief session. The two conceptual change support mechanisms we will be investigating how to incorporate into an ERS are; analogy and dissonance.

In the first workshop, the students will be given a detailed scenario of potential use of the system and design mock ups of how these mechanisms could be incorporated into an ERS. The initial design will come from a design study of the CLICK environment that we performed before we decided to incorporate conceptual change support mechanisms into the system. The initial design will reflect how we think the support mechanisms might be incorporated into the ERS. Then, we will ask the students to redesign the feedback environment; i.e, suggest if and how they want it to appear on the
screen, and what sort of characteristics they want associated with the mechanisms such as the type, design and position of tags for the recommended resources and if they want to see information about how much of the learning goals they have fully mastered. We will be video-taping these think-aloud redesign sessions so we can have a record of what features the students felt were useful and usable and how they would like to see it instantiated in the feedback environment.

In the second workshop, the new set of students will be given the redesigned mock ups and asked to use it, with the functionality implemented manually in a wizard of oz. Hopefully, at this point, we will have captured a lot of what students would like to see in such an educational recommender system and this new set of students will just make clear any design loopholes especially in transitions between screens, colors, labeling of information and perhaps instructions on how the different mechanisms work. We understand that we cannot design to every students’ specification but we would like to ensure that most of the students involved in the design workshops are comfortable with the final design of the environment. In the final debrief session, we will invite all the students back and have them review the final design.

**Anticipated Results**
The anticipated result for this study is an interface design and some design recommendations for how to integrate analogy and dissonance strategies into educational recommender system interfaces.

### 5.5 Study 5 - Qualitative learning study to examine student processes and outcomes (Proposed)

**Purpose**
This purpose of this learning study is to understand how CLICK2 influences learner processes and outcomes. We are interested in two outcomes which can be phrased as questions: (1) how did the students use the conceptual change support mechanisms in CLICK2? and (2), how did the use of CLICK2 affect students’ science understanding, specifically, their explanation of the reason for seasons?

**Materials**
The study will use CLICK2, the system resulting from the previous four studies. And the data to be collected will be the student essays and their associated misconceptions. In addition, we will have video tape recordings of the students using the system, their think-aloud utterances, detailed notes from observing the students while they are using CLICK2 and detailed notes from the debrief sessions.

**Methodology**
This study will be a pilot learning study with two stages and no control, only an experimental group. In the first stage, the students will be asked to write an essay explaining in as much depth as they can, the reason for the seasons. The essays will be analyzed by CLICK2 and the students misconceptions identified. Then in the second stage, the students will be given the CLICK2 feedback environment to review their misconceptions and rewrite their essay. After both stages, the students will be interviewed to ensure that what they wrote truly matches their understanding of the reason for the seasons. How students used the support mechanisms will be investigated by analyzing the video-tapes of the students during their use of CLICK2, analyzing their think-aloud utterances while they were using the system and then interviewing them about their use of the support mechanisms after they use the system. The
second question will be addressed by analyzing the essays that the students produced and interviewing them after they write each essay.

The student essays will be evaluated quantitatively and qualitatively. For the quantitative assessment, we will ask expert Earth science teachers to score both of each student’s essay holistically on a scale of 0 - 10. We will use taxonomic categories for the qualitative assessment. We will examine the misconceptions that were identified in the students’ essay in the first stage of the learning study and then examine their current understanding of those misconceptions in the second essay. We will code the misconceptions in the first essay and the current understanding in the second essay on a 4 scale. Level 1 will reflect a naive mental model, level 2, a lower synthetic mental model, level 3, an upper synthetic mental model and level 4, a scientific mental model, perhaps with more fine-grained coding within each level as was done in Plummer et al. (2011). We will analyze the codings to determine if and by how much students using the CLICK2 environment underwent conceptual change in their understanding of the reason for the seasons.

