EXPLORING STUDENTS' PATTERNS OF REASONING

by

MOJGAN MATLOOB HAGHANIKAR

M.S., University of Glasgow, 2002

M.S., Kansas State University, 2007

AN ABSTRACT OF A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Physics

College of Arts and Sciences

KANSAS STATE UNIVERSITY

Manhattan, Kansas

2012

Abstract

As part of a collaborative study of the science preparation of elementary school teachers, we investigated the quality of students' reasoning and explored the relationship between sophistication of reasoning and the degree to which the courses were considered inquiry oriented.

To probe students' reasoning, we developed open-ended written content questions with the distinguishing feature of applying recently learned concepts in a new context. We devised a protocol for developing written content questions that provided a common structure for probing and classifying students' sophistication level of reasoning. In designing our protocol, we considered several distinct criteria, and classified students' responses based on their performance for each criterion.

First, we classified concepts into three types: Descriptive, Hypothetical, and Theoretical and categorized the abstraction levels of the responses in terms of the types of concepts and the inter-relationship between the concepts. Second, we devised a rubric based on Bloom's revised taxonomy with seven traits (both knowledge types and cognitive processes) and a defined set of criteria to evaluate each trait.

Along with analyzing students' reasoning, we visited universities and observed the courses in which the students were enrolled. We used the Reformed Teaching Observation Protocol (RTOP) to rank the courses with respect to characteristics that are valued for the inquiry courses. We conducted logistic regression for a sample of 18courses with about 900 students and reported the results for performing logistic regression to estimate the relationship between traits of reasoning and RTOP score.

In addition, we analyzed conceptual structure of students' responses, based on conceptual classification schemes, and clustered students' responses into six categories. We derived regression model, to estimate the relationship between the sophistication of the categories of conceptual structure and RTOP scores. However, the outcome variable with six categories required a more complicated regression model, known as multinomial logistic regression, generalized from binary logistic regression.

With the large amount of collected data, we found that the likelihood of the higher cognitive processes were in favor of classes with higher measures on inquiry. However, the usage of more abstract concepts with higher order conceptual structures was less prevalent in higher RTOP courses.

EXPLORING STUDENTS' PATTERNS OF REASONING

by

MOJGAN MATLOOB HAGHANIKAR

M.S., University of Glasgow, 2002

M.S., Kansas State University, 2007

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Physics

College of Arts and Sciences

KANSAS STATE UNIVERSITY

Manhattan, Kansas

2012

Approved by:

Major Professor

Dean Zollman

Abstract

As part of a collaborative study of the science preparation of elementary school teachers, we investigated the quality of students' reasoning and explored the relationship between sophistication of reasoning and the degree to which the courses were considered inquiry oriented.

To probe students' reasoning, we developed open-ended written content questions with the distinguishing feature of applying recently learned concepts in a new context. We devised a protocol for developing written content questions that provided a common structure for probing and classifying students' sophistication level of reasoning. In designing our protocol, we considered several distinct criteria, and classified students' responses based on their performance for each criterion.

First, we classified concepts into three types: Descriptive, Hypothetical, and Theoretical and categorized the abstraction levels of the responses in terms of the types of concepts and the inter-relationship between the concepts. Second, we devised a rubric based on Bloom's revised taxonomy with seven traits (both knowledge types and cognitive processes) and a defined set of criteria to evaluate each trait.

Along with analyzing students' reasoning, we visited universities and observed the courses in which the students were enrolled. We used the Reformed Teaching Observation Protocol (RTOP) to rank the courses with respect to characteristics that are valued for the inquiry courses. We conducted logistic regression for a sample of

18courses with about 900 students and reported the results for performing logistic regression to estimate the relationship between traits of reasoning and RTOP score.

In addition, we analyzed conceptual structure of students' responses, based on conceptual classification schemes, and clustered students' responses into six categories. We derived regression model, to estimate the relationship between the sophistication of the categories of conceptual structure and RTOP scores. However, the outcome variable with six categories required a more complicated regression model, known as multinomial logistic regression, generalized from binary logistic regression.

With the large amount of collected data, we found that the likelihood of the higher cognitive processes were in favor of classes with higher measures on inquiry. However, the usage of more abstract concepts with higher order conceptual structures was less prevalent in higher RTOP courses.

Table of Contents

List of Figures

List of Tables

Acknowledgements

This research project would never have been possible without the talented helpers around me. I wish to express my heartfelt gratitude to my advisor, Professor Zollman whose extensive experience was enlightening. I astonished with his diligent support, quality research and awakening leadership. I gratefully acknowledge the occasions that he provided for me to develop my self-confidence that is a foundation for a career in physics education.

Deepest gratitude is also due to my advisory committee; Professor Rebello, whose knowledge and assistance was tremendously helpful, Professor O'Shea for his ongoing assistance and Professor Hodge for providing statistical knowledge.

Not forgetting Mrs. Kim Coy who was always there to help, especially for the plenty time she spent on developing the excel data file needed for this study. In addition, Special thanks to Allison Coy, for her contribution to my data analysis and using my rubric. I benefitted from the discussions I had with Professor Leigh Murray from Kansas State University Department of Statistics. She freely shared her time and expertise and gave thoughtful suggestions that shaped my statistical analysis.

Thanks also are due to Physics Education team at Kansas State University for participating in reliability studies and attending my seminars. I admire our numerous stimulating discussions that prepared me for future research.

This study was part of collaborative effort with other universities. I had productive conversations with Professors Dennis Sunal & Cynthia Sunal (University of Alabama) and Professor Cheryl Mason (San Diego State University) about the development of the procedure. I should also thank Donna P. Turner and Erika Steele (University of Alabama) who contributed to the RTOP measures.

I am also grateful to many brilliant researchers in physics education community, who paved the way and their talents intrigued my interest in physics education.

I wish to express my love and gratitude to my beloved mother who passed away during the completion of this work and to my sister for her immense understanding, endless love and encouragement through these years.

This research was based in part upon work supported by the National Science Foundation under Grant Number ESI-05 54594.

Dedications

I want to dedicate this thesis to loving memory of my parents who believed in me to more than I could be.

Chapter 1 - Students' Reasoning and Measures of Reform

The purpose of this study is to classify students' patterns of reasoning skills as they apply scientific concepts to answer written extended questions. To do this, we developed an assessment tool that is compatible with the proper implementation of inquiry-based learning that elicits and analyzes the students' ideas. We devised content questions, which are open ended and distinguished by their having applied recently learned concepts in a new context.

Science educators have developed different types of assessments that include ongoing evaluation on classroom discussions as well as the formative assessment of performance skills and experimentation. To apply higher stakes to a broader assessment, more attention to higher orders of thinking and performance skills is required (Resnick, 1987). There is, therefore, a definite need for more productive assessments that are not just testing bodies of knowledge; but the ability to inquire, showing competence in reasoning, developing explanations, making predictions, which also requires access bodies of knowledge (National Research Council, 1996).

However, assessing the correctness of the conclusions students reach, evaluating the skill with which they control variables, or measuring their conceptual knowledge may misinterpret the assessment of their ability to conduct inquiry. Such types of assessments are not sensitive enough for evaluating the important aspects of inquiry that are ultimately more valuable than the correctness of the conclusions students reach (Russ et al., 2008). One of those aspects is constructing a well-reasoned response that is characterized by successive statements that follow one another logically without gaps (National Research Council, 1996). To evaluate the logic of the responses, researchers have examined the students' constructions of the cause-and-effect relationship and their explanations of how a particular component of a system causes its actions. Furthermore, these researchers have focused on how the underlying process of cause and effect and its associations are explained by the students (Russ et al., 2008).

Building on such an approach, we designed an assessment tool that allowed us to evaluate the quality based on the learner's competence to reason a prediction, sophistication of a cause-and-effect relationship that appears in the responses, the students' prior knowledge and performance skills and the application of knowledge and skills in a new situation.

1.1 National Study of Science for Undergraduate Science (NSEUS)

Our study is part of a much larger nationwide overarching research project called NSEUS that explores the effectiveness of developmental intervention through different lenses. Due to the extensive nature of the NSEUS project, collaborators of the project were assigned to work on different research questions. Most of the research studies were centered on the measure of reform and its effectiveness on different learning outcomes.

NSEUS was a 5-year collaborative project conducted by three Co-PIs of the project, who were researchers from University of Alabama, San Diego State University and Kansas State University.

The NSEUS research group investigated a number of subthemes including the features of the reform courses and the effects on the end users, who were pre-service elementary school teachers. Moreover, they compared the extent to which the faculty used reformed methods In most of the universities, pre-service teachers, either enrolled in a traditional science course delivered by lecturing or they participated in a science course designed specifically for elementary education majors based on inquiry-oriented exploration of chemical, geological and physical characteristics of nature.

1.2 Thesis Organization

This thesis includes six chapters and subsequent appendices that center on exploring students' patterns of reasoning. In the first chapter, I introduce the overarching NSEUS project, types of research questions it addressed and methods used to investigate the impacts of interactive learning strategies implemented in science content courses for pre-service elementary education majors. We as a group in Kansas State University, studied the quality of students' reasoning in relation to a measure of reform, which is the subject of current thesis and its comprehensive details are reported in the following chapters.

1.2.1 *Roadmap to chapter one:*

In the current chapter, I give an overview of the overarching NSEUS project. I also present the theoretical framework behind NSEUS and the chronological development of the previous research that led researches to conduct NSEUS project. The precedent for NSEUS was an effort to promote inquiry teaching and was performed as a series of workshops a decade ago. I review the history of the events that progressively led to the organization of the above-mentioned nationwide project, the set of overarching questions that the project seeks to address, the principal research methodology, the instruments, and surveys applied in the project.

In the final part of this chapter, I discuss the specific goals that are the subject of current thesis. In particular, this thesis focuses on the following two major objectives:

a) Developing an assessment tool to code students' patterns of reasoning

b) Using the assessment tool to compare the scientific reasoning of students' who are instructed by inquiry-oriented strategies to that of students instructed by traditional teaching strategies.

1.2.2 *Other chapters*

After introducing NSEUS in Chapter 1, I review scientific literature in the following chapter. The review is centered on interactive learning theories, higher order thinking processes, and transferring of knowledge. In addition, I review the theoretical frameworks of the constructivism, and what has been learned from previous research on using inquiry-oriented assessment tools.

In Chapters 3, 4 and 5, I describe the protocol that we designed for developing content questions and discuss the underlying structures informing those questions. In addition, I discuss our content analysis strategy and present the results of the data that we analyzed. In the final chapter, I discuss conclusions and potentials for future work.

1.3 Inquiry-Oriented Teaching

During the decades following the initiation of the discovery learning movement based on Bruner's constructivist theory (1960), the education community has increasingly supported inquiry-oriented teaching strategies. This movement gained momentum after Vygoysky's (1978) proposed a cognitive theory of social constructivism, which favored socially interactive environments as being more conductive to learning than traditional teacher-centered classrooms. The discovery learning movement also drew on Dewey's (1933) definition of meaningful understanding as a rationale for the teaching strategies it advocated. For example, Wiggins and McTighe (2006) referred to the following remark by Dewey (1933): "To grasp the meaning of a thing, an event, or a situation is to see it in its relations to other things: to see how it operates or functions, what consequences follow from it, what uses it can be put to. In contrast, what we have called the brute thing, the thing without meaning to us, is something whose relations are not grasped …The relation of means-consequence is the center and heart of all understanding." Therefore, understanding is not a matter of knowing facts and memorizing procedures, but knowledge that is the result of exploring why and how things are happening (Wiggins & McTighe, 1998).

Inquiry learning, therefore, is an educational strategy in which students are placed in the position of scientists, so they can conduct their own investigations and explore phenomena as scientists approach a problem. However, to achieve this goal, teachers should select or design activities that are appropriate to the level of students' scientific skills repertoire (Keselman, 2003). In other words, learning activities should be occasions for investigation, such as partial and miniature inquiries (Schwab, 1958).

Inquiry-oriented teaching challenges the education and science communities to transform the teaching strategies. Furthermore, a plan for systemically reforming teaching strategies based on inquiry-oriented learning would affect a large number of schools. The purpose of the reform based on inquiry-oriented learning and discovery learning has been to align instruction, assessment and curriculum development and to reshape learning activities and professional training to the goals of inquiry. Therefore, Socratic dialogues based on students' questions and their prior knowledge replaced recitation and direct instruction. This classroom setting has enabled students to construct knowledge by posing questions, conducting relevant experiments, collecting data and making inferences. However, inquiry oriented teaching is not about using advanced technology; instead, it seeks to encourage understanding through collaborative work and peer interactions that allow students to explore natural phenomena by finding evidence for their theories and then allowing them to refine their theories and mental models.

However, literature has repeatedly reported the lack of commonly accepted definition for inquiry (Bybee, 2001; Southerland et al., 2003). The perception that inquiry-oriented learning is about teaching the scientific methods is misleading. According to Schwab (1958), if science aimed to seek unalterable truths, then education appropriate to such a view would be the mastery of those truths by practicing the application of those truths. However, in contrast to the straightforward and clear methodology of traditional instruction, the inquiry-oriented teaching method is ambiguous. As Klapper (1995) observed, the scientific process is fuzzy since scientific rules lack explicitness and the practice of science differs from science as a body of knowledge. Accordingly, the roles of teachers are crucial in this type of classrooms Teachers must precisely understand what inquiry means as well have a sufficient understanding of the scientific context for the techniques of inquiry. Teaching by inquiry often criticized, for being unstructured and disorganized. However, inquiry can be delivered as guided or structured inquiry, in a way that question and procedures can be provided by the teacher and students generate a hypothesis supported by the evidence that comes from their explorations. The role of teacher is to prepare conditions under which students can best learn. The teacher is a mentor and facilitator who leads students through the scientific process and not a source of knowledge who provide answers to all student questions (Wenning, 2011).

Since a consensus view as to how to teach, practice and assess inquiry does not exist, teachers should be guided by standards to consider what students should learn, how students should learn and what their abilities are. In addition to state and district instructional standards that specify teaching requirements, the National Research Council (NRC, 1996) and American Association for the Advancement of Science set forth standards to define a framework for systematic reform of how science should be thought. The documented standards advocate more emphasis on investigation and knowledge construction through inquiry and less emphasis on memorizing facts (Marx et al., 2004). These standards target teachers, department chairs, curriculum directors, administrators, publishers, and school committees.

1.4 Background History

To promote a large-scale reform in the way science is taught; NASA supported a series of extensive workshops conducted by University of Alabama. Beginning in 1996, the faculty from 103 institutions of higher education participated in these professional developmental workshops. This national reform project called NOVA (NASA Opportunities for Visionary Academics) was established to facilitate reform in higher education and to promote the science literacy of pre-service teachers (Sunal & Sunal, 2009). Aligned with the National Science Education Standards (NRC, 1996) and Benchmarks for Science Literacy (AAAS, 1993), the workshops were designed to reflect the activities of the proposed reforms and the use of inquiry-oriented science material. The effort was aimed at courses for elementary education majors, and the faculty members who participated in the workshops were involved with teaching pre-service teachers. Each team that participated in the NOVA workshops received funding for course development. Afterwards, the participants modified 150 courses to support the desired teaching reforms and to meet the needs of elementary education majors.

1.5 National Study of Education in Undergraduate Science (NSEUS)

After promoting reform-based instructional practices, it was reasonable to ask:

- What was the impact of reform-based instructional practices on the target population?
- What really happens in the classrooms and do the teaching practices reflect the framework portrayed in the standards?
- What are the teachers' perceptions about the reform and science?
- Was the intervention effective enough to change the teachers' beliefs about nature of science?

Therefore, conducting additional research seemed essential for characterizing the effectiveness of the reform-based instructional practices program. The research project called NSEUS has been investigating a number of subthemes including the features of the reform courses and the effects on the end users, who are undergraduate students and elementary school teachers. More specifically, the NSUES study was started with the following research questions:

1. How do the reform-based instructional practices change undergraduate science faculty members' teaching practices based on evidence of their descriptions of their curricular priorities, lesson planning, instruction and pedagogical decision-making?
2. How do the characteristics, learning environments, course structure, pedagogical content knowledge, and amount of collaboration differ between the

reform courses and comparison (traditional) courses and how do these differences affect the learning outcomes among undergraduate students?

3. How do the levels of the science courses, characteristics learning environments, course structure, pedagogical content knowledge, and amount of collaboration differ among the reform courses (treatment only) and how do these differences affect the learning outcomes among the undergraduate students?
4. How do the characteristics of reform and traditional science courses influence the in-service K-6 teachers as they conduct their own science classes? How can these influences be compared?

1.6 Research Methodology of NSEUS Project

After a detailed review of the research literature published in the area of inquiry-based instruction, NSEUS derived a list of criteria that could be used to define a course as a reformed course. These criteria include whether the course does the following (Sunal et al., 2008):

- Allows all students to learn science
- Considers the students' prior knowledge
- Promotes inquiry-based pedagogy and encourages student participation in the process of instruction
- Refocuses the role of instructor who participates in a learning environment to be a listener who coaches students and uses action research to improve teaching
- Uses integrated learning styles
- Uses interdisciplinary approaches

- Involves the interaction of faculty from different colleges (Arts and Sciences and Education) within the university
- Focuses on collaborative learning
- Uses technology to facilitate learning
- Uses assessment strategies that evaluate performance and meaningful understanding
- Emphasizes evidence-based learning, which uses actual data and employs the scientific method

In the research conducted by NSEUS, both qualitative and quantitative research methodologies are used. The data collection methods include multiple quantitative and qualitative instruments to analyze and conduct comparative and relational studies.

1.6.1 *Sampling*

The samples of universities drawn for the study conducted by NSEUS, were chosen from 103 higher education institutions that participated in the NASA/NOVA professional development program over the last 10 years. The universities sampled were stratified by institutional type. After choosing the universities, the sample of 35 science courses was selected from these institutions. Some of the chosen courses were modified by the faculty who participated in the NASA/NOVA professional development program. The rest of sample courses were not involved in the NOVA process. The target populations were the faculty of the chosen courses, their pre-service undergraduate students and K-6 in-service teachers who had taken these courses while they were undergraduates.

Most of the faculty, who intended to adopt inquiry-oriented approach, were among the original participants in the NOVA project. However, some of the faculty members did not receive professional development through NOVA, and they were teaching courses that they inherited from the NOVA instructors who had originally participated in the workshops.

In sum, not all the NOVA courses were practicing authentic inquiry even though that was the intent. Many teachers had wrong perceptions about inquiry, for example, they thought that asking students many questions constitutes inquiry or using hands on activities implies inquiry (Wenning, 2011). In between the purely traditional and reformed teaching practices, we found a spectrum with many discernible characteristics that differentiated various levels of reform. Therefore, we ranked classes by the degree that they were measured to be reformed. In the following chapters, I will discuss in detail the criteria that we used for this ranking process.

1.7 Instruments developed and used by NSEUS

The process of reform consists of different subsets and dimensions. For example, in one strand reform affects university faculty members, their collaboration with each other and their teaching performance. In another strand, reform concerns elementary education majors and in-service teachers. For pre-service teachers, the NSEUS team considered students' learning achievements and their interaction with the teaching environment. However, for in-service teachers the NSEUS team looked at their teaching strategies and attitudes toward learning and teaching. To investigate the effect of these characteristics, NSEUS researchers developed some instruments or used several contemporary well-known standardized instruments. Because the attributes that these instruments measure differ, our NSEUS collaborators subsequently choose an appropriate instrument for each attribute they desired to examine. The ultimate goal was to compare the outcomes for the reformed and traditional courses for each set of collected data. The types and the formats of instruments varied due to the different nature, characteristics and elements of each research question. Some of the instruments were in the form of Likert scale questionnaires; others included observational tools, interviews, open-ended questions, etc. In the following sections, I briefly review each instrument with respect to the relevant characteristic.

1.7.1 *Online of Questionnaire Type Instruments*

Thinking about Science Survey Instrument (TSSI)

Knowing how scientific knowledge has been integrated through history and by analogy can help students conceptualize scientific inquiry as the process by which scientific knowledge was developed (Lederman, 2004). We needed to know whether teachers regarded science as a way of knowing and appreciated how scientists approached learning science, or whether they considered science to be a set of truths, or whether they believed that authority or right answers would lead to practicing those truths (Lederman, 1992). Therefore, this part of study that was conducted by our NSEUS collaborators, aimed to examine whether the pre-service and in-service teachers changed their views toward learning science since the earlier mentioned intervention has employed

The “Thinking about Science Survey Instrument” (TSSI) is a 35-item Likert-scale ('Strongly agree', 'Agree', 'Disagree', 'Strongly disagree') questionnaire (Cobern, 2000). We administered TSSI in an online format to undergraduate students and to in-service K-6 teachers independently of the site visits.

Constructivist Learning Environment Survey (CLES)

The CLES is a survey that consists of 20 statements, which participants are asked to rate on a five-point Likert-scale from Strongly Disagree to Strongly Agree (Taylor et al., 1997). The purpose of the survey is to measure the development of constructivist approaches in the classroom (increasing the role of students in the classroom in helping to construct their own learning), as perceived from the teachers' and students' points of view. The survey questions can be classified into the following five categories: learning about the world and how students see the relevance of learning environment, learning about science and about the uncertainty of science, learning to speak out and be critical, learning to learn and learning to communicate. The survey was usually administered to undergraduate students (pre- and post) and to in-service K-6 teachers independently of the site visits.

Science Teaching Efficacy and Beliefs Instrument (STEBI-A and B)

STEBI-A and B (Riggs, 1990) are used to measure the efficacy of teachers' teaching science. STEBI-A was designed to measure elementary school teachers' self-efficacy beliefs. STEBI-B is a modified version of STEBI-A designed for pre-service teachers. NSEUS administered STEBI-B online to undergraduate education majors as a posttest and STEBI-A to in-service K-6 teachers during site visits. The survey includes two types of components. First, it measures the teachers' level of confidence in their own teaching abilities and, second, their beliefs about how their students' learning can be influenced by effective teaching. Each question on the survey is specific to one of the two components, but the two types are combined. The survey uses a 5-point Likert scale (Strongly

Agree/disagree). STEBI-A and STEBI-B were administered to elementary school teachers and pre-service teachers respectively.

1.7.2 *Reformed Teaching Observation Protocol (RTOP)*

RTOP (Pilburn et al., 2000; Swada and Pilburn, 2000) is an observational instrument that is used to measure the degree to which a science classroom is "reformed." For this observational protocol, the characteristics of reformed teaching practices are based on national standards and inquiry oriented math and science education. The characteristics are organized into the following five categories:

1. Lesson Design and Implementation
2. Content: Propositional Pedagogic Knowledge
3. Content: Procedural Pedagogic Knowledge
4. Classroom Culture: Communicative Interactions
5. Classroom Culture: Student/teacher Relationships

Each category includes five questions according to which the observer ranks the occurrence of the event on a scale of 0-4. Observations for the RTOPs took place during site visits in the middle of the semester. The university faculty members and K-6 teachers were participants whose classes we observed.

1.7.3 *Interviews*

We interviewed university faculty and in-service teachers individually and undergraduate students in small focus groups (5-6 students). We conducted semi-structured interviews in all cases. We asked both university faculty and in-service teachers about their teaching strategies and experiences. Our focus group interviews sought the students' opinions about the course and their attitudes toward science. Furthermore, we asked in-service teachers about their learning experiences, and ask if they see themselves as becoming teachers that are more effective in the future. All interviews took place during site visits during the middle of the semester.

1.7.4 *Open-ended type questions*

Draw a Scientist Test (DAST)

This DAST was used to assess the students' perceptions about the appearance of scientists. Originally developed by Chambers (1983), an open-ended projective test aims to explore children's perceptions of scientists. In using DAST, investigators ask students to draw how they imagine a scientist. There is a checklist for DAST, which represents a stereotypic characteristic derived from the reviews of literature relating to students' images of scientists. The checklist has 11 items and for each checked item, evaluators give one point. Each item on the protocol used to analyze the DAST data represents a stereotypic characteristic. Therefore, when more items are "checked" on the DAST analysis more stereotypes are displayed in the student's drawing. Having each student add short personal narrative can help evaluators interpret better drawings, since interviewing each student is not practical. This instrument was administered to pre-service teachers as pre- and post-test.

Science Content Assessment (SCA)

For the SCA, students answer an open-ended content question(s) on a major concept they have learned. This question aims to measure students' content and critical thinking achievement with respect to scientific concepts of the course. The question(s) should reflect an application and conceptual understanding of what they have learned when given a situation that is novel. The question was administered as a part of a quiz or examination embedded with other questions in the exam, or as an activity in a class session during the last couple of weeks of the course. Consequently, we developed SCA and used it only for post-testing. Each science content question should have the following characteristics:

1. Requires thinking beyond memorization, so the questions should be rich in detail and require some sort of logic and reasoning connecting the concepts that needs evaluation. This requires comparing and contrasting two or more concepts and discussing relationships and differences while explaining the phenomenon involved
2. Requires the student to embed newly learned concepts in a way that explanation and justification requires logical connections and to relate phenomena in which similar rules and concepts hold
3. Ascertains why students used the particular concepts and particular connections, and what led them to determine the relationships among concepts
4. Determines whether students know when to use an appropriate concept, and whether their answers reflect an understanding of the meaning of that concept
5. Determines whether students know when, why or how to use the rules, procedural skills and/or and principles related to the concepts. In particular, the question should allow the investigator to determine whether students know the procedural skill or rule, but have applied the wrong aspect of the rule or principle
6. Leads to a thinking process that first restructures and then synthesizes scientific conceptual knowledge, and then applies that knowledge to explain the new context

7. Includes qualitative rather than quantitative features

1.8 Our Group's Responsibilities in NSEUS Project

In the previous sections, I explained that NSEUS aims to investigate the impact of reform through different lenses. NSEUS conducted several parallel research projects to compare the achievements and quality of reformed and traditional courses. It is sensible that different research questions and responsibilities were distributed among three collaborating institutions of the NSEUS.

1.8.1 *Site visits*

We took this study into the field to explore what was really happening in the classes. Site visits provided us an opportunity to observe university faculty's and elementary school teachers' classes. Mostly, in every university, two instructors were involved, one who taught a traditional course and another who taught the inquiry-oriented course. The initial plan was to observe and to compare the two types in each university and elementary school. However, not every university and elementary school offered both types of courses and for some of the schools the traditional counterpart was missing. Consequently, as an indicator of the degree that the course has been reformed we assigned a set of RTOP scores to each class, and examined the comparison across the schools. As one of the collaborators of NSEUS, we at Kansas State University carried out almost one-third of site visits and conducted research to investigate the characteristics of the reform courses and their impact on students' reasoning skills of scientific contexts.

Mostly our team from Kansas State University visited those institutions that were located in the central region of the United States. During these visits, we carefully looked for the indications of inquiry-oriented strategies including the level of student engagement with the learning activity. In every site visit, we collected samples of instructional materials, conducted structured interviews with faculty members and in-service teachers and focus group interviews with pre-service undergraduates. In order to measure the level of reform, we observed the university faculty's and elementary teachers' classes and used the Reform Teaching Observation Protocol (Appendix-A). For each site-visit, we also wrote a comprehensive report on our visit. Substantial amount of data were collected by NSEUS research group and were distributed among collaborators of the project based on the parameters that were related to their research questions.

1.8.2 *Exploring the impact of reform on students' reasoning*

Educators involved in reform know that one of the most challenging aspects of teaching inquiry is assessing inquiry-oriented learning goals. Assessing inquiry is time consuming and different types of evidence are required to assess variety of goals. According to Marx et al., (2004), the evaluation of reform efforts was often limited to the data obtained through either teacher reports or students' performance on nationally normalized tests. However, those types of assessment are not designed to capture the implications of inquiry learning achievement goals. As such, they are not sensitive enough to detect the effects of reform (Porter & Smithson, 2001). According to Russ et al. (2006) a well-reasoned answer may end up with wrong conclusion. On the other hand, students may provide correct answers, but are not able to understand them or provide supportive arguments. As a result, an assessment of conceptual understanding that evaluates canonical knowledge is not necessarily appropriate for evaluating the quality of reasoning. This is mainly because the former tool is not designed to recognize the characteristics of a well-reasoned argument.

Therefore, this study attempted to classify different dimensions and levels of scientific reasoning for elementary education majors. To do this, we analyzed students' reasoning in the context of open-ended content questions.

1.8.3 *Research questions*

To align assessment with goals of inquiry, we carefully designed an inquiry-oriented assessment protocol. This process involved defining specific structures for developing content questions and providing descriptions of a coding scheme, sensitive enough to determine how well reasoned answers were as opposed to their correctness. Moreover, we defined parameters that controlled the level of abstraction and thought processes in our questions. To do this, we devised different questions in different disciplines requiring the same levels of cognitive processing, types of knowledge and structural format of the concept construction, namely concept category and type of concept links. In this way, the processes of thinking would be similar for the comparison groups although the contexts would be different. This was useful because often we needed to compare classes in two different disciplines.

Using our protocol, we addressed the following research questions;

Q1: How do we elicit students' reasoning using written content questions?

Q2: How do we classify students' reasoning in terms of the quality of their responses to written content questions?

Q3: What is the relation between the quality of students' thought processes and the degree to which course is considered to be reformed (in terms of RTOP measure)?

Q4: What is the relation between the conceptual structure of the responses and the degree to which course is considered to be reformed (in terms of RTOP measure)?

1.9 Implications and Summary

The overall aim of NSEUS was to investigate the quality of the reformed instruction, the implementation of inquiry and its impact on the different aspects and strands of learning. NSEUS has investigated a large number of subthemes that compare the achievements and quality of reformed and traditional courses. Therefore, it has involved several studies and various types of instruments, to probe each aspect of reform. The data have been collected through surveys, site-visits and through the administration of exam questions. The collaborators of the project received the collected data and analyzed the particular type of data related to their research questions.

In the remaining chapters, I discuss students' reasoning skills, assessment protocol, and our analysis of the results of our investigation. I specify the particular characteristics of a well-reasoned answer in our assessment. I also describe the protocol that we devised for designing questions that categorize characteristics of students' reasoning and define a rubric to detect these characteristics in the students' responses.

Chapter 2 - Literature Review

This chapter reviews literature related to this study. It begins by reviewing the contribution of psychology and constructivism to the theories of learning and teaching activities. The literature reviewed on constructivism in this chapter covers the different types of teaching practices including inquiry-based learning, cooperative learning, and project-based approaches. These teaching practices address the learner as an active participant in the learning environment. Constructivists believe learners should construct their own understanding and knowledge of the world through questioning, exploring and investigations.

The first section in this chapter summarizes several psychological approaches, learning models, and their implications for instructional planning. The second section reviews the debate concerning the meaning of inquiry, a fundamental concept informing the NSEUS project. In the literature we review, several issues emerge including the vagueness of the concept of inquiry that allows for different interpretations of the theory, which makes assessing class management and performance difficult. We also discuss another major theme in the literature concerning teachers' beliefs about the nature of science and inquiry methodology and its influence on the quality of their teaching.

The fourth section of this chapter briefly reviews types of transfer and compares the traditional views about transfer to more recent extended views of transfer. The issue of transfer is central to the exploration of a new phenomenon, and exploration is one of stages in the process of inquiry. Another implication of transfer to our study concerned the way our content questions were developed so that it requires application of newly learnt concepts in a new context.

In the closing sections of this chapter, we review other theoretical frameworks related to our methodology for analyzing students' scientific reasoning and developing science content assessments.

2.1 Constructivism

Piaget tried to sketch an integrated theory of child development with its biological basis of adaption (Piaget, 1961). He believed that each human being inherits a special style of intellectual functioning that interacts with the environment in order to develop the intellectual structure. Nevertheless, the fundamental process underlying the progressive development of human beings is identical and remains the same throughout life. Piaget stated that the development of intelligence is a dynamic process and is a consequence of adaption to the environment; he postulated that this adaption has a basic tendency toward the organization of thought (McNally, 1974).

In contrast with previous psychologists, Piaget regarded the child as an active participant of the process of intellectual development. Prior to Piaget, psychologists believed that the child's brain was similar to that of the adult in the way it functions. In terms of cognitive development, the child gained more knowledge commensurate with biological maturation and environmental learning. However, Piaget stressed the interactions of the child with the environment and the process that causes brain function to develop and evolve by the means of these interactions (Atkinson et al., 1993).

Piaget noticed that infants inherit natural reflexes and skills called sensory-motor skills, which allow them to react to the events and objects around them. In this way, an infant can interact with the world and construct exploratory theories regarding the result of his or her own experiences with environment. Piaget called these mini theories schemata (Atkinson et al., 1993) and expanded further on the thought processes that articulate the adaptation of the schemata (Hyde, 1970). According to Piaget, after gaining more knowledge, the child will construct a set of schemata and whenever s/he encounters a new situation, will attempt to fit the new information into existing schemes. If the new situation confirms the existing schemata, the child can perceive the phenomena in terms of the schemata that persist. On the other hand, if the schemata do not fit the new situation, the child will alter the existing schemata in response to the new stimuli. Piaget called the latter process assimilation and the former accommodation to describe schemata alteration (Piaget & Inhelder, 1969). When assimilation occurs, the new event fits the existing conception; in contrast, accommodation is the result of cognitive disequilibrium, when expectations are not met and events do not fit and there is tendency for existing beliefs to be weakened. In fact, assimilation and accommodation are two functions of adaption operating in two different directions.

Piaget believed that cognitive development is a process of constant *adaptation*, and *assimilation* of new experience to existing schemata and disequilibrium occurs when external reality does not match the existing schemata. Assimilating new information into pre-existing mental schemes and accommodating the new information to the altered schemata is denoted as an effort to maintain the balance that allows for cognitive development and effective development of thought processes. Equilibrium is the fundamental principle underlying Piaget's model. Maturation, for Piaget, involved reaching higher states of equilibrium between assimilation and accommodation.

2.1.1 *Piaget's Stages of Intellectual Developments*

Piaget observed that the child's cognitive structure develops and grows through a series of distinct stages. He categorized the cognitive development into the following four stages: Sensori-motor (birth to 2 years), Pre-operational (2-7 years), Concrete operational (7-11 years) and Formal operational (11 years on). In the first stage, the child learns to differentiate the self from objects and recognize the self as an agent of an action. In the next stage, the child develops skills such as learning language that involve classifying objects by features and colors. Ideas related to the formal learning of sciences do not really become important until the latter two stages. These two stages involve thinking logically and finding relationships between concepts and schemes.

Piaget regarded the concrete operational stage as a critical stage in the child's cognitive development because it marks the beginning of logical or operational thought processes. However, this logical thought is related to things that can be touched and sensed. During the years from about age 11 years onwards the individual's ability to understand systemically in order to analyze a situation and to test hypothesis develops steadily. In a sense, the ability to handle abstract notions develops at this stage.

2.1.2 *Comments on Piaget's Intellectual Development*

Piaget's ideas are now broadly accepted in general terms. Firstly, he demonstrated that, in cognitive terms, children are not miniature adults. There are specific ways in which their learning is different. Secondly, he offered a description of the stages through which children process as they move towards cognitive maturity. Piaget's work is a coherent set of observations but is not a theory. His description does not explain why people go through these stages. It does not refer to the development of memory with age or other factors. Piaget's work has been criticized for its definition of a set of well-defined age related boundaries with sudden transitions from one stage to the next (Ausubel et al., 1978). Later work has shown that people do not jump from stage to stage in neat ways. Indeed, a person can be operating at one stage in one context and at another stage in a different context (Novak, 1978; Jenkin, 1978; Dawson, 1978).

Renner and Lawson (1973) looked at the level of students' thought processes in relation to students' age and content of study. They administered tasks that can identify formal thoughts in the high school physics students. The results showed that the large number of the adolescent population were not in the formal operational stage. Renner and Lawson Suggested that by providing them inquiry-oriented science experiences the level of thought processes can be improved for high school students and Freshman College.

Heron (1975) looked at cognitive stages of first-year undergraduates. He found that not all students were at Piaget's top stage. He concluded that individuals could not learn to think formally unless they were required to make an effort to develop formal thinking skills. In simple terms, it is not enough to reach a certain age to be able to function at this stage. According to this explanation, the reason some students are not completely operating at formal stage is that they were never required to operate at this level.

2.2 Social Constructivism

Vygotsky (1978) advocated an approach that was very different from Piaget as he observed the process of knowledge construction in a social context. He regarded the process of learning as being more complicated than Piaget's stages and as being a process that should be analyzed. Vygotsky emphasized both individual and social aspects of learning and found that social interactions and supportive learning environments actually leads to cognitive development.

On this basis, Vygotsky postulated the *Zone of Proximal Development (ZPD)* idea. This refers to actual level of improvement as determined by independent problem solving [without guided instruction] in comparison to the level of development as determined by problem solving under the guidance of more knowledgeable person. One can measure the interval between the two levels by comparing the student's performance on both tasks. These results have significant implications for teaching and learning. For example, if we measure a learner’s ZPD for a particular skill, we can predict how that learner will independently utilize that skill in the near future. Therefore, zone of proximal development is the optimum range for effective learning and designing instructional materials. If the level of instruction is above the proximal level, learning will not occur or will be incomprehensible; on the other hand, if the level of instruction is below this region, it will not be challenging. As learning improves, students need less assistance and are able to solve more of the problems independently, but they may need help on more challenging tasks. Therefore, the ZPD continues to change as the novices develop their abilities.

Using Vygotsykian framework, Bruner (1966) suggested the idea of scaffolding that is providing right amount of support to a learner to accomplish a task. According to Bruner, there is a way to communicate ideas to children that is appropriate to the particular age and it is futile educationally simply to wait passively for the children to grow in readiness. In addition, Shayer and Adey (2002) proposed the idea of cognitive acceleration while they were trying to find a way to help students to go through the stages faster. They reported that some acceleration (in Piagetian terms) occurred because of using their effective set of teaching materials and strategies. However, these materials and strategies may not be beneficial to all students and their potential for boosting acceleration may be limited. Both of these approaches favored Vygotsky's (1978) developmental theory.

2.3 Inquiry Oriented Teaching

One of the most discussed topics in science education is the idea of teaching by inquiry. The theoretical and philosophical rationale underlying inquiry teaching is constructivism. The studies of Vygotsky (1978) and Paiget (1964) have been enormously influential in the formation of interactive methodologies and idea of inquiry. For example, inquiry oriented teaching can benefit from Vygotsky's ideas of scaffolding. To encourage students to think on their own, below the point of quitting, teachers withdraw the scaffolding while keeping the students in their ZPD level (Bandura, 1986). The inquiry-oriented method of teaching often employs student centered and constructivist strategies of learning. Fostering Socratic dialogue and exposing students to cognitive dissonance are examples of the teaching methods that can be used in a constructivist approach. These techniques have evolved from Piaget's theories of assimilation and accommodation and have appeared to be effective to help learners to bring out misconceptions and faulty reasoning (Hake, 1992).

The history of the inquiry also goes back to John Dewey (1910), who proposed shifting the emphasis in teaching from teaching aimed at helping students accumulate knowledge toward teaching aimed at helping students acquire methods of thinking and a new attitude of mind. Dewey (1938) considered the scientific method of thinking as comprised of four stages: induction, deduction, mathematical logic and empiricism. Dewey's ideas served as a cornerstone in developing guided discovery or inquiry approach. Later on inquiry instruction received attention in curriculum design and a large number of inquiry-oriented curricula were developed.

Further studies set out to determine the nature of inquiry and important aspects of teaching inquiry. The National Science Education Standards (NSES) established a set of standards for K-12 educational system to promote inquiry teaching and prepare people to be scientifically literate citizens (National Research Council, 1996). The standards targeted areas such as: instruction, assessment, professional development and curriculum. Ebenzer et al. (2011) categorizes the content of NSES and inquiry teaching based on three main themes of learning practices that address the following:

- Scientific conceptualization
- Scientific investigation
- Scientific communication

Scientific conceptualization

According to Ebenezer et al. (2011), knowledge scaffolding can foster scientific conceptualization that can be achieved through: eliciting ideas, revising work while completing activity, using simulations that enhance modeling and visualization. Ebenezer et al. also believe that concept conceptualization can be achieved in other ways such as using online concept maps to find relational links between the concepts and using virtual models to learn concepts that are invisible and abstract. Many interactive computer simulation models have been developed to facilitate visualization of the abstract concepts and to engage students in active inquiry, which promotes conceptual understating

(Finkelstein et al., 2005). In a simulation environment, students are engaged in active inquiry; by arranging, a virtual set up, and changing variables, students can observe the effect of their actions and evaluate their predictions and reasoning (Jaakkola et al., 2011).

Scientific Investigation

Inquiry standards (NRC, 1996) highlighted abilities concerned with procedures of inquiry conducing scientific investigation. These abilities involve question making and testing hypothesis, using instruments, using mathematical and statistical tools, designing an experiment and explaining the results. Klapper (1995) suggested that the process of questioning should be emphasized more at the outset than hypothesis making. This approach, Klapper claims, is closer to what happens in science research. Therefore, students should develop questioning skills to pose appropriate research questions and direct an investigative approach to draw conclusions.

Schwab (1958) argued that the methods of teaching science should be consistent with the way that modern science research has naturally progressed. Schwab suggests: "The nature of scientific inquiry now controls research. Science is no longer a set of truths for verification but a revisable structure of knowledge that is under continuous assessment and modification."