Anticipated Results

The anticipated results of this study is an understanding of how the support mechanisms do or do not influence students’ understanding. I will also work to make the data collected, reusable.
Chapter 6

Limitations, Risks and Mitigation

Risks and Mitigation

As with any research, there is a possibility that things will not go as planned and I could end up with unexpected results. Below, I discuss the possible risks I see associated with the proposed dissertation and how I plan to mitigate them.

1. **Automatic misconception detection algorithm does not perform well enough**

   There is a possibility that the misconceptions that CLICK will detect will not really be misconceptions or could be misconceptions that I do not worry about in this research such as spelling or punctuation errors. This can affect two of my studies, Study 3 - the misconception prioritization study and Study 5 - the pilot learning study. If there is a problem with the misconception detection module, I will have a high school teacher, PhD student in Earth science education or Earth science digital library professional go through the students’ work and identify the misconceptions i.e, I will employ a manual method to get the misconceptions instead of an automatic method.

2. **No effect in Learning Study**

   There is a chance that I will see no effect in the final study and hence will not be able to show that creating an educational recommender system using the conceptual change theory can improve students’ understanding of science. I acknowledge that this might happen, especially since I plan to run the learning study with only about 6 students. However, even if I do not find any effect, this work will still be useful to:

   (a) show how we can use a learning theory to drive the design of an educational recommender system.

   (b) show how the research-based conceptual change mechanisms can be translated into a different context.

   (c) produce data i.e. representative samples of college freshmen understanding of seasons, which can be used in further research.
Limitations

As with all research, this dissertation proposal has several limitations

1. The system (CLICK2) will be evaluated only on the topic of seasons and the results may not generalize to other domains. Investigating generalizability would be for future work.

2. The system does not evaluate all the components of the system using the same topic. The first three studies are carried out on the topic of plate tectonics while the design of the interface and the learning study are done on seasons.

   I started out with plate tectonics because that was the focus of the original CLICK work. But when I adopted conceptual change theory, I decided to switch to seasons because it is a very richly researched topic in terms of conceptual change. In the larger CLICK project, we are studying the generalizability of the algorithms for Earth science, weather and biology. Therefore, it is reasonable to assume that the algorithms that I assessed with plate tectonics can be used for seasons because our research indicates that algorithms developed for plate tectonics work fine in a near domain like weather.

3. Again while this research will yield great preliminary evidence about the utility of my conceptual framework, to limit the scope of my research, I focus on the cognitive aspects of conceptual change theory. Future research should bring in the affective aspects.
# Chapter 7

## Timeline

<table>
<thead>
<tr>
<th>Year</th>
<th>Tasks and Work Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summer 2012</td>
<td>- Finish up Study 2: sequencing of learning goals</td>
</tr>
<tr>
<td></td>
<td>- Conduct Study 3: prioritization of student misconceptions</td>
</tr>
<tr>
<td></td>
<td>- Develop protocol for Study 4 and Study 5</td>
</tr>
<tr>
<td></td>
<td>- Get IRB approval for Study 4 and Study 5</td>
</tr>
<tr>
<td></td>
<td>- Write introductory dissertation chapters</td>
</tr>
<tr>
<td>Fall 2012</td>
<td>- Conduct Study 4: design workshop for ERS</td>
</tr>
<tr>
<td></td>
<td>- Construct CLICK2 environment</td>
</tr>
<tr>
<td></td>
<td>- Finalize task for Study 5: pilot learning study</td>
</tr>
<tr>
<td></td>
<td>- Write chapters on first 3 studies</td>
</tr>
<tr>
<td>Spring 2013</td>
<td>- Conduct Study 5: pilot learning study</td>
</tr>
<tr>
<td></td>
<td>- Analyze study 5 data</td>
</tr>
<tr>
<td></td>
<td>- Continue writing</td>
</tr>
<tr>
<td>Summer 2013</td>
<td>- Finish up dissertation, submit and defend</td>
</tr>
<tr>
<td></td>
<td>- Graduate by the beginning of August</td>
</tr>
</tbody>
</table>
## Appendix A