Scientific Communication

Scientific communication is the ability to communicate research and participate in ongoing classrooms discussions (Ebenzer et al., 2011). Scientific communication also includes explanation and the ability to articulate reasons to justify conceptual understating. Through critical responses or doubtful questions from peers or experts, a deeper conceptual understanding may emerge. Aufschnaiter et al. (2008) examined the progression of argumentation capabilities for the junior high school students. They asserted that developing argumentation in a scientific context requires very specific knowledge of the phenomena and students can engage in argumentation only if the content of argumentation and its level of abstraction are familiar to them. Aufschnaiter, et al. stated the ability to undertake higher levels argumentation depends whether the students can relate the content of argumentation to their prior knowledge.

2.3.1 *Scientific method and scientific inquiry*

Finley and Pocovi (2000) distinguished between the traditional notions of the scientific method and scientific inquiry. Traditionally, scientific method was taught as the different steps involved in recognizing the research problem, forming a hypothesis, conducting experiments, collecting and analyzing data, and presenting the final theory, which can be tested again by other scientists. These steps seem like a prescription for a kind of thinking that is learnable and better suited to the time constraints of teaching in the classroom. However, Finley and Pocovi claimed that scientists necessarily do not always follow those steps, and the research strategy they choose depends upon the topic or problem under investigation. Consequently, Finley and Pocovi (2000) sought instances from the history of science to verify their claim. For example, they noted that Roentgen observed that x-rays accidentally passed through a shielded tube while he was exploring cathode rays and Faraday discovered benzene's structure from an oily film that was deposited from the gas used for lighting. Therefore, the traditional view of the scientific method's formulated set of procedures does not always comport with the way in which scientific discoveries are made. Many scientific discoveries have been made without following steps in the scientific method, and experiments are not always the core of research. Some problems are empirical while some others are conceptual and theoretical; for example, black holes were first predicated theoretically. In fact, sometimes ideas have emerged from the theories and were tested only later in the natural world. Students, should therefore, learn that science involves more than just experimentation.

Another problem is that the performance of the scientific method's steps can be influenced by the prior theories or biases that students or researchers have acquired about the natural world. Senses and filters of prior knowledge can limit the way one chooses to observe and interpret the observations. Consequently, students should learn to discuss the limitations that experiments impose on solving scientific problems and consider that science is a human effort and no aspect of the scientific method is immune to the human limitations.

An unfortunate consequence of the mindset imposed by the dogma attached to the scientific method is that the outcome of an experiment is misconstrued frequently as an ultimate truth about nature. Quite to the contrary, however, the revisionary character of inquiry necessitates continual assessment and modification. Finley and Pocovi (2000) noted that scientific inquiry is an unending process and new ideas should be the starting points for new discussions, replications and modifications. The falsifying or proving of a hypothesis by deduction is not sufficient, because, existence of contradictory evidence is always possible and ambiguity in data can influence the assumptions scientists make. Therefore, student should have in mind that scientific truth is not absolute but a collective reasoning of observed phenomena, previous theories and research.

Finley also noted that scientific problems can be either theoretical or empirical, and each requires its own strategies and resolutions. This means solving some problems requires observation and experimentation while the solutions to other problems depend on theories and revising concepts about nature, and students need to distinguish between these two.

2.3.2 *Inquiry in action - Sources of difficulties*

Many times research has shown that learning outcomes are in the favor of inquiry oriented methods (e.g. Shymansky,1984), but in reviewing of teaching methods, Hurd et al., (1980) reported a negligible evidence of using inquiry oriented teaching methods in most of the biology classrooms. Based on these findings, Costenson and Lawson (1986), conducted interviews with experienced teachers to discover why they were hesitant to use inquiry. Costenson and Lawson reported common critics such as time and energy, slow speed of the inquiry, interfering with previous teaching habits, immaturity of the students etc. Most of the teachers complained about finding the time for developing materials that challenge students and keep them enthusiastic and interested. Teachers need to comply with school district curricula and cover prescribed amount of material, and the inquiry method by its nature decreases the pace with which the prescribed topics in science can be covered. The risk of implementing inquiry-based learning in science is often perceived as too high for administrators who are required to evaluate teachers' performance based on outcomes. An additional obstacle to the implementation of inquiry-based teaching practices is the immaturity of students that prevents them from coping with the method's higher cognitive demands. Teachers also resist changing their teaching habits and relinquishing control of what is happening in the class. Shortages of resources and required materials, pose another difficulty to the practice of inquiry-based science learning. Using inquiry means exchanging the coverage of material with developing inquiry and meta-cognitive skills that can be used in different content areas.

In addition, Costenson and Lawson (1986) reported the reading difficulties that students had shown in inquiry classes. They speculate that these difficulties are due to the higher reasoning demands that exist in reading inquiry rather than placing demands on students' memorization ability. They also criticized textbooks for representing a distorted view of scientific investigation leaving students confused with a little understanding how to participate in the process. For example, "Mendel's Law of Segregation" was a hypothesis, which was introduced as a law in the scientific textbooks. Costenson and Lawson observed that the simplification of the analysis of data often summarized as scientific laws that distorted the aim of his inquiry. Instead, they argued the textbook should have described the steps in Mendel's reasoning that led from data to the hypothesis and conclusion.

2.3.3 *Multiple perspectives teaching inquiry*

Inquiry learning and its implementation have been interpreted in a variety ways. Bybee (2000) compared three classrooms that implemented three different strategies based on inquiry. In the first class, students were asked to examine an event (the changing water level in an open container) by designing a simple investigation. This approach toward inquiry aimed to improve their ability to reason based on evidence. The second class was guided through series of questions and observations in which the students compared two populations of fossils. This approach toward inquiry aimed to facilitate discussions, which concerned types of the fossils and how they were formed, genetic variability and evolution. In the third class, students reviewed stories about scientists and discuss their approach toward scientific method of investigation. These examples may clarify how teachers may adopt different styles toward inquiry.

After presenting these three examples, Bybee encouraged readers to think about their perceptions of inquiry, the strategies and activities that they may use to promote inquiry. One may think implementing inquiry is conducting experiments in laboratories. Another may see inquiry as developing higher order thinking levels and performance skills. He concluded that teaching by inquiry may have multiple meanings and teachers should evaluate their perception of inquiry, their approach and the outcome.

Implementing inquiry and not only includes teachers and students, but also includes, administrators, and curriculum developers.

2.4 Transfer

Transfer is the process of applying knowledge learned previously to a new situation. The degrees of similarity between the two situations can differ and, therefore, the ease of transfer can be affected by the degree of similarity or difference between the situations.

Barnett and Ceci (2002) developed taxonomy with nine-dimensions of transfer. In their taxonomy, the subject of transfer could be classified in terms of the type of required learning skill, the style of performing the learning skill and the cognitive demands that are required in the new situation. They defined three types of learning skills (Procedure, Representation, and Principle), three styles of performance (Speed, Accuracy and Approach) and three levels of cognitive demands (Execute, Recognize and Execute, Recall and recognize and Execute).

According to Barnett and Ceci, among three types of learning skills, (Procedure, Representation, and Principle), transfer would be easier if it involved a specific routine procedure rather than a more general principle or strategy. For example, solving algebraic equations is easier than applying "control of variables strategy".

The second part of their taxonomy classifies the degrees of similarities between two contexts along six dimensions. The six dimensions are the *knowledge domain, physical context, temporal context, functional context, social context* and *modality*. The *knowledge domain* refers to the subject matter as in transferring knowledge from algebra to physics. The *physical context* refers to any change in the environment of the learner such as changing classes, teachers, or settings such as everyday life, classroom or interview settings. The *temporal context* refers to the time delay that occurs between the training and the transfer; for example, whether the evaluation of the transferred knowledge occurs right after the training or sometime later. *Functional context* here refers to the purpose of attempting to work on a task, whether it is for an academic, financial, work or play purpose. The fifth dimension, *social context*, means whether a task is performed individually, in a pair, or in a group. Collaborative learning can influence the transfer, therefore, *Social context* is one of constituent components involved in transfer (Druckman & Bjork, 1994). The final dimension is *modalit*y; for example, changing the format of a question from an essay or short answer type question to a multiple choice type question. Other examples of modalities that can affect transfer include oral exams or experimentation. For each of the six dimensions above, Barnett and Ceci considered different levels of similarities.

Other researchers (e.g. Mayer, 1975) distinguished between near and far transfer. The near transfer refers to the events involving greater degrees of similarity, whereas the far transfer involves situations that are very much different from each other. With respect to each of the dimensions -- *Knowledge domain, physical context, temporal context, functional context, social context* and *modality* -- we can consider situations that are more similar or similar.

More recently, researchers have begun thinking about cognitive theory to describe transfer (Royer et al., 2005). They have attempted to define a set of conceptual tools that could be used to further explore the issue of transfer. Long-term memory can be regarded as a network consisting of conceptual representations or semantic representations. Semantic representations are structures involving nodes and links between nodes. The nodes represent objects or concepts and the links represent relations between nodes. Accordingly, if a conceptual similarity between two situations exists, the network with the conceptual knowledge would be activated.

Traditionally, researchers have designed pairs of problems whose solutions involve performing tasks with some degree of commonality and look for evidence of transfer. For example, they considered pairs of tasks with different surface features and structural similarities between underlying concepts and the evidence for transfer was the recognition of the underlying conceptual structure in a new context.

Studies based on such traditional views of transfer showed that transfer rarely occurred (Rebello et al., 2005). To account for the lack of transfer, researchers then reconsidered the definition of transfer. The traditional view was researcher-centered and focused on the expert's definition of transfer and looked for implications that were imposed by the researcher. In this approach researchers failed to detect what students were actually attempting to transfer (Royer et al., 2005). Recently, however, those who have reconsidered the way the earlier research had specified transfer have challenged the work of earlier researchers. In this approach, (Beach, 1999; Bransford & Schwartz, 1999; Dyson, 1999; Greeno, Smith, & Moore, 1993; Lave, 1988; Lave & Wenger, 1991) the research question shifted from "What is transferred" to "How students apply and activate prior knowledge".

Lobato (2003), compared these two theoretical perspectives, summarized the distinctions and proposed an actor-oriented model of transfer. Her research design emphasized recognition of the similarities between two circumstances, rather than the transferring of knowledge from one situation to another. In that way, the perspective shifted from the expert-centered perspective to the learner's perspective. In this newer view, the concept of "dynamic production of sameness" has replaced the concept of "static application of knowledge". Using this view, researchers are attempting to explore how students recognize similarities between two situations, or to find how they connect one situation to another. Consequently, the definition of transfer has changed from "The application of knowledge learned in one situation to a new situation" to "The personal construction of relations of similarity across activities" (Lobato, 2003).

2.4.1 *Dynamic transfer*

Rebello et al. (2005), elaborated on the expanded view of transfer to provide a framework for dynamic transfer. Specifically, they looked at the common themes that emerged from the perspectives of Bransford and Schwartz (1999), Greeno et al. (1993) and Laboto (2003), all of whom advocated on expanded view of transfer. Based on the common themes that emerged from the contemporary research on transfer, Rebello et al. (2005) developed a framework to analyze students' responses in an interview. Their model of dynamic transfer consists of external inputs that are provided by interviewer and interview materials. In addition, the model considers the associations that occur between prior knowledge (*Source tools*) and pieces of knowledge that learners obtain from the input or new situation (*Target tools*). Another component of the framework is called *(Workbench)* which includes the dynamic mental processes that help learner to associate the source and target tool. However, learners can select part of the input information (*Readouts*) that they find relevant. Association can be controlled by factors such as motivation and epistemological beliefs. Based on the external input including messages from the interviewer, a learner may activate particular epistemic mode. For instance, a student who believes that knowledge comes from authority more likely transfer knowledge that s/he acquired from 'authoritative' sources such as textbook or an instructor, rather than from life experiences social construction of knowledge. Epistemic resources can control the type of tool that students use.

2.5 Bloom's Taxonomy

Educational taxonomies were originally formulated to classify outcomes for different domains of learning. Furthermore, they were used for categorizing instructional objectives and evaluation plans. Bloom et al. (1956) were among the pioneers of classifying educational goals. Among learning taxonomies, Bloom's Taxonomy is the best-known example of one that characterizes the elements of understanding in a scaffold hierarchy. The components of the taxonomy progress from simplest to most sophisticated level of cognitive reasoning. Bloom's taxonomy was first created for organizing instructional objectives. Teachers can use Bloom's framework as a template for assessment, defining teaching objectives, lesson planning, and curriculum development. Bloom's concern was the promotion of higher forms of thinking. He believed that most teaching tended to be in the lowest level of training that is information recall. Consequently, Bloom sought to articulate the cognitive domains of learning (Bloom et al., 1956). His hierarchy consists of six levels within the cognitive domain ranging from the lowest level, which is recall or recognition of facts, more complex and abstract levels of thinking and the highest order, evaluation. Table 2-1 shows Bloom's original categorization of cognitive domains. Further descriptions for each component are discussed below.

Table 2-1 Bloom's cognitive learning taxonomy

Bloom's cognitive learning taxonomy:
Knowledge Comprehension Application Analysis Synthesis Evaluation

Knowledge

Knowledge concerns the remembering of previously learned material including facts, specific information, theories, concepts, and even principles and methods. The verbs that can describe learning outcomes for knowledge are: define, describe, identify, label, list, name, recall, reproduce, select and state.

Comprehension

Comprehension concerns the ability to grasp the meaning of something. The evidence for comprehension could be changing representation, for example interpreting meaning from a table of data, a chart, or a graph, or paraphrasing verbal into material to mathematical formats or justifying procedures. This learning outcome is a higher level than remembering but still represents the lowest level of understanding. The verbs that can be used to describe learning outcomes for comprehension are convert, defend, distinguish, estimate, explain, infer, paraphrase, predict, rewrite, and summarize.

Application

Application refers to use of a learned material in a new context. This can involve using a concept, principle, rule or method or involve executing a procedure. By applying physical laws and theories to practical situations, learners show that they can solve the real-world problems. Application represents a higher level of learning in Bloom's taxonomy. The verbs that can be used to express leaning outcomes for application include: change, compute, demonstrate, discover, manipulate, prepare, produce, relate, show, and use.

Analysis

Analysis refers to the ability to divide a structure into its constituent pieces, study the components and their interrelationships, and identify the principles that informs those relationships. Analysis represents a higher level of understating than comprehension and application because it requires the understanding of both content and structural relationships. The verbs that can be used to express learning outcomes include: break down, diagram, differentiate, discriminate, distinguish, outline, point out, relate select, separate, and subdivide.

Synthesis

Synthesis is the ability to combine disparate ideas and find coherency among them. In another words, synthesis involves generalizing from a set of abstract relationships. This may require unifying, integrating or formulating new patterns, or planning a new set of operations. The execution of synthesis may require developing a well-organized theme or a well-organized speech as well as integrating knowledge and learning skills from different areas into a strategy for solving new problems and formulating or developing new schemes for classifying knowledge. The verbs that can be used to express learning outcomes for synthesis include: compile, compose, create, devise, design, generate, modify, organize, plan, rearrange, reconstruct, relate, revise, rewrite, and write.

Evaluation

Evaluation refers to being able to determine or estimate the value of a material for the purpose for which it created. In other words, evaluation involves accepting or rejecting something by making judgments about the validity and quality of the object of the evaluation based on a set of criteria. Evaluation is the highest cognitive domain in Bloom's hierarchy, as it requires the integration of all the components of the taxonomy. The verbs that can be used to express learning outcomes for evaluation are*:* prove, criticize, decide, assess, disprove, estimate and dispute.

2.5.1 *Bloom's taxonomy (affective domain and psychomotor domain)*

Later, Krathwhol et al. (1964) added another domain to Bloom's taxonomy, which they called the affective domain. This taxonomy classified the development of feelings, values, appreciation, enthusiasms, self-awareness, motivations and attitudes. They defined the five main categories covering the development of affective domain from the simplest to the sophisticated is shown in Table (2-2) below:

Table 2-2 Bloom's affective learning taxonomy

Bloom's affective learning taxonomy:
Receiving
Responding
Valuing
Organizing
Internalizing

In the lowest level, "Receiving", students passively listen to others or remember an experience; however, in the next level they react in a certain way, for example by participating in the class discussions or asking questions. In the third level, students attach a value to an event, object or behavior. By accepting responsibility or valuing ethical issues, students can progress to the higher levels of taxonomy. In the highest level of the affective domain, students develop a value system that controls their behavior. They are more confident, self-dependent, and ready to work individually.

Others also have tried to expand on the Bloom's taxonomy. Harrow et al., (1972) for example built on Bloom's taxonomy in other domains to categorize psychomotor skills. Their taxonomy starts with the low levels of reflex and basic fundamental movements, while other physical activities and skilled movement make up the highest levels.

2.6 Revised Bloom's Taxonomy

After 40 years of development of a growing body of literature on meaningful understanding a revision of Bloom's original taxonomy (for the cognitive domain) to incorporate research outcomes into this framework (Anderson et. al., 2001) was appropriate. Recently, research has shifted the attention more toward a knowledge construction view of learning, the focus has changed from knowledge acquisition toward cognitive engagement that is active cognitive processing, organizing and integrating incoming knowledge with the previous knowledge (Mayer, Bransford, 1999).

Because of this shift of focus and numerous research outcomes, teachers are confronted with exceedingly large number of objectives that are stated in the content standards and curriculum standards. Therefore, a learning taxonomy was required to serve as a bridge between standards, objectives and teaching plans. Making objectives more contextualized can help teachers to understand the objectives more specifically, which leads to more appropriate instructional designs. Consequently, a group of educational experts revisited the taxonomy. They included the representatives of three groups: cognitive psychologists, curriculum instruction theorists, and assessments specialists. Their revised version, added a knowledge dimension to the Bloom's taxonomy. The knowledge dimension included four types of knowledge: factual, conceptual, procedural, and meta-cognition. This hierarchy proceeds from the lowest level of factual knowledge to more complex and abstract levels of conceptual and procedural knowledge.

The authors of the revised taxonomy sought to be consistent with the terminology of contemporary educational psychology; therefore, they changed some of the terminology that previously was used in the original taxonomy. Based on an extensive review of the objectives including the original Bloom's taxonomy and also more updated achievements of research (e.g. Stenberg, 1998), Anderson et al., selected 19 cognitive processes to provide a broader and more specific category of objectives in compare to the original Bloom's taxonomy. The cognitive process of 'Remember' promotes retention, whereas, other cognitive processes are associated with transfer of learning. Because in the current research our goal is to foster transfer; therefore we should devise assessments that involve cognitive processes beyond 'Remember'.

In the revised version of the taxonomy, 'Create' is associated with the highest level of cognitive processing. Unlike other categories, which involve cognitive processing with given sets of elements, in 'Create', students should put many resources together and generate a novel pattern in compare to their prior knowledge (Anderson et al., 2001).

As one can see from Table 3, the intersections of the six cognitive processes (Remember, Understand, Apply, Analyze, Evaluate, and Create) and the four knowledge dimensions (Factual, Conceptual, Procedural, and Meta-Cognitive) form a grid with twenty-four separate cells. Curriculum planners or instructors could use this grid to organize the objectives of a lesson plan. To do this, they should identify their objectives and the thought processes they considered in that lesson plan. The relevant thought processes can be identified on Anderson's grid by selecting the row for the relevant knowledge type and the column of the relevant cognitive type and noting their intersection. For example, the grey cells in Table 3 represent two tasks with slightly different levels of thought processing. One task requires applying conceptual knowledge to a new situation and another requires understanding some type of procedural knowledge.

Table 2-3 Anderson Grid (Blooms revised two dimensional taxonomy)

Revised Taxonomy (Anderson & Krathwoll, 2001):						
The Knowledge Dimension	**The Cognitive Process Dimension**					
	Remember	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Create**
Factual Knowledge						
Conceptual Knowledge						
Procedural Knowledge						
Meta-Cognitive Knowledge						

2.6.1 *Definitions for knowledge dimension categories*

Although knowledge is domain specific, knowledge from different disciplines shares some commonalities. Anderson, et al. classified knowledge into different types, based on the features that were similar in every domain.

Many researchers before Anderson, distinguished different types of knowledge such as; content knowledge, conditional knowledge, explicit knowledge, strategic knowledge, schematic knowledge, semantic knowledge, prior knowledge, domain knowledge, and tacit knowledge. Anderson's four categories of knowledge not only capture all these definitions but also is simple and practical. These four categories are *Factual knowledge, Conceptual knowledge, and Procedural knowledge* and *Meta cognitive knowledge*.

Subtypes are associated with each type of knowledge. The definitions for knowledge types and their associated subtypes are described below.

Factual Knowledge

In the revisited taxonomy, factual knowledge is the lowest level of knowledge and consists of basic elements students must know to be acquainted with a discipline or solve a problem. In other words, this kind of knowledge includes bits of information or informative parts of scenarios that are similar to declarative knowledge. This type of knowledge is about "knowing what" and "knowing that." The subcategories of factual knowledge include knowledge of terminology and knowledge of specific details, such as a list of information pertinent to a subject based on real occurrences, including statements about events or something that has demonstrated or has existed.

Conceptual Knowledge

Learning every subject matter starts with collecting facts or basic information pertinent to that subject. The next level is attributing relevant properties to the facts and discovering the interrelationships that exist among the facts. Anderson (2001) illustrates this with an example of seasonal changes and shows how the bits of knowledge become interrelated to explain a phenomenon. For example, by applying the concept, that earth rotates around the sun and earth has a tilted angle relative to the plane of its orbit, learners can understand the seasonal changes by constructing relationships among these three concepts.

- <u>Conceptual Knowledge/Knowledge of Classifications and Categories</u>

One of the subtypes of conceptual knowledge is knowledge of *classification and categories*. This means knowing that facts with similar properties can be classified into specific groups and groups can be combined into categories based on more general features to make larger groups. The difference between experts and novices is mainly in the way they classify and organize the knowledge. Novices hold pieces of disconnected information; however, experts have a coherent organized structure of conceptual schema. Therefore, when they perform problem-solving tasks, novices try to find the piece of information that matches the problem; in contrast, the expert thinks of the underlying principle developed from the interconnected concepts (Bransford et al., 1999).

Knowledge of classification explains some of the misconceptions related developing students' conceptual schema. These deficiencies are due to the novices' misclassification of the concepts by incorrectly associating particular features with the concepts. For example, novices may incorrectly categorize processes in physics, such as heat or current as substances, and cognitively it is not easy for them to re-categorize these concepts (Chi & Slotta, 2006). In fact, misclassification explains a large number of misconceptions and difficulties students have in physics. Treating vectors as if they were scalars or dealing with conservative forces, as if they were non-conservative, or describing a phenomenon applying classical theory, while it should be described using the quantum realm are common examples of misclassification.

Classification of knowledge requires being able to identify a group of facts and concepts with certain properties and associate those facts and concepts with certain methods, features and strategies. At higher levels of thinking, the cluster of concepts are then generalized into knowledge of principles and unified into theories. Being able to organize and classify knowledge correctly is a skill that distinguishes the novice from the expert. Anderson et al., defined three subcategories for conceptual knowledge including;

knowledge of classification and categories, knowledge of principles and knowledge of theories.

- Conceptual Knowledge/Knowledge of Principles

Knowledge of principles is generalized from knowledge of classification and categories; having this knowledge means being able to discern patterns among different categories and classes. For example, the principles of the conservation of energy and momentum are patterns that have been observed in many systems and interactions that occur in them.

- Conceptual Knowledge/Knowledge of Theories

The organization of ideas, classified knowledge and principals can be formulated in a way to generate a theory to describe a complex phenomenon. The theory of evolution in biology or the theory of special relativity are examples of knowledge of theories.

Procedural Knowledge

Procedural knowledge is knowledge of how to perform a task or a sequence and set of steps. Some examples of this type of knowledge include techniques and subject-specific skills or knowing the rules and methods related to a certain theory or principle. Every domain requires specific procedural knowledge. For example, mathematical procedures, such as vector algebra or solving equations, are different from procedures required for computer programming or for solving problems in Newtonian mechanics by applying Newtonian laws and drawing free body diagrams.

Meta-Cognitive Knowledge

Meta- cognitive knowledge is an awareness of one's own cognition. Students with meta-cognitive abilities are more responsible for their knowledge and thought processes. Anderson et al., emphasized the distinction between knowledge of cognition and knowledge of monitoring, controlling and regulating the cognition. In agreement with Flavell (1979), Anderson et al., discussed three subtypes for the meta-cognitive knowledge including; strategic knowledge, conditional knowledge and self-knowledge.

- <u>Meta-Cognitive Knowledge/Strategic knowledge</u>

Strategic knowledge consists of variety of learning strategies for learning, problem solving, comprehending meaning from a text or tactics for understating better in a classroom. Examples of strategic knowledge are strategies for memorizing for better retaining information, strategies for organizing such as drawing diagrams or outlining and elaborative strategies such as summarizing and paraphrasing.

- <u>Meta-Cognitive Knowledge/ Contextual and conditional knowledge</u>

Another component of meta-cognitive knowledge is conditional knowledge, which can be viewed as knowledge of context difficulty with respect to the memory demands associated with a particular context. In other words, conditional knowledge is the skill of using appropriate learning strategies in different situation and knowing when and why to use different cognitive tools.

- <u>Meta-Cognitive Knowledge/ Self knowledge</u>

Self-knowledge is awareness of one's strengths and weaknesses about cognitive or learning abilities. In addition, having self-knowledge means to be aware of one's goals in motivations in performing a task as well as personal interests.

2.6.2 *Definitions for cognitive processes dimension categories*

Recall

The cognitive process of recall simply involves the ability to store and retain the information from the memory. At this level of reasoning, students retrieve the stored information to produce definitions, facts, or lists, or recite material.

Understand

- Understand/ Interpret

In the context of questions related to science, to interpret means being able to paraphrase the given information into scientific terms, symbols, rules or equations. It also means being able to change representations of data from one format to another, as in the case of being able to derive information from graphs, images, chemical reactions, equations, and transforming that to another form such as verbal information or vice versa.

- Understand/Classify

The cognitive process of classifying should not be confused with the knowledge of classification. In this process, the emphasis is on how to classify or group a set of data by recognizing common features.

- Understand/Infer

Being able to infer means being able to recognize a relationship within a set of events, as in the case of being able to recognize causes and effects. This process involves making comparisons between events, and identifying a pattern that can be used to predict an outcome.

- Understand/Compare and Contrast

Being able to compare and contrast means being able to define or ascertain the commonalities or changes among two or more concepts, models, situations, or methods. It involves determination of the similarities and differences between objects, events, ideas, characteristics, structure, etc. Comparing can facilitate the processes of inferring and reasoning. For example, in collisions, we compare the situations before and after the interaction, or in genetics, we compare the products of mutations to the products of the original genes using a model derived from theories and initial assumptions.

- Understand/Explanation

Being able to present an explanation, means being able to perform the process of showing the explicitness of one's thoughts. That means presenting a well-supported argument, one that is fully justified and coherent and not a sketchy collection of fragments of one's thinking. Such an argument must be a cohesive set of statements with subtle connections between assumptions and theories and all types of knowledge so that all the different parts of the argument lead coherently to a conclusion. Such an argument may also involve making inferences from observations. Finally, a well-developed explanation in science might require constructing a chain of cause and effects for major events.

Apply

The knowledge domain signified by the term "apply" is used when a procedure is performed in a given task. Anderson (Table 2-4) distinguished between two types of Apply namely Execute and Implement. The latter refers to applying familiar concepts in a new context, and the former to applying a procedure in unfamiliar task. Because familiar concepts can be hidden in new contexts, students need to have skills to ascertain or discern the relevant concepts.

Analyze

Analyze involves finding structural relationship of the elements and constituents parts. This cognitive category includes three cognitive processes of ‘Differentiate’, ‘Organize’ and ‘Attribute’. The subcategory of ‘Differentiate’ is to divide a process to its major steps or to distinguish the constituents and elements of a whole process in order to recognize the relevant and irrelevant parts of a subject. The subcategory of ‘Organize’ usually occurs in conjunction with ‘Differentiate’ in which student should differentiate the relevant and irrelevant parts of a structure and then to find the systematic, coherent relationship between the constituents in which structure fits. The cognitive process of ‘Attribute’ occurs when a student is able to ascertain the underlying intention of a communication.

Evaluate

‘Evaluate’ includes making judgments based on criteria and standards. The criteria that most often are used include; quality, effectiveness, efficiency and consistency.

Create

The cognitive process of ‘Create’ starts with a divergent phase that a student pulls out information from different resources and follows a convergent phase to put together the information to generate a coherent functional product. Unlike other cognitive processes, ‘Create’ requires drawing information from various sources, whereas, other cognitive processes involve the given information. ‘Create’ has the essence of divergent thinking that is to consider alternative possibilities and various thinking.

Table 2-4 Knowledge Dimensions of the Revised Bloom's Taxonomy

1. Factual Knowledge	The basic elements that students must know to be acquainted with a discipline or solve problems in it. Aa. Knowledge of terminology Ab. Knowledge of specific details and elements
2. Conceptual Knowledge	The interrelationships among the basic elements within a larger structure that enable them to function together. Ba. Knowledge of classifications and categories Bb. Knowledge of principles and generalizations Bc. Knowledge of theories, models, and structures
3. Procedural Knowledge	How to do something; methods of inquiry, and criteria for using skills, algorithms, techniques, and methods. Ca. Knowledge of subject-specific skills and algorithms Cb. Knowledge of subject-specific techniques and methods Cc. Knowledge of criteria for determining when to use appropriate procedures
4. Meta cognitive Knowledge	Knowledge of cognition in general as well as awareness and knowledge of one's own cognition. Da. Strategic knowledge Db. Knowledge about cognitive tasks, including appropriate contextual and conditional knowledge Dc. Self-knowledge

Table 2-5 Cognitive Dimensions of the Revised Bloom's Taxonomy

1. Remember	Retrieving relevant knowledge from long-term memory. 1.1 Recognize 1.2 Recall
2. Understand	Determining the meaning of instructional messages, including oral, written, and graphic communication. 2.1 Interpret 2.2 Exemplify 2.3 Classify 2.4 Summarize 2.5 Infer 2.6 Compare 2.7 Explain
3. Apply	Carrying out or using a procedure in a given situation. 3.1 Execute 3.2 Implement
4. Analyze	Analyze--Breaking material into its constituent parts and detecting how the parts relate to one another and to an overall structure or purpose. 4.1 Differentiate 4.2 Organize 4.3 Attributing
5. Evaluate	Making judgments based on criteria and standards. 5.1 Check 5.2 Critique
6. Create	Putting elements together to form a novel, coherent whole or make an original product. 6.1 Generate 6.2 Plan 6.3 Produce

2.7 From Bloom's to Marzano's Taxonomy

Marzano and Kendall (2007) tried to extend the original Bloom's taxonomy to turn it into more effective classification scheme. One of the major differences between Bloom's taxonomy and the Marzano's taxonomy is that the latter is two dimensional model concerning knowledge dimension and levels of processing. The objectives of the Marzano taxonomy are classified across three types of knowledge domains that can interact with six levels of thought processing. In effect, the assessment task can be designed across different types of knowledge and thought processes.

2.7.1 *Knowledge Domain*

The domain of knowledge includes three categories concerning 'Information', 'Mental procedures' and 'Psychomotor procedures'. Domain of information concerns declarative knowledge that is the knowledge of specific details and organizing ideas. A 'Mental' procedure is the knowledge of specific skills, procedures and processes that are performed mentally and psychomotor skills include performing psychomotor processes.

2.7.2 *Levels of Thought Processing*

Marzano and Kendal (2007) introduced six levels of mental operations that are not necessarily hierarchical in nature. In contrast to Bloom's revised taxonomy, Marzano and Kendal made no claim about the hierarchical levels of complexity specially in higher levels. The six levels of thought processing include Self-system, Meta-cognitive, Knowledge utilization, Analysis, Comprehension and Retrieval, Each level has been described with few sublevels. For example 'Retrieval' includes recognize, recall and execute and 'Comprehension' consists of integrating, and symbolizing. Matching, classifying, error finding, specifying and generalizing are subcategories of Analysis. The 'Knowledge utilization' level concerns decision-making, problem solving, experimenting and investigating. The fifth level is classified as specifying goals, process monitoring, clarity and accuracy monitoring. The 'Self system thinking' is about examining importance, efficacy, emotional responses and motivation.

According to Marzano and Kendal (2007), these six levels of thought processing interact with the three knowledge domains described in section 2.7.1.

2.8 From Taxonomies to Rubrics

Traditional assessments are inadequate tools for evaluating the results of ongoing inquiry and meaningful understanding as they were usually applied to measure essential knowledge skills or performances. Understanding is a sophisticated multi-dimensional process that can vary with respect to the depth of focus and breadth of knowledge. Any instructional setting has desired goals and preferences that emphasize different dimensions of learning. Assessment strategies should be aligned with the goals of instructional setting.

Wiggin and McTighe (1998) describe understanding as a multi-dimensional process with various levels of mastery along different dimensions of learning. They defined understanding as a continuum or a matter of degree, which combines misconception and insight, as well as skill and awkwardness, and is not something that can be acquired immediately or absolutely; there is no bright line dividing knowledge and ignorance.

To apply these considerations to curriculum planning and assessment, Wiggins and McTighe developed a multifaceted framework to classify the progression from naïve to sophisticated understanding. For their teaching design objectives, Wiggins and McTighe (2000) considered six dimensions (*Explanation, Interpretation, Application, Perspective, Empathy and Self-Knowledge*), which they called the six facets of understanding. They also defined five levels of accomplishment (*Sophisticated, In-Depth, Developed, Intuitive, and Naive*) for each component of their taxonomy. Their six facets of understanding is one of the assessment tools that give a good sense for how learning taxonomies can be developed into rubrics. These of rubrics are critical assessment tools for alternative evaluation.

One limitation with Wiggins and McTighe's framework is overlooking the ability of insight i.e. sudden awareness of a likely solution. By focusing on well-articulated answers as an indicator of content knowledge, they may neglect the evidence in answers that suggest genuine insights. In addition, the abilities of some students to express themselves may be limited and while they may have some understanding, they are not able to articulate what they know.

In another approach, Bennett and Dewar (2007) proposed a taxonomy with eight dimensions (*Interest, Confidence*, *Factual, Procedural, Schematic, Strategic, Epistemic* and *Social*) for their learning goals. In the same approach as Wiggins and McTighe, they developed their rubric and categorized three levels of accomplishment (A*cclimation, Competence* and *Proficiency*) for the components that they considered for their learning goals. Initially, this assessment tool was used for evaluating the mathematical knowledge that students transfer to their lives.

As an example of using taxonomies in designing physics problems, I reviewed a study conducted by Teodorescu et al., (2008) who designed a classification scheme for categorizing physics problems. Based on Marzano's taxonomy (Marzanoand Kendall, 2007), Teodorescu et al., designed a classification scheme called Taxonomy of Introductory Physics Problems (TIPP) to rank the cognitive level complexity of the physics problems. This classification scheme can serve as a guideline for selecting problems from textbooks. According to Teodorescu et al., with considering student's background and the overall goals of the course, the instructor can decide to what level of thinking complexity (how high in the taxonomy) the course should be taken. Teodorescu et al., (2009) compared level of thought processing that should be planned for conceptual physics courses for non-science majors with minimal or no algebra (regarded as the "easiest" course), algebra-based introductory courses for life science and pre-med majors (with "medium" difficulty) and calculus-based courses for engineering and physical science majors (the "hardest" course). They classified physics problems with respect to three criteria:

1. They used Marzano's definitions of knowledge types and classified the type of knowledge that was involved in the problem in two major groups of information and mental procedures.
2. They determined the highest complex cognitive process that is necessary to solve a problem (for both information and mental procedures)
3. They considered the *number of intermediate complex cognitive processes* required to solve it (with respect to information and mental procedures).

2.9 Concept Construction

Lawson et al. (1989) proposed a framework for categorizing concepts. Originally, they identified three levels of concepts in the context of biology. The first type, descriptive concepts, are those that can be sensed and observed (e.g., organism, population, magnetism, heat and focal length), The second type, theoretical concepts, are ones that cannot be observed or sensed directly, but they can be explained by deductive logic, analogies, or derivations from other theories. As examples, atoms and genes would fall into this category. The third type, hypothetical ones are intermediate to descriptive and theoretical ones. These concepts are not usually observable, but they could become so over time. Examples of these concepts would include natural selection, or evolution or other explanations of events that manifest themselves on a geological time scale. With a minor modification, we can expand on Lawson's definition of hypothetical concepts to include those that can be indirectly measured or observed such as voltage, magnetic fields, and electric fields (McBride et al., 2010). Scientists very often have used scientific models to represent a concept or to provide a better way of understanding a phenomenon. Scientific models can also be used to provide a better visualization, which in this case they can be categorized as hypothetical concepts as they simplify the visualization of the abstract reality.

Lawson and Thomson (1988) observed a relation between acquisition of knowledge and the level of concepts students employ. Lawson hypothesized in any new field of study the perception of the descriptive concepts happens first. This result has significant implication for instruction that construction of descriptive concepts, in any field of study, precedes the construction of theoretical concepts and instruction should introduce descriptive concepts prior to other type of concepts.

Inspired by Ausubel's (1977) definition of meaningful understanding, Nieswandt and Bellomo (2009) expanded on the Lawson's classification of concepts. They proposed not only categories of concepts but also types of connections among concepts contribute to their levels of abstraction. They analyzed the extended written responses of 12th-grade biology students to questions. Nieswandt and Bellomo's analysis showed that the students, had difficulties not only demonstrating types of concepts but also linking the concepts, the latter proving to be even more challenging.

Their findings have important consequences for the assessments of students' abilities to reason scientifically because the evidence of understanding does not just involve demonstrations of the various types of knowledge but being able to link these types of knowledge. Effective links occurred only when the discrete concepts were connected meaningfully with plausible links among the chains of cause and effects. In short, this occurred only when the students were able to explain what happened, why it happened and how the causes related to the effect.

Nieswandt and Bellomo distinguished three types of links among concepts. A one-concept-level link refers to connections between two concepts from the same category (e.g. descriptive and descriptive). Cross-concept-level connections are connections between two different categories of concepts (descriptive, hypothetical) and multi-concept-level links occur when all three categories of concepts (descriptive, hypothetical and theoretical) are connected. The least sophisticated type of concept link is a one-concept-category link while the most sophisticated is the multi-concept-category link. Nieswandt and Bellomo postulate that meaningful answers must reflect multi-concept-level links among the concepts.

2.10 Previous Attempts at Assessing Students' Reasoning

Prior to developing our approach, we reviewed a few other methods that are often used to probe students' scientific reasoning skills and conceptual knowledge and determined they were not appropriate for this study. We believe that understanding in the sense of a constructivist perspective involves more than the acquisition of knowledge and being able to state correct answers.

Consequently, in the discussion that follows, I review several well-known instruments often used to assess conceptual understanding and students' reasoning and explain why they were not effective for the measurements required in our study.

Concept Inventory

Concept inventories are usually multiple-choice tests designed to evaluate specific sets of concepts and are based on extensive research. The Force Concept Inventory (FCI), developed by Hestenes et al. (1992), is a well-known concept inventory with the questions, responses and distracters that are constructed based on actual interviews with students, so the outcomes and the students' commonly held misconceptions are known in advance. The FCI is multiple-choice test designed to evaluate students' conceptual understanding of force and related kinematics and intended for pre-and post-testing.

In similar ways, Thornton and Sokolof (1998) assess student learning of Newton's laws and introduced a research-based, multiple-choice assessment of student conceptual understanding. For their approach, they developed a multiple-choice assessment of student conceptual understanding of Newton's laws of motion, which is called Force and Motion Conceptual Evaluation (FMCE) (Thornton et al., 2009). The Conceptual Survey of Electricity and Magnetism (CSEM) is another concept inventory surveying students' conceptual knowledge of electricity and magnetism (Maloney, et al., 2001). Commonly concept inventories that developed for physics courses are multiple-choice assessment instruments used mostly in introductory physics courses, but differ in content domains and also use different representational formats.

Subsequent development of concept inventories in other disciplines included inventories in biology (Anderson, 2002), astronomy (Zeilik et al., 1999), chemistry (Mulford, 1996) and geology (GCI) (Libarkin, 2008).

Because we were investigating students' reasoning across the disciplines, discipline-specific concept inventories would not have been appropriate tools. Developing an individual concept inventory for the purpose of our study would not have been practical as we were dealing with different subjects of chemistry, astronomy, biology and geology since the development of the types of questions for these types of inventories would have required extensive research.

In addition, according to (Libarkin, 2008) the relationship between scores in different inventories are not clear. For example, learning gains measured by the FCI, are linked to specific teaching approaches useful for physics instruction that may not be appropriate for another discipline. In other words, it is not clear how the learning gains or effect sizes measured by different inventories are comparable.

Furthermore, it was not possible for us to know precisely what topics were covered in the different courses for which we collected data. The faculty who were teaching the courses designed the courses syllabuses. Another contributing factor is the time that we collected data during our site-visits. We visited the courses at different times of the semesters through several years. As such, we were not assured that students have enough knowledge for all the content that were assessed by the inventories.