### AAAS plate tectonics learning goals

<table>
<thead>
<tr>
<th>#</th>
<th>AAAS plate tectonics learning goal</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>PT-BMK-MS1</td>
<td>The interior of the Earth is hot. Heat flow and movement of material within the Earth cause Earthquakes and volcanic eruptions and create mountains and ocean basins.</td>
</tr>
<tr>
<td>(2)</td>
<td>PT-BMK-MS2</td>
<td>Matching coastlines and similarities in rock types and life forms suggest that today’s continents are separated parts of what was long ago a single continent.</td>
</tr>
<tr>
<td>(3)</td>
<td>PT-BMK-MS3</td>
<td>The Earth first formed in a molten state and then the surface cooled into solid rock.</td>
</tr>
<tr>
<td>(4)</td>
<td>PT-BMK-MS4</td>
<td>There are a variety of different land forms on the Earth’s surface (such as coastlines, rivers, mountains, deltas, and canyons).</td>
</tr>
<tr>
<td>(5)</td>
<td>PT-BMK-MS5</td>
<td>Some changes in the Earth’s surface are abrupt (such as Earthquakes and volcanic eruptions) while other changes happen very slowly (such as uplift and wearing down of mountains).</td>
</tr>
<tr>
<td>(6)</td>
<td>PT-BMK-MS6</td>
<td>Vibrations in materials set up wavelike disturbances that spread away from the source. Sound and earthquake waves are examples.</td>
</tr>
<tr>
<td>(7)</td>
<td>PT-BMK-HS1</td>
<td>The theory of plate tectonics provides an explanation for a diverse array of seemingly unrelated phenomena, and there was a scientifically sound physical explanation of how such movement could occur.</td>
</tr>
<tr>
<td>(8)</td>
<td>PT-BMK-HS2</td>
<td>Earthquakes often occur along the boundaries between colliding plates, and molten rock from below creates pressure that is released by volcanic eruptions, helping to build up mountains. Under the ocean basins, molten rock may well up between separating plates to create new ocean floor. Volcanic activity along the ocean floor may form undersea mountains, which can thrust above the ocean’s surface to become islands.</td>
</tr>
<tr>
<td>(9)</td>
<td>PT-BMK-HS3</td>
<td>Ocean-floor plates may slide under continental plates, sinking deep into the Earth. The surface layers of these plates may fold, forming mountain ranges.</td>
</tr>
<tr>
<td>(10)</td>
<td>PT-BMK-HS4</td>
<td>The Earth’s plates ride on a denser, hot, gradually deformable layer of the Earth.</td>
</tr>
<tr>
<td>(11)</td>
<td>PT-BMK-HS5</td>
<td>The slow movement of material within the Earth results from heat flowing out from the deep interior and the action of gravitational forces on regions of different density.</td>
</tr>
<tr>
<td>(12)</td>
<td>PT-BMK-HS6</td>
<td>The solid crust of the Earth—including both the continents and the ocean basins—consists of separate plates. The crust sections move very slowly, pressing against one another in some places, pulling apart in other places.</td>
</tr>
</tbody>
</table>
Appendix B

Learning goal data 2 (LGD2)

1. In particular, four major scientific developments spurred the formulation of the plate-tectonics theory: (1) demonstration of the ruggedness and youth of the ocean floor; (2) confirmation of repeated reversals of the Earth magnetic field in the geologic past; (3) emergence of the seafloor-spreading hypothesis and associated recycling of oceanic crust; and (4) precise documentation that the world’s earthquake and volcanic activity is concentrated along oceanic trenches and submarine mountain ranges.

2. In an ocean-continent convergence, the collision of ocean and continental plates causes the accretion of marine sedimentary deposits to the edge of the continent.

3. The Earth’s internal heat source provides the energy for our dynamic planet, supplying it with the driving force for plate-tectonic motion, and for on-going catastrophic events such as earthquakes and volcanic eruptions.

4. A single seafloor mountain chain circles Earth and contains some of Earth’s tallest mountains.

5. Plate tectonics is the theory that Earth’s outer layer is made up of plates, which have moved throughout Earth’s history.

6. Over millions of years, plate tectonics has changed the appearance of the Earth’s crust.

7. Shaping the Ocean Floor at the Mid-Ocean Ridges

8. These ridges, formed as the Earth’s plates separate from each other, rise from the deep sea floor as volcanic mountains.

9. Geologists came to the conclusion in the 1960’s that the Earth’s rigid outer layer (crust and outer, rigid layer of the mantle) was not a single piece, but was broken up into about 12 large pieces called plates.

10. This drives the oceanic plates deep into the mantle destroying the oceanic plates.

11. If the magma reaches the surface of the Earth, a volcano forms.

12. Plate tectonics tells us that the Earth’s rigid outer shell (lithosphere) is broken into a mosaic of oceanic and continental plates which can slide over the plastic aesthenosphere, which is the uppermost layer of the mantle.

13. Hot volcanic material rises from the Earth’s mantle to fill the gap and continuously forms new oceanic crust.
14. The science of the shaping of the Earth’s crust goes by the name ”tectonics,” and the process described here is the essence of ”plate tectonics” by the Earth’s crust consists of distinct plates which are continually rearranged, sometimes carrying along continents or parts of continents.

15. According to the plate tectonic model, the surface of the Earth consists of a series of relatively thin, but rigid, plates which are in constant motion.

16. The surface layer of each plate is composed of oceanic crust, continental crust or a combination of both.

17. Most of the Earth’s tectonic, seismic and volcanic activity occurs at the boundaries of neighboring plates.

18. These plates are in constant motion causing earthquakes, mountain building, volcanism, the production of ”new” crust and the destruction of ”old” crust.

19. The motion of the Earth’s plates help scientists to understand why earthquakes, volcanoes, and mountain building occur.

20. It moved hundreds of miles in 135 million years at a great speed (4 inches per year!!!) The Indian plate crashed into the Eurasian plate with such speed and force that it created the tallest mountain range on Earth, the Himalayas!

21. The pieces of the shell are Earth’s tectonic plates – there are 12 major ones – and they float across a layer of soft rock like rafts in a stream, their motions driven by forces generated deep in the Earth.

22. Some scientists, such as David James of the Carnegie Institution of Washington, believe that the continents are anchored into the mantle by deep keels of rock that extend hundred of miles below the surface, and the continental crust and mantle therefore move in concert).

23. In the oceans, magma reaches the surface at the boundaries between plates called spreading centers, like the Mid-Atlantic Ridge, and there new oceanic crust forms.

24. The scraping of one plate on another generates powerful earthquakes; the heating of the plate within the depths of the mantle releases fluids which melt the rock over it, producing blobs of molten rock, or magma, that surface as volcanoes.

25. As the plates continue to move, and more crust is formed, the ocean basin expands and a ridge system is created.

26. Structure of the Earth
   History of plate tectonics
   Plates Plate boundaries Forces in the Earth
   Faults Hypercard Resources

27. The Earth’s surface is covered by a series of crustal plates.

28. This heated layer is the source of lava we see in volcanoes, the source of heat that drives hot springs and geysers, and the source of raw material which pushes up the mid-oceanic ridges and forms new ocean floor.

29. Deep in the Earth’s interior, convection of the rocks, caused by temperature variations in the Earth, induces stresses that result in movement of the overlying plates.
30. As the denser plate of oceanic crust is forced deep into the Earth’s interior beneath the continental plate, a process known as subduction, it encounters high temperatures and pressures that partially melt solid rock.

31. Most of the 600-plus active volcanoes on Earth are associated with the boundaries of the tectonic plates, the seven great plates that carry the oceans and continents. They are especially common in subduction zones, which occur when one plate dips beneath another.