Concept Maps

The concept map is an alternative assessment tool often used in inquiry-based settings, and it asks students to depict graphically the relationships among concepts Concept mapping was originally devised by Novak (1996). In the concept map model, the concepts are enveloped by students in circles or boxes and connected with lines, arrows and relationships that can be specified with a short description.

Novak based concept mapping on the Ausubel's theory of meaningful learning and a theory of knowledge. This kind of assessment that requires students to find the links between of areas of knowledge is grounded in the cognitive theory research as it uncovers what learners know, as well as hierarchy and relationships of their conceptual structure.

According to Beaty (2002), not only the amount of knowledge distinguishes expert and novice, but also the way they both structure and organize the knowledge. Beaty (2002) elicited information about conceptual structures and mental maps by probing students' conceptual knowledge structures and investigating the inter-connections, nodes and links within those conceptual structures. In another attempt, Vaides et al. (2005) suggested a type of concept map in which the concepts were given, and the students connected the concepts with one-way arrows and labeled the arrow with short phrases describing the relationships.

We decided concept maps were not practical assessment tools for our study because the design and the complexity of concept maps can change from one question to another. Designing a concept map carefully enough so that it could be used with the same degree of complexity across the disciplines with a same would have been too difficult. Moreover, not all the students are familiar with using concept maps and training them to be able to use them would not have been practical for our research setting. Although concept mapping requires linking concepts, it does not involve students in the higher levels of thinking that were included in the learning goals of our study. These goals also included applying concepts in new contexts, inferring, comparing, and other types of reasoning.

Lawson Classroom Test of Scientific Reasoning

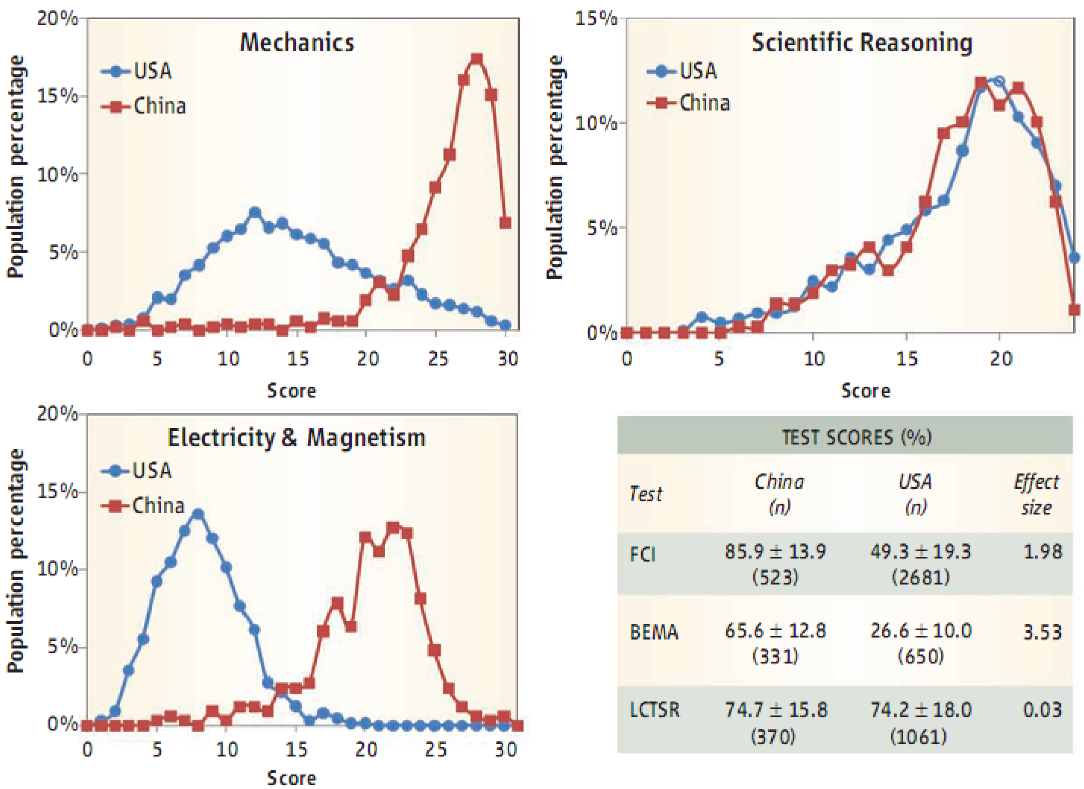
Based on interviews with students, Lawson (1978) developed a classroom test of reasoning. The classroom test was designed to allow teachers or researchers to classify student reasoning from concrete to formal developmental levels. The test included items that measured concrete levels of thinking such as understanding the principles of the Conservation of Weight (Piaget & Inhelder, 1962) and Displaced Volume (Karplus & Lavatelli, 1969). The items on the test measured higher levels of formal thinking including cognitive knowledge of Proportional Reasoning (Suarez & Rhonhelmer, 1974), Controlling Variables (Inhelder & Piaget, 1958), Combinatorial Reasoning (Deluca, 1977), Combinatorial Reasoning and Permutations (Longeot, 1965) and Probability.

Colleta and Philips (2005) conducted research to find correlations between the FCI and Lawson's Scientific Test of Reasoning. They administered Lawson's classroom test of scientific reasoning to 65 students and found a significant, positive correlation between students' normalized FCI gains and their Lawson's score. Because of this study, they speculate that variations in average reasoning ability in different student populations are associated with the students' FCI gain. They suggest that students who have not learned the concepts previously and have higher levels of scientific reasoning are more likely to achieve higher FCI gains. In another study, Bao et al. (2009) used quantitative assessment tools to compare Chinese and American students in their conceptual understanding and reasoning abilities. Bao et al. used the concept inventories of FCI (Force Concept Inventory) and BEMA (Brief Electricity and Magnetism Assessments) to compare their conceptual and content knowledge, and they used LCTSR (Lawson Classroom Test of

Scientific Reasoning) to evaluate students' reasoning abilities. Their study shows (Figure 3) that the distribution of reasoning abilities is almost identical; however, the results from the FCI and BEMA also show that having had rigorous and numerous physics courses enhanced the performance in content knowledge for Chinese students.

Both studies (Colleta & Philips, 2005; Bao et al., 2009) show that the Lawson score is not always correlated with the students' content knowledge. According to Colleta and Philips (2005), the correlation factor between FCI and Lawson's test varies with respect to the population being assessed. However, Bao's study (2009) shows that Chinese and American college students differ with respect to their content knowledge, but not with respect to their abilities to reason scientifically. Therefore, we concluded that the Lawson test would not allow us to investigate students' reasoning in the scientific contexts.

Figure 2-1 Comparison of Chinese college students and US shows differences in content knowledge but not on tests of scientific (Ed, 2009)



TEST SCORES (%)			
Test	*China (n)*	*USA (n)*	*Effect size*
FCI	85.9 ± 13.9 (523)	49.3 ± 19.3 (2681)	1.98
BEMA	65.6 ± 12.8 (331)	26.6 ± 10.0 (650)	3.53
LCTSR	74.7 ± 15.8 (370)	74.2 ± 18.0 (1061)	0.03

2.11 Summary

In this chapter, I reviewed the literature that has already shaped current discussions of reform, standards, assessment and quality of students' reasoning. We adopted NSES professional development standards (National Research Council, 1996) as a model of practicing scientific inquiry. I reviewed the literature related to inquiry because we were investigating if students in inquiry oriented courses reason better than those in traditional courses. Development of inquiry is intimately linked with learning theories that are consistent with contemporary emphasis on constructivism. Many of the learning theories postulate that meaningful understanding involves transfer of learning. We hold the view of previous researchers (Bransford *et al*., 1999) that placed importance on both the cognitive processes and knowledge that students construct in a new context.

I reviewed a few types of taxonomies that discussed hierarchal nature of cognitive processing. Anderson et al. (2001) expanded on Bloom's taxonomy and distinguished among cognitive processes and knowledge processes that are used for retention and transfer.

In addition, I reviewed previous studies that have addressed the issue of assessing students reasoning; however, if we value transferability of knowledge, getting evidence of understanding also means designing assessments that probes transfer of learning. In the next chapter, I explain the research methodology that we adopted for probing and classifying students' reasoning in diverse settings.

Chapter 3 - Research Methodology

3.1 Introduction

In this chapter, I discuss our use of the findings, previous research, educational goals and requirements of National Science Education Standard (NRC, 1996) to design an assessment tool that classifies students' reasoning abilities. As a part of the NSEUS project, we developed a protocol to explore and compare the students' reasoning skills in the content of science courses. The participants in our study were pre-service elementary teachers who were given open-ended questions on their final exam.

Our assessment protocol included a template for question design and a rubric to analyze responses. In the discussion that follows, I describe our research conditions and clarify our goals. In addition, I discuss the rationale informing our assessment protocol, our philosophical perspectives, and the validity and the inter-rater reliability of our assessment tools. Finally, I present a few examples to show how we used this assessment tool.

3.2 Research Method

The nature of our research was qualitative because it involved a deep understanding of a complicated phenomenon, namely students' reasoning skills. Creswell (1998) classified qualitative studies according to five traditions, those being narratives, grounded theory, case study, ethnography and phenomenology, and discussed the structures of these different approaches. Each of the five traditions of the qualitative research methodologies might be appropriate for certain type of study in relation to the research question and nature of the study. Creswell also suggested that a mixed approach might in some instances address better the different aspects of the study.

We adopted the approach of grounded theory (Glaser and Strauss, 1960) in which we developed a theory that was grounded in the data we collected. We used both qualitative and quantitative methods to generate a model from the data. Our first step was data collection through variety of methods. We analyzed the data through different lenses and classified data using a hierarchical coding system. From the categories that emerged, we developed a model that defined the nature of relationship between students' reasoning and the extent that the courses are inquiry oriented. While through qualitative methods we classified the data, we also had to adopt statistical methods to predict the probability of occurring of the categories which some of them were dependent and independent variables. Therefore, we adopted a mixed methodology involving qualitative and quantitative methodologies.

3.3 Research Setting

This research was conducted at Kansas State University as a part of collaborative project with University of Alabama and San Diego State University. The data were collected from sample universities by the teachers of the relevant courses and were sent to us for analysis.

3.4 Participants and type of sampling

This comparison was conducted for 18 science courses located at 13 universities across all scientific disciplines. The sampling process had two stages. First, based on the NSEUS sampling criteria, the universities that met certain characteristics were selected from among 130 universities that participated in the professional developmental workshops. The characteristics of these universities are described in the Table 3-1 below. Initially, NSEUS researchers planned to choose randomly among those universities that met their selection criteria; however, the final participants were just those universities that agreed to participate in the study. Kansas State University did not control the sampling process and our collaborators at the University of Alabama arranged participation with the universities who were willing to participate in the study. The target population was selected from two strata of reformed and traditional courses offered for elementary education majors. We analyzed the data from all the students who participated in the selected courses and compared the results.

Table 3-1 Characteristics of selected universities

Characteristic	*Average*
Years course offered at institution	**6**
Times offered in a year	**2**
Course enrollment	**range 20 -275**
Minority enrollment	**25%**
Credit hours	**4 (range = 3 - 5)**
Sections per semester	**1.5 (range = 1 - 6)**

On most campuses, a traditional course and an interactive engagement course at the same level and in the same subject area do not exist, making direct comparisons on subject matter learning impossible.

Because of the size of the study, we could not use interviews as research tools and had to develop another type of instrument to evaluate students' reasoning. We focused on those methods that would elicit students' reasoning rather than those would measure the degree of correctness of their answers. Therefore, we compared the reasoning skills within the content that the students had learned. Comparing students learning skills across disciplines is not a straightforward endeavor. Since the characteristics that would determine a well-reasoned answer and the categories we would use to address those characteristics would be independent of the question's context, we created a template to classify the sophistication of reasoning in different contexts and employed the template to compare students' reasoning across different disciplines.

3.5 Philosophical Perspectives and Assessment Design

The whole approach toward developing the assessment protocol resembles the "backward design" strategy developed by Wiggins and McTighe (1998). In this approach, one starts with the end goals or standards that are desired and then designs an assessment tool to reflect the achievements of those predetermined goals. We defined three stages for developing our assessment design, namely identifying desired results that are due to the inquiry learning goals, determining what are the acceptable evidences for achieving the desired goals, and planning for the assessment design accordingly. To adapt the three stages to the context of our study rewrote them as following questions:

- What is worthy or requiring of a well-reasoned response?
- What is the evidence of a well-reasoned response?
- How we can design our content question to cultivate accurate information about how well students are reasoning?

3.5.1 *Stage one: Inquiry oriented assessment goals*

In the first stage, we used the National Science Education Standards (NRC, 1996) and previous research to identify the priorities for the desirable reasoning skills in the assessment of inquiry. Once we clarified our theoretical perspective toward assessing students' reasoning, we refined our instructional goals and broke down each objective into specific goals through the lens of the standards that were appropriate for our target population. Finally, we used the revised Bloom's taxonomy to make our objectives more contextualized and more specific, for the purpose of devising appropriate assessment designs.

Our approach was informed by the findings of research into schemata theory (Mayer, 2002). Accordingly, we believed that the sophistication level of reasoning relates to the pieces of knowledge and cognitive abilities that students bring to a new context and the way they connect and organize pieces of information. In other words, we believed reasoning can be defined in terms of the thought processes and knowledge types that students bring to a new context. Because inquiry oriented teaching approaches always involve exploration, and therefore, require that students be exposed to new contexts, an inquiry-oriented assessment tool must aim to measure the transfer of learning.

Transfer requires that the students recognize relevant facts and concepts in a new environment and discover how they are interrelated. The interconnections can be generalized to the knowledge of theories and principles. Consequently, students may rethink the structure of their conceptual schema, select the relevant concepts, and find associations between concepts, theories, procedures by inference, induction, deduction, analogical, relational, or cause-and-effect reasoning. Each of these tasks may require higher levels of cognitive processing than recall.

Science Content Standards (National Research Council, 1996) have also emphasized the use of more sophisticated assessments that evaluate higher order skills. Developing these types of assessments is not easy and requires the consideration of different variables. Because the science content standards advise shifting the focus from evaluating factual knowledge toward measuring the conceptual understanding and procedural skills, activities that verify science content should be modified to activities that investigate and analyze science questions. In assessments that value inquiry, assessors should also focus on measuring multiple process skills, cognitive and procedural skills and how these skills are applied in new contexts. In addition, the National Science Education Standards emphasize being able to communicate scientific explanations rather than providing answers to questions. According to those standards, statements that follow one another logically without gaps from statement to statement characterize a well-reasoned response. In other words, students' responses should represent the complete chain of "What", "Why" and "How".

Along the same lines, the American Association for the Advancement of Science (AAAS) developed Benchmarks for Science Literacy (1993) and included the *"knowledge of how*" in each set of benchmarks. Schematic and strategic knowledge (knowledge of how and why) involves greater degrees of complication and requires deeper levels of understanding. Russ et al. (2006), acknowledged these standards when they drew attention to the construction of cause-and-effect relationships in students' responses. They examined students' ideas about how particular components of a system cause its actions. Russ et al. emphasized not only the association of cause and effect, but also the underlying process that explains how the cause and effect are associated.

Anderson et al. (2001) were concerned with the lack of consistency and severe misalignment that exists between instructional goals and assessment design. In addition, judging correctness, evaluating the use of controlling variables, or measuring students' conceptual knowledge may not effectively assess the students' gains from inquiry learning (Russ et al., 2008).

Consequently, we needed to design an assessment tool that values the qualities of inquiry and is sensitive enough to examine achievement with respect to our goals. As such, our assessment tool needed to be able to differentiate between knowing and applying knowledge.

3.5.1.1 Specific goals- A framework for objectives

Based on a review of knowledge construction (Mayer, 2002; Bransford et al., 1999), Anderson et al. (2001) expanded on Bloom's taxonomy (Bloom, 1956) and added another dimension to it. The matrix they constructed also shows the relationships among the types of cognitive processes and knowledge development.

To provide a better visualization of the objectives that we discussed earlier and to show the organization of our classification scheme, we mapped our goals for the assessment task according to the hierarchies of thought processes in revised Bloom's taxonomy (Anderson et al., 2001).

Table 3-2 Selected components from Bloom's revised taxonomy

The Knowledge Dimension	**Cognitive Dimension**				
	Remember	**Understand**			**Apply**
		Interpret, Compare	Infer, Explain	Exemplify, Classify	Implement (New context)
Factual Knowledge					
Conceptual Knowledge Conceptual schema, Classification, Principles, Theories					
Procedural Knowledge					

In their scheme, complexity of the cognitive dimension increases from “Remember” to “Apply” while hierarchies of knowledge proceed from the lowest level of factual knowledge to the more complex and abstract levels of conceptual, procedural, and meta cognitive knowledge. Sometimes, however, conceptual knowledge develops out of procedural proficiency and vice versa. Thus, procedural knowledge may not be more abstract than conceptual knowledge in all cases.

Many studies (Kim, 2001; McDermott, 2001) have shown that facility in solving problems following mathematical procedures is not an evidence for conceptual understanding. Furthermore, these studies have shown that students’ abilities to comprehend different types of knowledge are not the same. Accordingly, one may think of distinctions between different types of knowledge or whether higher grades of mental engagement are required for some types of knowledge in relation to other types.

For example, students may experience more difficulty when they encounter a task that requires comparing ideas and situations than the difficulty they encounter when they are retrieving knowledge from long-term memory. Similarly, paraphrasing and changing representations, finding cause-and-effect relationships or inferring and drawing a logical conclusion are different levels of cognitive processing that each may pose cognitive challenges with different degrees of difficulty.

Based on Anderson’s revision of Bloom's taxonomy, we classified every type of knowledge and cognitive processes in terms of subtypes that are more specific. For instance, ‘Conceptual Knowledge’ has four subtypes: the knowledge of the interrelationships between facts (*conceptual schema*), the knowledge of classifications, the knowledge of principles, and the knowledge of theories and structures. The subcategories of the cognitive process of “Understand” include changing representation, exemplify, classify, summarize, infer, compare, and explain. Therefore, when we refer to the category of “Understand”, we needed to be more specific about the subcategories we chose to incorporate in the question.

The two-dimensional framework displayed in Table 3-2 consists of 15 separate cells. Each separate area is created by the intersection of rows and columns that belong to a certain type of knowledge and cognitive process. In other words, the cells represent the types of knowledge and cognitive processes that are required for thinking about a situation. For example, if student recalled only facts, their level of reasoning was placed at the top left of the grid. On the other hand, classifying students' answer to be located in the cell in the right bottom corner, represents a good level of reasoning as the intersection shows the application of the knowledge had occurred and the features of the new situation were intertwined with the construction of knowledge.

3.5.1.2 Scientific goals-Level of abstraction

As another indication of understanding, we used questions that elicited students' conceptual structure that is their ability to exhibit multi-level-links. Such ability was exhibited if the response showed the student was able to construct an argument based on descriptive, hypothetical, and theoretical concepts. In order to show a more in-depth level of reasoning, we classified concepts into three types: Descriptive, Hypothetical, and Theoretical (Lawson et al., 2000), and categorized the level of abstraction of the responses in terms of the types of concepts and the links between them. We used Nieswandt and Bellomo's (2009) method, which was described in the previous chapter, and administered questions that included several concepts on different levels (descriptive, hypothetical, and theoretical) and would require students to link them to articulate a comprehensive answer.

3.5.2 *Stage two: Evidence of Reasoning*

3.5.2.1 Evidence of transfer

The constructive process of transfer depends on many prerequisite steps as specified in Anderson's taxonomy. Based on the definitions in Anderson's grid, we needed to clarify what parts and pieces of the responses represented--whether they represented factual knowledge, conceptual schema and knowledge of classification, theories or, procedural knowledge. We also applied Anderson's cognitive dimension to students' answers and identified the type of cognitive processes they employed in connecting the concepts, facts, or types of knowledge.

At the lowest level, connections may have occurred by recall. However, many connections occurred because students found similarities and differences between two facts or concepts. If the responses involved converting one representation format to another, we could assert that cognitive process of interpretation had occurred to some degree. Some responses demonstrated the ability to identify a concept within a certain category as the responses showed the students had recognized common features among the concepts and properties within a certain category. Higher levels of reasoning included "inferring" and "applying," which involved the recognition of concepts, relationships and common patterns in new situations.

3.5.2.2 Evidence for level of abstraction

Nieswandt and Bellomo offered a multi-stage analysis for evaluating the links.

Here is a brief and modified description of their data analysis method that we used to ascertain evidence of conceptual knowledge.