32. As the plate dives into the mantle – the layer of hot, flexible rock on which the plates glide – it gradually is heated.
Appendix C

Pedagogical sequencing of LGD2 by two subject experts

<table>
<thead>
<tr>
<th>Level</th>
<th>Expert1-Traditional1</th>
<th>Expert1-Traditional2</th>
<th>Expert2-Traditional</th>
<th>Expert1-Evidence</th>
<th>Expert2-Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>27, 9, 16</td>
<td>27, 9, 16</td>
<td>9, 27</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>(2)</td>
<td>3</td>
<td>3</td>
<td>5, 14</td>
<td>9, 27</td>
<td>9, 27</td>
</tr>
<tr>
<td>(3)</td>
<td>29</td>
<td>29</td>
<td>15</td>
<td>16</td>
<td>5, 14</td>
</tr>
<tr>
<td>(4)</td>
<td>5, 15</td>
<td>5, 15</td>
<td>21</td>
<td>5, 15</td>
<td>15, 21, 12</td>
</tr>
<tr>
<td>(5)</td>
<td>14</td>
<td>14</td>
<td>12</td>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td>(6)</td>
<td>21, 12</td>
<td>21, 12</td>
<td>16</td>
<td>12, 21</td>
<td>22</td>
</tr>
<tr>
<td>(7)</td>
<td>28</td>
<td>28</td>
<td>22</td>
<td>28</td>
<td>1</td>
</tr>
<tr>
<td>(8)</td>
<td>17, 18, 19</td>
<td>17, 18, 19</td>
<td>1</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>(9)</td>
<td>22</td>
<td>22</td>
<td>6</td>
<td>13, 23</td>
<td>17</td>
</tr>
<tr>
<td>(10)</td>
<td>6</td>
<td>6</td>
<td>17, 18, 19</td>
<td>8</td>
<td>19</td>
</tr>
<tr>
<td>(11)</td>
<td>2, 20</td>
<td>4</td>
<td>17, 18, 19</td>
<td>8</td>
<td>19</td>
</tr>
<tr>
<td>(12)</td>
<td>10, 32</td>
<td>13, 23</td>
<td>25</td>
<td>17, 18, 19</td>
<td>25</td>
</tr>
<tr>
<td>(13)</td>
<td>30</td>
<td>8</td>
<td>8</td>
<td>24</td>
<td>8</td>
</tr>
<tr>
<td>(14)</td>
<td>11, 31</td>
<td>25</td>
<td>31</td>
<td>11, 31</td>
<td>31</td>
</tr>
<tr>
<td>(15)</td>
<td>4</td>
<td>2, 20</td>
<td>2</td>
<td>10, 32</td>
<td>2</td>
</tr>
<tr>
<td>(16)</td>
<td>13, 23</td>
<td>10, 32</td>
<td>20</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>(17)</td>
<td>8</td>
<td>30</td>
<td>3</td>
<td>2, 20</td>
<td>3</td>
</tr>
<tr>
<td>(18)</td>
<td>25</td>
<td>11, 31</td>
<td>29</td>
<td>22</td>
<td>29</td>
</tr>
<tr>
<td>(19)</td>
<td>24</td>
<td>24</td>
<td>28</td>
<td>28</td>
<td>6</td>
</tr>
<tr>
<td>(20)</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>29</td>
<td>10</td>
</tr>
<tr>
<td>(21)</td>
<td></td>
<td></td>
<td>30</td>
<td>3</td>
<td>30</td>
</tr>
<tr>
<td>(22)</td>
<td></td>
<td></td>
<td>32</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>(23)</td>
<td></td>
<td></td>
<td>24</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>(24)</td>
<td></td>
<td></td>
<td>13</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>(25)</td>
<td></td>
<td></td>
<td>23</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>(26)</td>
<td></td>
<td></td>
<td>11</td>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>
Bibliography


Cartier, J. and Center, E. R. I. (2000). *Using a modeling approach to explore scientific epistemology with high school biology students*. University of Wisconsin-Madison, National Center for Improving Student Learning and Achievement in Mathematics and Science.