1) They divided answers into segments, each of which reflected an individual idea

2) In each segment, they highlighted the scientific term and classified it as one of three concepts levels (descriptive, hypothetical or theoretical)

3) They coded the effective links and differentiated three types of links; linking of concepts on the same level (descriptive to descriptive), called "one-concept-level links"; linking of concepts on two different levels (theoretical to hypothetical), called "cross concept-level links;" and linking of concepts on all three levels (descriptive, hypothetical, and theoretical), called "multi-concept-level links."

3.5.3 *Stage 3: Developing Content Questions*

In previous sections, we characterized our goals and defined what would indicate evidence of understanding. In this section, we explain how we developed the questions that would elicit responses to help us achieve our desired goals. It was critically important that we should design a scenario that would be new to the students, otherwise; a rote learner would be able to use memorization to give thorough and complete answers. On the other hand, if we went beyond rote questioning and designed content questions that required the application of newly learnt concepts in a new context, the rote learners would no longer be able to answer them. Our goal, however, was, is to investigate how students would proceed from the initial step of knowledge construction to applying that knowledge to a new context. As such, we needed to ask questions that led students through the dynamic process of knowledge construction. Given the limitations on our research, we could not design questions that required many types of knowledge structures and cognitive processes. Our content questions were placed on final exams and only 10 to 15 minutes were allowed for answering each question. In addition, to compare reasoning skills across disciplines, we had to follow the same structural format for our knowledge types and cognitive processes. Therefore, we had to select from the subcategories. To design questions for the category of “conceptual knowledge” we mostly considered the subcategory of “conceptual schema” and for the cognitive process of "understand," we examined students’ ability to demonstrate “infer," "compare," and "explain." For the category of “apply”, a higher level of cognitive processing, we selected the subcategory of “implement” which requires applying different knowledge types to unfamiliar tasks.

3.6 Template for developing content questions

To be consistent with our assessment goals and criteria for evidence of understanding, we determined that content questions in different contexts such as biology, physics, chemistry, astronomy and geological sciences would need to have following characteristics:

- Require that students explore a phenomenon based on the notion that there are some similarities between earlier learnt skills and concepts and the feature of the new scenario. We were mindful that Bloom (1956) had argued that if the situations were to involve application, the situations should have new elements that differ from the situation in which the abstraction was learned.
- Evaluate whether students could recognize relevant scientific concepts and the relationships among the concepts. Consequently, students would be required to activate

the relevant scientific concepts in the scenario and recognize the relevant subject specific skills, rules and principles.

- Be designed to trigger higher levels of cognitive processes. The question scenario would need to be designed to include basic ingredients of students' pre-existing knowledge with some variations that could affect the features of the scientific concepts. The question would need to encourage students to investigate the changes and variations that had occurred to the system and how and why the changes could affect the outcomes. The appropriate information processing would require combining the following abilities: recognizing, changing representation, comparing and inferring.
- Be designed to avoid oversimplification and evaluate elements of abstraction. To consider this aspect of understanding, we designed questions with multi-level-concept link structures. We aimed to expose students to situations that would involve multi-level concepts (descriptive, hypothetical, and theoretical), require applying different types of knowledge, and produce multi-level links to infer the relationships between the cause and effects and to reinvent a theory. Nieswandt and Bellomo, (2009) showed students predominantly exhibit one-concept-level links (Descriptive-Descriptive), and had more difficulty demonstrating cross-concept-level links (theoretical-hypothetical) while multi-level-links rarely occurred.
- Evaluate students' subject-specific skills. This knowledge could be either replicable or applicable meaning that the students could show that they could have learned sets of rules and skills by practice or knew how to retrieve the appropriate procedures. However, in the higher levels of understanding the students would have to

interact with a new subject and recognize how to apply the appropriate procedures in a novel context.

- Examine the students' communication skills by exposing them to situations that would require explanation. Because an assessment should be able to discern "what" students know and "why" and "how" they arrived at certain conclusion in well-crafted successive statements, our written questions would need to require answers that explained a coherent plausible pattern of thoughts with inter-related pieces of knowledge and a logical chain of cause and effect.

We used this predefined structure for different contexts and disciplines and used a common language for the protocol of devising our content questions across disciplines. As a result, we used a structure that would control the level of abstraction in our questions without being dependent on the context. In other words, we designed content questions with structures that were similar across the disciplines in terms of their thought processes and knowledge types.

3.7 From assessment goals to rubric

Up to this point, our discussion has concerned with how we devised questions that would meet our goals and elicit the different types of knowledge and conceptual structures that interested us. Here we describe our approach to analyzing and comparing the responses.

Using the specific techniques of qualitative research (Cresswell, 1997) and grounded theory (Strauss & Corbin, 1990), we analyzed the data, figured the range of explanations that students provided and explored different categories, which emerged from students' responses. We classified responses to distinct levels and analyzed each level separately. To develop a rubric that we could use for the analysis, we considered eight different possible steps of reasoning that was usually a common pattern in the students' responses. The first column in the Table F-1 in Appendix F lists the different levels of reasoning and the second and third columns the corresponding knowledge types and cognitive processes associated with each level of reasoning. Therefore, the extent to which the students fully implement the steps of reasoning reflects the degree of mastery they have of each type of knowledge and cognitive process. In a similar fashion, the second table in Appendix F also emerged from students' responses in order to show the characteristics of answers that proceed from naïve to the higher levels of thinking. The second and third columns show the corresponding knowledge types and cognitive processes. The fourth column represents the concept link structure (Nieswandt and Bellomo, 2009) that may be associated with the seven types of the answers.

The categories that were obtained from the students' responses had certain characteristics that are described in the second table in Appendix F. These two tables together provided us with plenty evidence that guided us to distinguish the levels of quality and proficiency of the students' responses. We then developed a rubric that described the progression of the knowledge types and cognitive processes. Our goal was to develop several independent criteria, which we could use separately to judge the quality of the responses. In their work, Wiggins and McTighe (2006) distinguished between holistic scoring and analytic-trait scoring. In holistic scoring, the assessors report their overall impressions of performances while analytic-trait scoring evaluates students' achievements with respect to several distinct criteria. Consequently, in analytic-trait scoring the performance is examined more than once through the lenses of different criteria.

Fundamentally, this approach of developing a rubric followed a procedure that had been used several times previously. As we discussed in the literature review, Wiggins and McTighe considered different aspects of understanding and derived six facets (*Explanation, Interpretation, Application, Perspective, Empathy,* and *Self-Knowledge*) for understanding. In their approach, the authors viewed understating as a matter of degree on a continuum and defined five levels (*Sophisticated, In-depth, Developed, Intuitive* and *Naive*) of accomplishment for the six facets they derived for understanding.

In a very similar approach to Wiigins and McTigh, Bennett and Dewar (2007) developed a rubric to evaluate the mathematical knowledge that students transfer to their lives. They derived eight facets (*Interest, Confidence*, *Factual, Procedural, Schematic, Strategic, Epistemic* and *Social*) for their learning goals and constructed their rubric by categorizing three levels of accomplishment (A*cclimation, Competence* and *Proficiency*) for the eight criteria that they defined for their educational goals.

Following the same approach, we differentiated the students' performances with respect to seven traits of factual knowledge, conceptual schema, and procedural knowledge, compare, infer, explain and apply. These seven traits map out the kind of evidences we need to classify the quality of reasoning. To provide a framework of making distinctions for reasoning levels, we described criteria for proving or disapproving that certain trait occurred. The rubric that I present in section 3.7.1, articulates the list of criteria we adopt to classify students reasoning.

3.7.1 *Rubric*

We construct our rubric by classifying two levels of accomplishment for each trait marked by evidence and no evidence. The abbreviation "E" and "NE" stands for evidence and no evidence.

Factual Knowledge

For any particular subject, factual knowledge refers to discrete pieces of information that are basic constituents of that subject. For the evidence of factual knowledge, we look to see if factual types of information are mentioned verbally or symbolically or they can be inferred from other statements.

NE= The students' written answer displays negligible evidence for having access to basic premises and discrete pieces of information that are basic constituents of the subject that are required for construction of conceptual schema.

E= The students' written answer displays some evidence of having access to the basic facts and discrete entities that are required for construction of conceptual schema, including cases that lacks some facts in other words account that is identified by mixture of relevant and irrelevant facts.

Conceptual Schema

Conceptual schema is one of the subcategories of conceptual knowledge in Anderson's table. For this type of knowledge, we investigate if the appropriate concepts are employed and if the definitions of concepts are clear and correct meanings are attributed to the concepts. Moreover, we evaluate the understanding of relations between the concepts and concept links and we look whether special attributes, specific features are associated with appropriate categories or classes.

NE= The students' written answer displays negligible evidence of knowing the meaning of the concepts or knowing the relations between the concepts; employing wrong concepts and attributing wrong meaning to the concepts without understanding of the relations between them or introducing the concepts by recall without showing their meanings.

E= The students' written answer displays some evidence of having access to relevant concepts; understanding the meaning of the concepts in relation with other concepts; constructing partially appropriate schemas in which student clarifies the relationship between the concepts or account that includes relevant concepts that are mixed with some irrelevant concepts.

Concept Level Links

Based on Nieswandt and Bellomo's multi stages analysis, we draw a symbolic representation of concept links by categorizing the level of the concepts students employed in their answers in terms of descriptive, hypothetical and theoretical. We can show the links with abbreviations T (for theoretical), H (for hypothetical) and D (for descriptive) and show the links students employed as H-H, H-D-T and so on.

a) Descriptive concepts: Concepts that can be inferred or observed with direct senses e.g. magnets, organisms, food chain

b) Hypothetical concepts: the concepts that cannot be observed directly but indirectly with employing a model or if the observational time period were extended e.g. magnetic fields or fossils

c) Theoretical concepts: The concepts that cannot be observed and the meanings come from the theories which ideas originate e.g. atoms and genes.

Procedural Knowledge

For this knowledge, we look to see if the students have the skills, knowing the techniques and algorithms or knowing rules and steps of applying a principle. Procedural knowledge manifests in different faces, such as algorithm, trigonometry, geometry, physics formula, vectors and so on. In the context of chemistry, procedural knowledge is attributed to the, the knowledge of writing chemical equations and balance the chemical reaction on paper or the procedures that students perform to cause chemical reactions to occur. In the context of genetics, procedural knowledge refers to the knowledge of knowing how to find probabilities or using combination rules of probabilities. In broader sense, procedural knowledge also concerns the knowledge of the prescribed steps of solving particular problems, or knowledge of steps to execute a process, or series of steps that are required for verifying a principal.

NE= The students' written answer displays negligible awareness of subject specific skills and techniques to implement the procedures or rules.

E= The students' written answer displays some evidence of being skillful or having some knowledge in using subject specific skills and techniques and knowing the series of steps to execute the procedures.

Compare and Contrast

The cognitive process of compare contributes to reason by analogy (Anderson, et. al, 2001). We looked to see whether the students were able to determine analogous situations or patterns, and we assessed their ability of finding one-to-one correspondences and to recognize the familiar situations in a new context.

NE= The students' written answer displays comparisons by recall, comparisons of irrelevant aspects which cannot establish a reasonable connection between cause and effect or concrete comparisons of superficial features of subject.

E= The students' written answer displays going beyond superficial aspects and comparing more in-depth features inferred from given information. They compare those aspects and features that are fundamental for justifying cause and effect changes, or compare variables that provide plausible evidence for why and how and what changes occurred, also including cases that some meaningful comparisons exists but there is a lack of compared entities for a plausible connection for what and why and how things happened.

<u>Understand (Infer)</u>

We assess if the answer recognizes the patterns between series of the events and instances.

NE= The students' written answer displays a nonsense conclusion including fragmentary segments, fail to relate assumptions and conclusion, or the links between assumptions and conclusions are either by recall or has been constructed concretely.

E= The students' written answer recognizes a pattern and finding a reasonable and plausible connection and developing some insightful relations between cause and effect with some evidence of plausible relationship between what and why and how.

<u>Explain</u>

We look for a cohesive and convergent argument that leads to reasonable conclusion.

NE= The students' written answer displays a descriptive and superficial or borrowed idea; providing an answer without supporting, based on personal assumptions or concrete idea, non cohesive and a fragmentary and sketchy argument.

E= The students' written answer displays some evidence of well supported by argument that shows explicit thought, subtle connections between assumptions, theories and types of knowledge required in the problem, also showing some justification and good sketch of "What" to "How" and "Why, judged from comparing specific and in-depth features of the subject. An argument has segments supported by another and cohesively leads to a reasonable conclusion, including cases that showed incomplete internalizations or contradictory statements that coexisted with meaningful connections.

Apply

We assess if students can recognize the information, the relevant concepts, principles and the relations between the facts, concepts and principles in the new context.

NE= The students' written answer displays an association of facts, concepts, procedures that are not explored in the context of question's scenario.

E= The students' written answer displays an association of facts, concepts, procedures and features of questions' new context that are partially reconstructed or association of facts, concepts, procedures reconstructed in connection with the features of question scenario to present a plausible answer.

3.8 Inter-rater Reliability

In this section, we report the results of assessing the inter-rater reliability in using a coding scheme we developed for classifying students' patterns of reasoning. We conducted the study in two phases and checked the inter-rater reliability along the three dimensions of our analysis, namely the students' knowledge level, cognitive skills, and types of conceptual structures in their responses.

3.8.1 *Phase one*

In the first phase, we conducted the study with two separate groups. With the first group, we checked our agreement in using the rubric and with the second group; we checked our agreement on conceptual structure coding. Each part of the study had its own group of raters consisting of a senior graduate students and research associates in the field of physics education who were working in our research group. The raters coded the students' responses individually and compared their codes with mine.

3.8.2 *Training sessions*

Before conducting the study, both groups underwent separate training sessions. The first group had two sessions of training until the raters felt confident about using the coding scheme. During the training sessions with the first group, we reviewed a few of the sample questions and responses. The sample questions were about moon phases and energy saving in the home in which white and black curtains were used. Raters coded few sample responses individually using the instructions were provided. Then they compared and discussed their scores together.

While listening to their discussions and interpretations, I identified the discrepancies in their assessments. After they resolved their disagreements among themselves, they compared their coding to mine, and we discussed the discrepancies until we reached a good level of agreement among ourselves.

With the two other participants in the concept link group, we followed the same procedure. The two sample questions were about conservation of momentum and about energy saving in the house in which white and black curtains were used. The participants coded a few sample responses. We discussed few cases and quickly reached to an agreement about our coding.

3.8.3 *Results*

Following the training session, the first group was given two questions with 10 sample answers for each question. The questions were about light and color and heat transfer and the participants brought the results back two days later. On the light and color question, our overall agreement was 83% with the first rater and 80% with the second rater.

On the heat transfer question, our agreement was 89% with the first rater and 72% with the second rater. On average, our inter-rater reliability on this measure was 81%.

Table 3-3 Inter-rater reliability result for light and color question

Factual	Conceptual/ Schema	Procedural	Understand/ Compare	Understand/ Infer	Understand/ Explain	Apply
90%	90%	90%	70%	85%	60%	95%

Table 3-4 Inter-rater reliability result for heat transfer question

Factual	Conceptual/ Schema	Procedural	Understand/ Compare	Understand/ Infer	Understand/ Explain	Apply
90%	89%	N/A All agreed	90% *N [1]	75%	80%	85%

[1] *N= At least one of the inter-rater disapproved that the trait occurred. The inter-rater reliability score reported in the table is among other raters who approved the occurrence of the trait.

For the concept link group we used 10 sample answers to a question on the conservation of momentum and 10 sample answers to a question on heat transfer. For this type of coding, the raters needed to divide the responses into the segments that represented same idea, to find the key concepts in the responses and to categorize the key concepts as either descriptive, hypothetical, or theoretical based on Lawson's classification of concepts. According to the nature of this type of coding, the inter-rater reliability could not actually be scored or quantified, but the results showed a significant consistency among the raters in terms of selecting the segments, finding the key concepts, categorizing them, and representing them as concept links.

3.8.4 *Phase two*

In order to validate our reliability test and check its consistency we repeated the test with a separate group of raters consisting a senior professor and a research associate in our physics education research group. The raters coded the students' responses individually and compared their codes with mine. The inter-rater reliability test was on three questions with 10 sample answers for each question. The questions were about a charged particle in a capacitor, moon phases and replication of DNA. All three raters examined the capacitor question and the overall inter-reliability score on that question was 78.5%. The inter-rater reliability test for the moon phase question was conducted between research associate and me and we reached at 70% total agreement. The results for the three examined questions are presented in the tables below:

Table 3-5 Inter-rater reliability result for capacitor question

Factual	Conceptual/ Schema	Procedural	Understand/ Compare	Understand/ Infer	Understand/ Explain	Apply
89%	89%	90% N*	63% N*	76%	80%	60%

Table 3-6 Inter-rater reliability result for moon phases question

Factual	Conceptual/ Schema	Procedural	Understand/ Compare	Understand/ Infer	Understand/ Explain	Apply
90%	90%	50%	70%	60%	70%	50%

Table 3-7 Inter-rater reliability result for DNA question

Factual	Conceptual/ Schema	Procedural	Understand/ Compare	Understand/ Infer	Understand/ Explain	Apply
73%	73%	*N	67%	80%	87%	93%

3.8.5 *Reliability overall*

The average reliability for each trait is shown in the table below. We can see, most of the time the average reliability was more that 75%, however, in some cases there was a question whether procedural knowledge or cognitive process of compare and contrast applies. For the cases that are marked by N in the table 3-8, one of the inter-raters disapproved that trait occurred. As such, listed reliability in the table 3-8, is the average of the scores rated by two other inter-raters.

Table 3-8 Inter-rater reliability overall

Factual	Conceptual/ Schema	Procedural	Understand/ Compare	Understand/ Infer	Understand/ Explain	Apply
86.4%	86.2%	77% & *N	72% & *N	75%	75.4%	77%

3.9 Summary

In this chapter, we focused on the goals of inquiry that we believed were relevant to assessing students' reasoning and described the type of assessment that would elicit different levels and qualities of proficiencies that define a well-reasoned response. We refined and combined several theories to make our goals more specific and contextualized so that we were enabled to design a protocol to evaluate students' responses.

Based on previous research of others, we discussed how we could control the level of thought processing and abstraction and design questions that would probe specific hierarchies of thought processing and multi-level conceptual construction. Consequently, we devised questions in different disciplines whose answers would require the same levels of cognitive processing, types of knowledge, and structural forms for the concept construction. In this way, the processes of thinking would be similar for different comparison groups, even though the contexts were not.

Aligned with the goals that we used for our question design, we developed a rubric that analytically examined those responses through different lenses. Our rubric describes seven traits that we consider to be evidence of students' reasoning.

Chapter 4 - Applying the Question Template and Rubric

In the previous chapter, I described the methodology that we developed to elicit and analyze students' reasoning. We provided instructions for posing specific types of content questions with special characteristics designed to elicit the students' thought processes and probe their ability to construct knowledge in new situations.

The weight we could assign to the students' arguments depended on the quality of the questions we asked. If we designed question that did not probe for a certain kind of knowledge, then the students would not have made an effort to learn that kind of knowledge. To explain the effectiveness of the different types of questions and the features of their design I discussed examples of the content questions that we administered to the target students. To illustrate the full range of questions we developed, I selected an example of a question we used for each discipline.

In this discussion, the "quality" of the question means the degree to which the question's structure is aligned with the content question design protocol I explained in the previous chapter. The product of our analysis is meant to be a measure of the students' reasoning skills that effectively assesses students' prior knowledge and the extent to which higher levels of reasoning were activated as manifestations of knowledge having been transferred. Consequently, a well-designed question would also need to have accounted for a level of prior knowledge that would be deemed appropriate. The degree of the student's familiarity with the content is also an important indicator of the likelihood of the student's higher-level thought processes having been engaged. As Vygotsky's notion of *Zone of Proximal Development (ZPD)* asserts, we believed that the higher-level thought processes would be activated when a novel problem was constructed based on scaffolding on a familiar scientific context so that students could relate the content of questions to their prior knowledge.

For the questions discussed below, I also reported examples of students' responses. I tried to select responses that were representative of the categories of students' reasoning levels. The categories emerged from each set of data when we employed our rubric to code the responses. For each of the sample responses, I have explained how we used our rubric to interpret the responses in relation to specific knowledge or skills that the students possessed or the cognitive processes that the students performed.

One limitation in our study was that we could not always administer questions that agreed with our protocol. For example, at some institutions questions on final exams were strictly designed by the faculty, and some professors were not willing to use questions based our template. Therefore, some schools administered questions with lower-level cognitive demands, which were not desirable for our study. We excluded questions that did not fit our protocol. In spite of this limitation, in many cases, we controlled the question design process and higher-level cognitive demands were involved in the questions. As a result, we collected enough data, and got to the point of data saturation.

4.1 Examples of four typical questions

Questions were designed in four disciplines: physics, biology and genetics, geology, and positional astronomy. In this section, I discuss the cognitive loads and kinds of knowledge that we attempted to examine with these questions. In addition, using Lawson's (2000) and Nieswandt and Bellomo's (2009) classifications, I discuss the types of conceptual structures the questions elicited and how they were embedded in each of the students' responses.

4.1.1 *Question 1-Moon Phases*

The following question was given to a traditionally taught astronomy course for non-science majors at a small midwestern university. Seventy-eight students answered this question on their final exam.

Q1: In the middle of the night, a student notices a quarter moon rising due east. Remember the earth rotates counterclockwise. Is this the first quarter or third quarter of the moon? Explain how you can tell. Your explanation may include a diagram.

Type of knowledge required:

<u>Factual Knowledge</u>

Knowledge of the various moon stages, the earth's spin and moon's orbit, the sunlit and shadowed portions of the moon and the earth are considered as basic entities needed for this question. These basic entities are like a corner stone for constructing other types of knowledge in answering this question.

<u>Conceptual Schema</u>

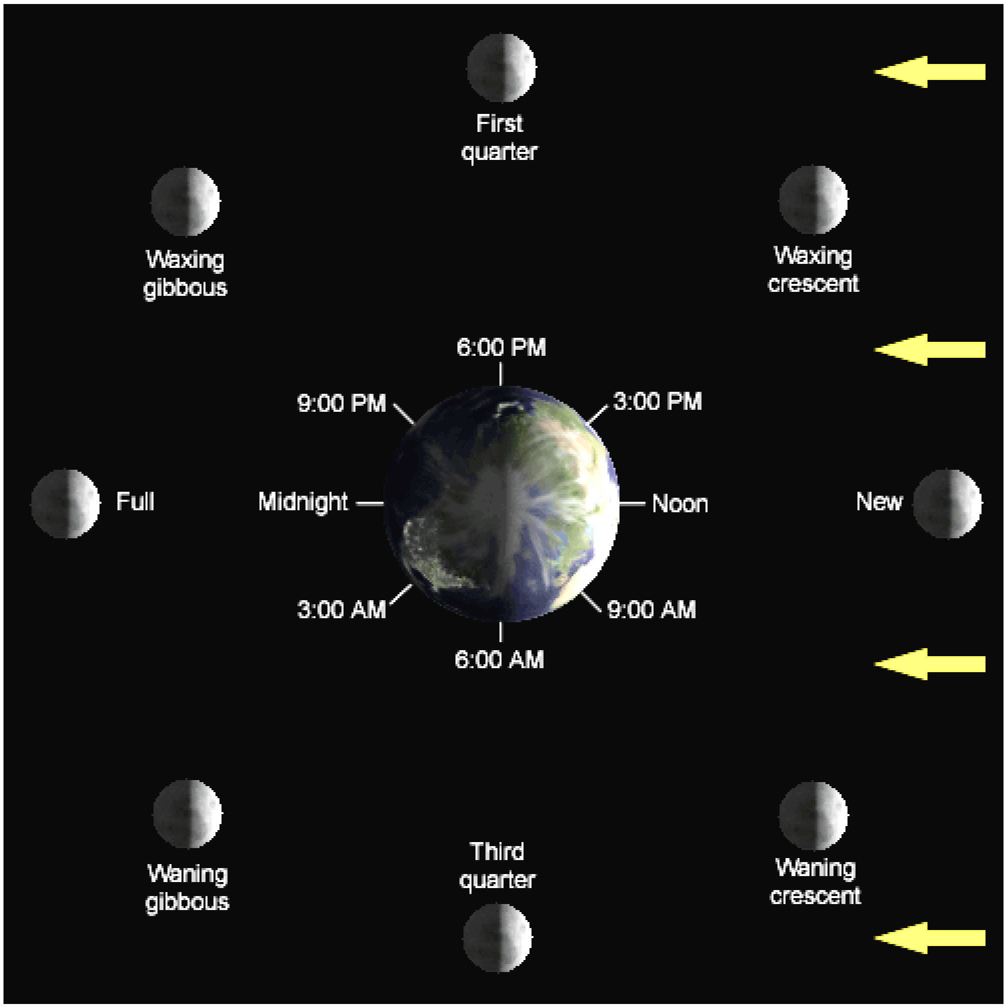
The development of conceptual schema requires constructing a mental model that consists of several interconnected concepts. For example, one can elaborate on the discrete factual knowledge of the sunlit and shadowed portions of the various stages of the moon to recognize the systematic sequences of the increasing and decreasing sunlit portions of the lunar cycle.

Procedural Knowledge

One of the most distinctive aspects of this task is being able to visualize the geometrical configuration of the sun, earth and the moon cycle. This requires a mental image of the relative positioning of the sun, moon and the earth during the various stages of the moon cycle. In addition, this task requires being able to understand the angles between the lines of sight with respect to the earth, the moon, and the sun. As the moon revolves in its orbit around the earth, these angles change constantly, the effect of which is the emergence of the various stages of the moon. To understand the moonrise and moon set phenomena, one should consider that the Earth-Sun-Moon geometry constantly changes. When the earth rotates on its axis at different time intervals the line of sight of the moon passes over different regions of the earth. The moonrise is the starting moment for this time interval, and the ending moment is when the moon set occurs.

This question also requires the skill of using cardinal directions to point to the orientations of the north, south, east, and west. This means that students must know east and west are at right angles to north and south, with east being in the direction of rotation of the earth and west being the opposite as shown in Figure 4-1-4 below.

Figure 4-1-4 Lunar Phases (Lunar Phases, Wikipedia, 2011)



Type of cognitive processing required:

<u>Compare and Contrast</u>

First, students needed to understand the following: 1.The Earth-Sun-Moon geometry constantly changes. The first and third quarters occupy different locations in the moon's orbit. 2. The moonrise and moonset occur at different times depending on the moon's phase. 3. The time of occurrence and the location of the phases depends on the positioning of the earth and moon with respect to the sun.

To proceed to the higher-level cognitive process of making inferences, the students would need to be able to distinguish how features that are related to first quarter are different from the third quarter and then relate those differences to the information given in the question.

Infer:

After distinguishing the features that are related to first and third quarter of the moon against the given information in the question, students would need to proceed to the cognitive process of inference to recognize and choose the answer that combines well with the other premises given in the question.

Explain

While students in some cases might insightfully infer a correct conclusion, they would not necessarily have been transparent with respect to the steps of their thought processes. The cognitive process of being able to explain in this question would be exemplified by their reflections concerning their knowledge of "What" and "Why" and "How". Their knowledge of "What" would be demonstrated in discussion showing their understanding the subject. Their knowledge of "Why" and "How" would be demonstrated by the quality of their discussions of the cause-and-effect relationships and the ability to explain how and why differences may cause different effects. For example, moon stages being located in various geometrical positioning may have different orientations with respect to the earth, which in result, first and third quarter moon rising at different times.

Apply

In a more progressive argument, students would be able to_use the constructed model of relative positioning of the Moon-Earth-Sun geometry and match the model to the context of the question. Using the given information, students would be able to locate an observer on the Moon-Earth-Sun model when it is midnight and think about the question from the perspective of the observer.

Type of conceptual structure required:

Using Lawson et al.'s (2000) definition of concept categorization, we classified the different phases of moon as descriptive concepts. However, the systematic sequences of the increasing and decreasing the sunlit portions of the lunar cycle would not be able to be seen until the phases had been observed and recorded. Although the phases can be seen directly, the trend in the changes in phases cannot be seen unless the single phases are observed and compared during the entire cycle. For example, by observing all the phases, a student would be able to detect differences in the shapes of the moon phases and then s/he would be able to create a model for the moon cycle. Even though individual phases can be seen, creating a model for entire cycle, involves comparison, organization and inference of visual information.

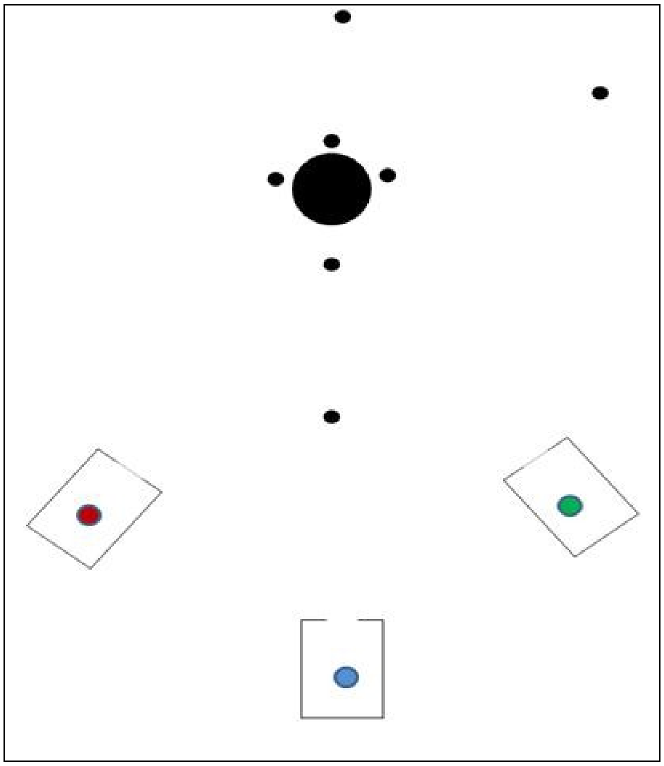
Earlier in Chapter 2, we explained that by modifying Lawson et al.'s definition of concept categorization slightly, we could expand their definition of hypothetical concepts to include those that can be measured or observed indirectly, such as voltage, magnetic fields, electric fields or concepts that can be seen through a model (McBride et al., 2010). Therefore, the concept of lunar cycle is a hypothetical concept. Similarly, the Moon-Earth-Sun geometry and the configuration of the Earth and the Sun or the concept of Earth's revolution are types of concepts that cannot be seen directly either. Yet, one can visualize these concepts using constructed models such as diagrams, physical models, or multimedia visual aids. Therefore, concepts such as moon cycle, earth's rotation or

Moon-Earth-Sun geometry can be classified as hypothetical concepts. To explain their answers to this question, students needed to be able to explain the relationships among the moon's phases, the Moon-Sun-Earth geometry and the Earth's rotation. This meant that, the types of conceptual links applied in this question combined both descriptive and hypothetical concepts.

Question 2 Light and color

The following question was administered in a reformed physics course designed for non-science majors at a small midwestern university. Thirty-eight students answered this question on their final exam in the course.

Q2: Three light sources are placed on a piece of white paper on top of a table. One of the light sources produces a red beam, another blue, and a third green as shown. The beams are aimed toward a vertical rod that blocks light. The beam of light from each light source falls on both the rod and on the white paper on the table. (Your may treat the light bulbs as points sources of light.) At each labeled point what color will you see? Explain your answer.



Type of knowledge required:

Factual Knowledge

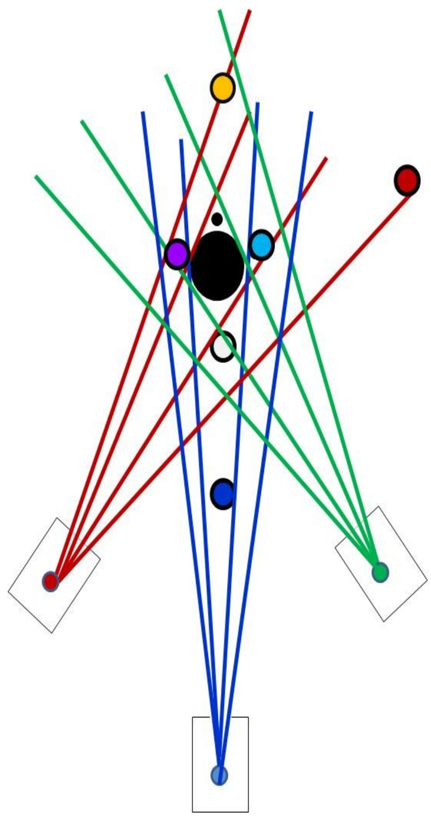
At the factual level, this question required the knowledge of RGB additive light color model that defines secondary colors as combinations of primary colors.

Conceptual Schema

At the conceptual level, this question required that students be able to understand why some regions were illuminated and others shaded and that the shaded regions were created by the blocking rod. Moreover, students needed to realize that the color at any point was a combination of colors received from the different light sources illuminating that point.

Procedural Knowledge

At the procedural knowledge level, this question required that students be able to sketch the light beams that emerged from the light sources, so they could distinguish where the light from each source of different color spread and where the light of that color was obscured in order to distinguish which regions were shaded from or illuminated by each light source--red, blue, or green.



Type of cognitive processing required:

Compare and Contrast

We did not expect students at this level to distinguish the darker and brighter parts of the shadowed regions. They also lacked knowledge of light intensity and the inverse square law. Therefore, we did not expect them to apply the cognitive process of comparing and contrasting in this question.

Infer:

In this question, students were required to predict the colors of the different regions that would be illuminated by different colored light sources. This act or process of inferring required deriving logical conclusions by ascertaining the links of the cause-and-effect chain based on the following cause-and-effect relationships:

- Discerning the shadowed and illuminated areas by tracing out the light rays that emerge from each particular light source.
- Identifying shadowed regions as places where the original light from the colored light source could not reach due to the obstruction of the rod.
- Explaining how two primary light colors can combine to create the secondary color at a given point.

The performance of the above thinking skills would allow the students to predict the colors of the different regions.

Explain

The knowledge of "What", "Why" and "How" in the context of this question meant that students would be able to deduce the colors of the different regions illuminated by the light sources of the different light colors. To show how and why regions would be illuminated or darkened, students would have to demonstrate that they trace out the light rays that emerged from each particular source and find the regions that were exposed to those light rays.

Apply

In this question, the cognitive process of apply would be manifested if students were able to trace out the spreading of the light rays in three different orientations along with corresponding geometrical shadow of the obstruction. Therefore, to proceed to cognitive process of apply, they would need to understand the concepts of shadow, illumination, the RGB additive light color model and have the procedural knowledge to create a geometrical ray diagram of the configuration given in the question.

Type of conceptual structure required:

The concepts of shadow, illumination, and light color combination are all observable phenomena and, therefore, fall into the category of descriptive concepts. However, the process of separating shaded and illuminated regions cannot be observed directly, but would only be visible with the aid of a geometrical ray diagram. Sometimes, non-observable concepts can become visible with the aid of laboratory experiment set-ups. Clearly, in this case because students were working on this question during their final exam, they did not have access to experimental set-ups. Consequently, in this question, we considered the process of discerning shaded and illuminated regions to be hypothetical concepts.

Question 3: Predicting the color of the plants based on their genes

The following question was administrated to the students in three biology courses. One of the courses was a traditional large lecture course and the other two were smaller, located at another institution, and reformed to different extents.

Q3: You are given four plants. They are all the same species but have different seed pod shapes and flower colors (yellow or white). You wish to determine which variations of flower color and pod shape are the dominant variations. You begin by breeding the plant with the swollen pods and yellow flowers to the plant with pinched pods and yellow flowers. The result is some plants with swollen pods and white flowers and others with swollen pods and yellow flowers.

A) What hypothesis can you form regarding which traits are dominant and which are recessive from this first cross? Justify your hypothesis based on the evidence

B) Next, you breed a plant with pinched pods and yellow flowers to a plant with pinched pods and white flowers and examine 8 plants produced by this cross. The result is all eight plants with pinched pods and white flowers. Does this result support your hypothesis? Does it disprove it? Explain your answers.

Type of knowledge required:

<u>Factual Knowledge</u>

At the factual level, students may refer to discrete pieces of information that are basic constituents of genetics and heredity. For example, we look to see if specific content elements such as allele, genes, phenotypes, genotypes and heterozygous, homozygous, recessive, and dominant alleles were mentioned.

Conceptual Schema

At the conceptual level, understanding the concepts of genotype, phenotypes, and alleles was essential prerequisite knowledge for being able to answer this question. Moreover, understanding the characteristics attributed to the alleles such as recessive and dominant, and dominate traits mask the recessive traits was also essential. Furthermore, students needed to understand the principles of the breeding that associate types of alleles in phenotypes with manifestations of properties or behaviors that appear in the phenotype. For example, students would have needed to understand that if the offspring of the plant discussed in the experiment above had a recessive allele for white and a dominant allele for yellow, the color of the plant would be yellow.

By developing a conceptual schema, students should have, consequently, been able to develop ideas about breeding the plants and the biological traits that would be passed to the offspring. For example, the breeding of two phenotypes may have different outcomes that depend on the genotypes that participate in the interaction. When the genotypes of parents come together, they come in pairs of alleles and as they meet, they split from the original pair and join with a new pair. As a result, one of the two alleles of the new genotype would be manifested. In other words, the types of genotypes (heterozygous, homozygous) and types of alleles (dominant, recessive) taking part in the breeding determine the product that is the manifested phenotype.

Procedural Knowledge

The procedural knowledge required for solving this problem is the mathematical skill of finding probabilities. This skill can be identified as being able to find the joint probability or combinations of possibilities that would correctly predict the probabilities of the possible events, each being probabilistically independent and each having an equal chance. Consequently, students would need using the joint probability mathematical model to find the various combinations of the genotypes in which different types of alleles might be participated. Likewise, students could apply another type of procedural knowledge in which they demonstrate the skill of using Punnett Square diagram that is equivalent to a joint probability mathematical model. The Punnett Square is a 2×2 matrix that shows every possible combination of two homozygous genotypes, in this case of those being studied in the cross.

Type of cognitive processing required:

<u>Compare and Contrast</u>

After developing conceptual schema and procedural knowledge of the possible alleles that participated in the breeding, students should have been able to state the possible occurrences of the phenotype and predict the possible traits that could appear in the offspring. However, if students know with certainty, which alleles participated in the breeding, they would have been able to determine with certainty, which traits would appear in the offspring. In this question, the breeding already resulted in certain traits being manifested in the offspring. Therefore, students should have been able to compare the phenotypes that appeared to the possible outcomes that would be predicted for the traits and eliminate the possible occurrences that did not happen in the breeding, to determine the type of interaction that actually happened.

Infer:

After comparing the outcomes to what would have been predicted possibilities, the students should have been able to infer which alleles actually participated in the breeding. Among the possible outcomes, one possible scenario has one parent being homozygous dominant and the other homozygous recessive, in which case all of the offspring would display the dominant trait. Another possible scenario has parents being heterozygous for a trait, in which case offspring will be produced with some displaying the dominant trait and some displaying the recessive trait in a 3:1 ratio respectively. Comparing the results from the first experiment to the predicted possibilities yielded the interaction below for the type of pods:

- Swollen pod parents (*PP*) x pinched pod parents (*pp*) → all swollen pod offspring (*Pp*), swollen dominates over pinched

Comparing the results from the first experiment to the predicted possibilities also yielded the interaction below for the type of flower color:

- Yellow flowers (*Ff*) x yellow flowers (*Ff*) → offspring with white (*ff*) and yellow flowers (*FF* and *Ff*)

Comparing the results from second experiment to the predicted possibilities yielded the interaction below for the type of pods:

- Swollen pod parents (*PP*) x pinched pod parents (*pp*) → all swollen pod offspring (*Pp*), swollen dominates over pinched

Comparing the results from the second experiment to the predicted possibilities yielded the interaction below for the type of flower color:

- White flowers (*ff*) x White flowers (*ff*) → offspring with yellow flowers (*FF*)

Pod shape in this cross-supported the hypothesis that the parents in this situation are homozygous recessive if the pinched trait is recessive.

However, flower color in this situation did not support the hypothesis being posed. If the yellow flower color were a dominant trait, then the yellow flowered plants would have needed to be either homozygous dominant (*FF*) or heterozygous (*Ff*);and plants with white flowers would have needed to be homozygous recessive (*ff*). In this case, plants with yellow flowers were crossed with plants with white flowers, all the offspring are white, and no yellow flowered plants appear. Therefore, students would need to infer that something other than simple Mendelian dominance must have been at work.

Explain

The process of explanation in this question referred to the students' ability to show their thought processes in meaningful way that connected the cause, the breeding of the genes, and the effect, and the outcome of the cross. In an insightful argument, student would have needed to identify the factors that affected the chain of cause and effect. For example, the observable traits, such as the manifestation and appearance of the phenotype, would need to have been explained in association with sets of rules according to which information of the genotypes was encoded.

Apply

Thus far, we have considered the different types of knowledge and cognitive processes that were required to be applied in the context of question only. The higher cognitive processes of application required that the students interpret the process of the experiment in terms of Mendelian principles using the concept of genes. Furthermore, the process of application activated in this question included scaffolding on a mathematical model of joint probability distribution and using the mathematical model in a new context. In this scaffolding, students needed to be able to both construct the model and then apply it to the context of genes to find the combinations of genotypes with different types of alleles to predict the possible outcomes. However, because a set of specified conditions could be inferred from the question, students could select among the possible outcomes and eliminate the rest.

Type of conceptual structure required:

This question required that students understand genes, alleles, genotypes, phenotypes, and types of alleles, which are all theoretical concepts. The appearances of the phenotype, in this question being flower color or type of pod were observable concepts that fall into the descriptive category.

Question 4 Predicting the direction of the wind based on the surface features of the land

Q4: Answer the following question using complete sentences. You may use diagrams to illustrate your points. Be sure to explain any diagrams. Recently an extra solar planet was detected. The planet was most similar to the earth yet discovered. A (hypothetical – not actually done) probe on the surface measures the wind direction from east to west during the planet's day and west to east during the planet's night. Using your knowledge of earth's winds, "describe and explain" what surface features you might expect to find near the probe?

Type of knowledge structure required:

Factual Knowledge

At the factual level, this question required that students be familiar with specific content elements that included temperature, pressure, cold air, and warm air convection.

Conceptual Schema

At the conceptual schema level, this question required that students understand convection currents produced by unequal heat absorptions rates of seawater and land. The sea has a greater heat capacity than land and therefore absorbs more heat than the land, so the surface of the sea gets warm slower than the land's surface. During the day, the temperature of the surface of the land rises; the land heats the air above it. The warm air is less dense and so it rises. As a result, the cooler air over the water replaces the warm risen air over the land. The conceptual schema also refers to understanding of the features attributed to the air pressure and the differences in the air pressure and their relation to the wind's directions, meaning that air (or gases) flows from lower pressures to higher pressures. As the air over the land rises, its pressure decreases, but the air above the sea is cooler and has higher pressure. As a results air current flows towards the land into the lower pressure, creating a cooler breeze near the coast.

Following the same chain of reasoning, students could argue that at night the direction of airflow is reversed because the sea is warmer than land.

Some students may discuss the situation from the perspective of change in altitude, because pressure differences could also be associated with changes in altitude. As such, higher altitudes have lower pressures and vice versa. Yet, in this question, the concept relating different altitudes to differences in pressure could not be used to explain why the change of the wind direction was happening.

Procedural Knowledge

No procedural knowledge was required to answer this question.

Type of cognitive processing required:

Compare and Contrast

Students would have needed to be able to compare the differences in the characteristics of planet's surface properties such as varying specific heat capacities for different planet surfaces and discuss the consequences that followed from For instance; the surface of the sea and the surface of the shore were two examples of surfaces that have different heat capacities. This meant that when a surface had a smaller specific heat capacity, it could be warmed more easily and would cause the air above that surface to rise more easily.

Infer

The process of inferring in this question related to being able to determine the factors that affected the chain of events. Students would have needed to be able to find links between surface properties and the factors that controlled air pressure and determined the direction of airflow.

Explain

The process of explanation in this question required the students to be able to describe the surface features they considered important and describe meaningfully the connections between the different characteristics of land surfaces and the factors that affected the flow of air and its direction.

Apply

Application in this problem was considered to be the students' ability to use the different properties of the planet's surface structures that affected airflow, pressure differences and wind direction.

Type of conceptual structure required:

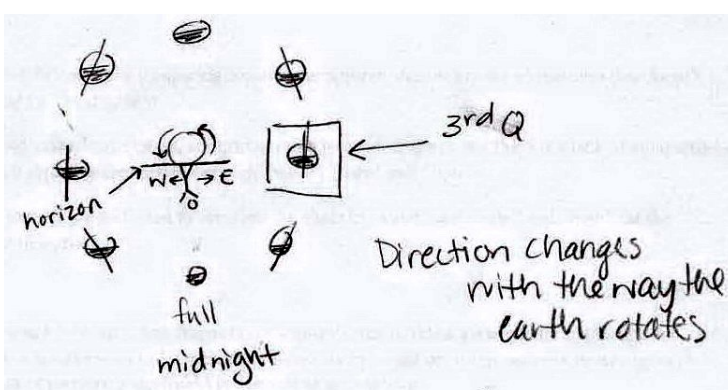
The concepts pressure and conduction are both examples of hypothetical concepts. On the other hand, the direction of the wind, hot and cold weather, and the surface features of the planet were descriptive concepts. The concept of specific heat capacity, which explains why different features have different temperatures during day and night, is a theoretical concept.

4.2 Scoring the responses using our protocol

In the previous sections of this chapter, I discussed four examples of content questions that were designed in different scenarios for different disciplines. In this section, for each question, I report on a few examples of the students' responses. I have tried to select responses that were representative of the categories of the students' mastery level. These categories emerged from each set of data to which I employed our rubric to code the responses. In addition, for each sample response, I have explained how we used our rubric to interpret the response in relation to specific knowledge or skills that the students possessed or the cognitive processes that the students performed.

4.2.1 *Sample responses for the moon phases question*

Response 1) *The moon that rises in the east at midnight is the 3rd quarter moon. The moon would be 1st quarter if it were setting in the west at midnight.*



Response 2) *I know that a first quarter moon is highest overhead at 6 pm and a third quarter moon is highest overhead at 6 am. If the student saw a moon rising in the middle*

of the night, it would have had to be a third quarter moon, because the 1st quarter moon is setting.

Response 3) *Third quarter moon, if it were a first quarter moon then it would not rise in the middle of the night. It would be daytime.*

<u>Evidence of knowledge types:</u>

<u>Factual Knowledge</u>

The first response indicates that student is familiar with various stages of the Moon, the Earth's spin, the Moon's orbit, the sunlit portions of the Moon and the shaded portions of the Moon. Therefore, the basic facts that are required for constructing other types of knowledge are accessible to this student whereas, the second student referred to a limited number of facts including the Moon's stages and the specific times of their appearance in the sky. The third student knew even fewer facts. With limited factual knowledge, the second student was not able to support an argument to justify why two of the Moon's phases appear highest overhead at certain times and the third student could not justify the prediction that the Moon in third quarter appeared on the left side of the diagram that the student drew. However, students demonstrated enough factual knowledge that allowed them to proceed to conceptual schema.

Conceptual Schema

The first student was able to elaborate on the factual knowledge of which of the illuminated and shadowed portions of the Moon appear during the various stages and demonstrated an understanding of the concept of the systematic sequences of increasing and decreasing sunlit portions of the moon during the lunar cycle. In the student's diagram, the sunlit portions of the moon were clearly facing the Sun and the observer on the earth at midnight was located opposite to the Sun. In addition, the student ascertained the relationships among the cardinal direction, the observer's location and the Moon's location during the lunar orbit and the Moon's phases.

The second response indicates that the student was only able to associate moonrise/moonset times with various stages of the Moon's phases. The third student drew the sunlit portions of the Moon so they were clearly facing the sun and the observer at midnight as being located opposite to the sun. In addition, the student associated the moonrise/moonset times with the different phases of the moon. Therefore, all of the students showed some level of conceptual schema.

Procedural Knowledge

The first student represented the geometrical configuration of the Sun, Earth and Moon cycle and demonstrated an understanding of the relative positioning of the Sun, Moon and Earth for the various stages of the lunar cycle. Furthermore, in that student's diagram, we can see that the orientation of the Moon's axis changes as it orbits the Earth. In addition, that student knew how to use the cardinal directions—north south, east, and west--and knew that, knowing east and west are at right angles to north and south, with east being in the direction of rotation and west being directly opposite. Obviously, the second and third students' responses showed no evidence for this knowledge and these skills. Neither student appeared to possess knowledge of the Earth-Sun-Moon geometry, which is the geometrical configuration of different phases of the Moon in relation to the Earth and the Sun.

Evidence of cognitive processing:

Compare and Contrast

The first student fully compared the relative locations of the Moon in the first and third quarters and showed how the alignments of the Moon, Sun and Earth affect how much of the Moon's sunlit portion appears to the observer. Moreover, the student appears to have related the change in the location of the Moon to the way the Earth rotates.

The second and third student both compared some of the properties related to the appearance of the Moon in the 1st and 3rd quarters and associated high overhead locations of the Moon as well as moonrise events to the specific times. However, neither compared other features that would explain why the Moon rises or appears high overhead at those times. Nonetheless, all of the students showed some level of mastery in using the cognitive process of compare and contrast.

Infer:

To evaluate how well the process of inferring was activated in this question, we looked at the cause-and-effect relationships that the students depicted for the chain of events. For example, the first student based inferences on the following:

- The student's own sketches of the geometrical positioning of the Sun, Earth and the Moon and as they appear the first and third quarter in the student's configuration
- The positioning of the observer at midnight on the student's Sun-Earth-Moon model
- The direction of the rotation in relation to east and west
- The use of cardinal directions

Building these as premises, the first student ascertained cause-and-effect relationships that led to the answer's conclusions.

On the other hand, the second student drew conclusions based on the following statements:

- The first quarter Moon is highest overhead at 6 pm.
- The third quarter Moon is highest overhead at 6 am.
- The Moon rises in the middle of the night.
- It would have had to be a third quarter Moon.

Although the statements are related, albeit weakly, the initial statement is an unsupported argument based only on memorization. The student has not indicated any event that would cause the Moon to appear at a certain time.

The third response includes a few informative statements about the Moon's phases appearing at certain times but no indication that any event has caused another event.

<u>Explain</u>

Students may insightfully come to a correct conclusion, but from the perspective of inquiry learning theory (Bybee, 2000). An appreciation of how they know, what they know and why they believe in their prediction is an essential quality of reasoning skills.

The first student demonstrated an initial understanding of the positions of the Earth, Moon and Sun relative to each other and was able to recognize the different factors that determined the location of the Moon in the first and third quarters. Furthermore, that student supported the argument so that it could be used plausibly to show how those factors would affect the moonrise and moonset events. However, the second student's statement relating the Moon's phases to their positions in the sky at specific times did indicate of knowledge of "How" and "Why" and showed no more than an ability to recall certain information.

Apply

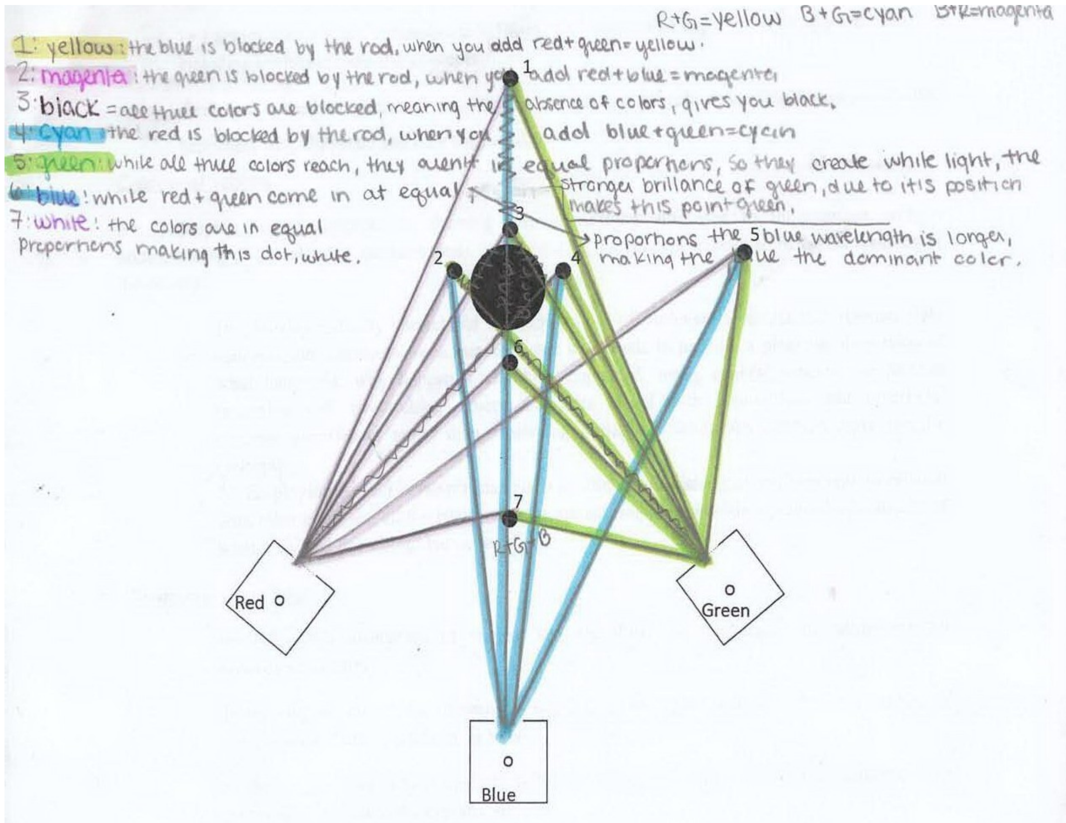
The first student constructed a model of relative positioning of the Moon, Earth, and Sun and matched it to the context of the question. Using the information that was given, the student located the observer's position at midnight on the Moon-Earth-Sun model and from the perspective of the observer discussed the question. The second student also applied the conceptual schema though to a lesser extent, and associated the motion of the moon from moonrise to the Moon's position high overhead. The third response failed to make any association of the concepts and procedures to the question.

Evidence of conceptual structure:

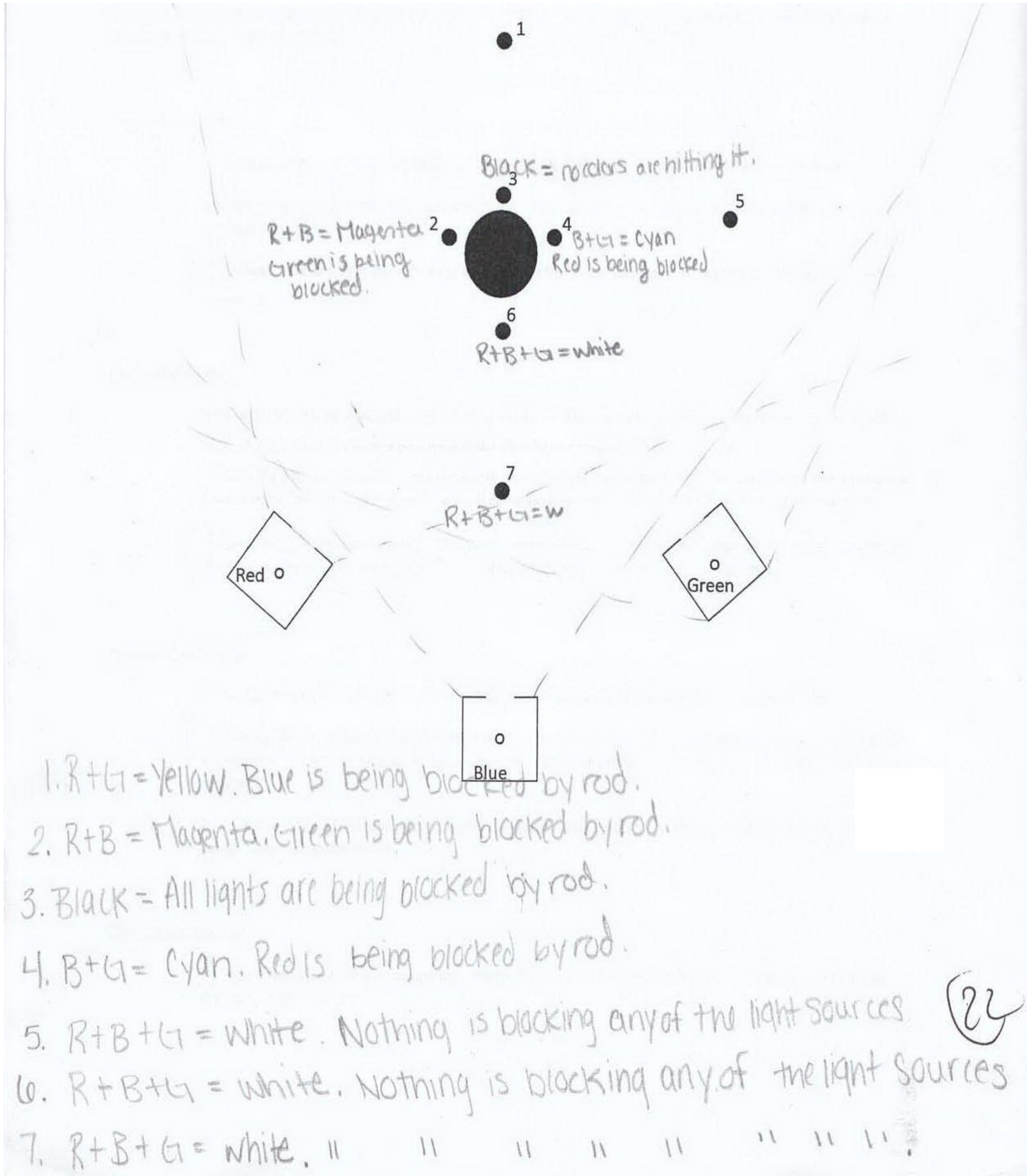
The first student connected three concepts: the appearance of the Moon's phases, the cycle of the Moon's phases, and the Earth's rotation. Both the cycle of the Moon's phases moon- phases and the Earth's rotation are hypothetical concepts because they cannot be seen directly, but they can be visualized with a model ; on the other hand, the appearance of the Moon's phases is a descriptive concept. The second and third student associated the Moon's phases with the time of their appearance, which is a descriptive concept.

4.2.2 *Sample responses for the light and color question*

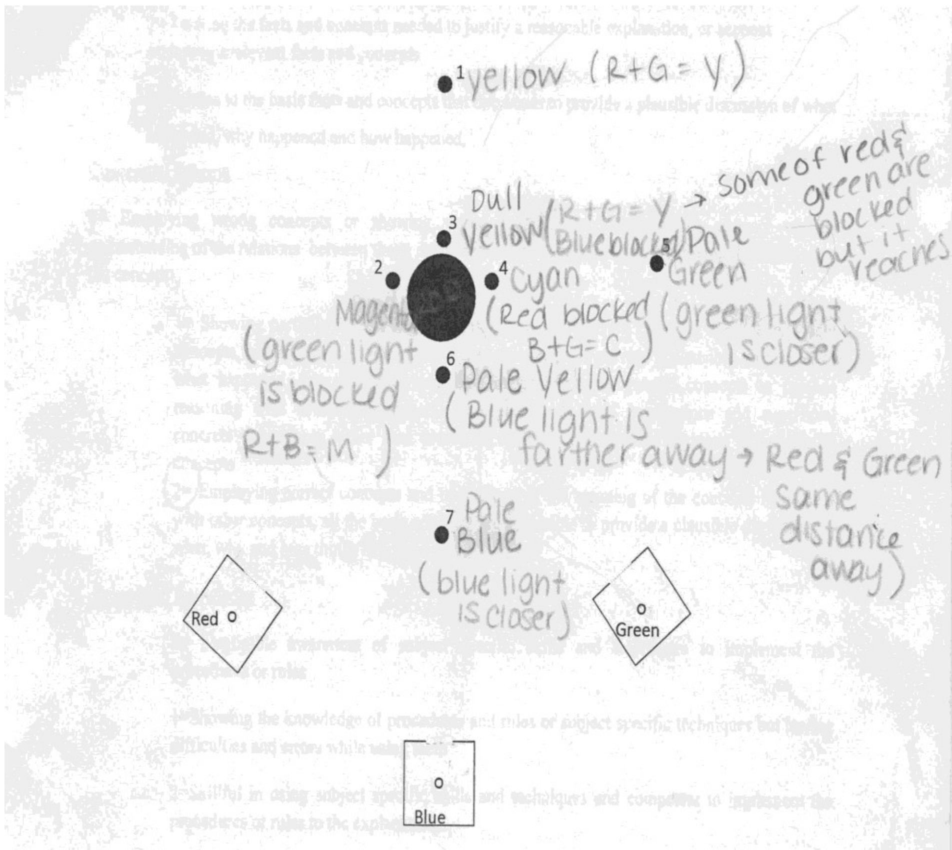
Response 1)



Response 2)


1
Black = no colors are hitting it.
3
R+B = Magenta 2
Green is being blocked.
4 B+G = Cyan
Red is being blocked
5
6
R+B+G = white
7
R+B+G=W
Red
Green
Blue
1. R+G = Yellow. Blue is being blocked by rod.
2. R+B = Magenta. Green is being blocked by rod.
3. Black = All lights are being blocked by rod.
4. B+G = Cyan. Red is being blocked by rod.
5. R+B+G = white. Nothing is blocking any of the light sources
6. R+B+G = white. Nothing is blocking any of the light sources
7. R+B+G = white. " " " " " " " ".
22

Response 3



Evidence of knowledge types:

Factual Knowledge

Based on all three responses, we can conclude that the three students possessed knowledge of the Red-Green-Blue (RGB) additive light color model that allowed them to make informative statements about combinations of primary colors and their resulting secondary colors.

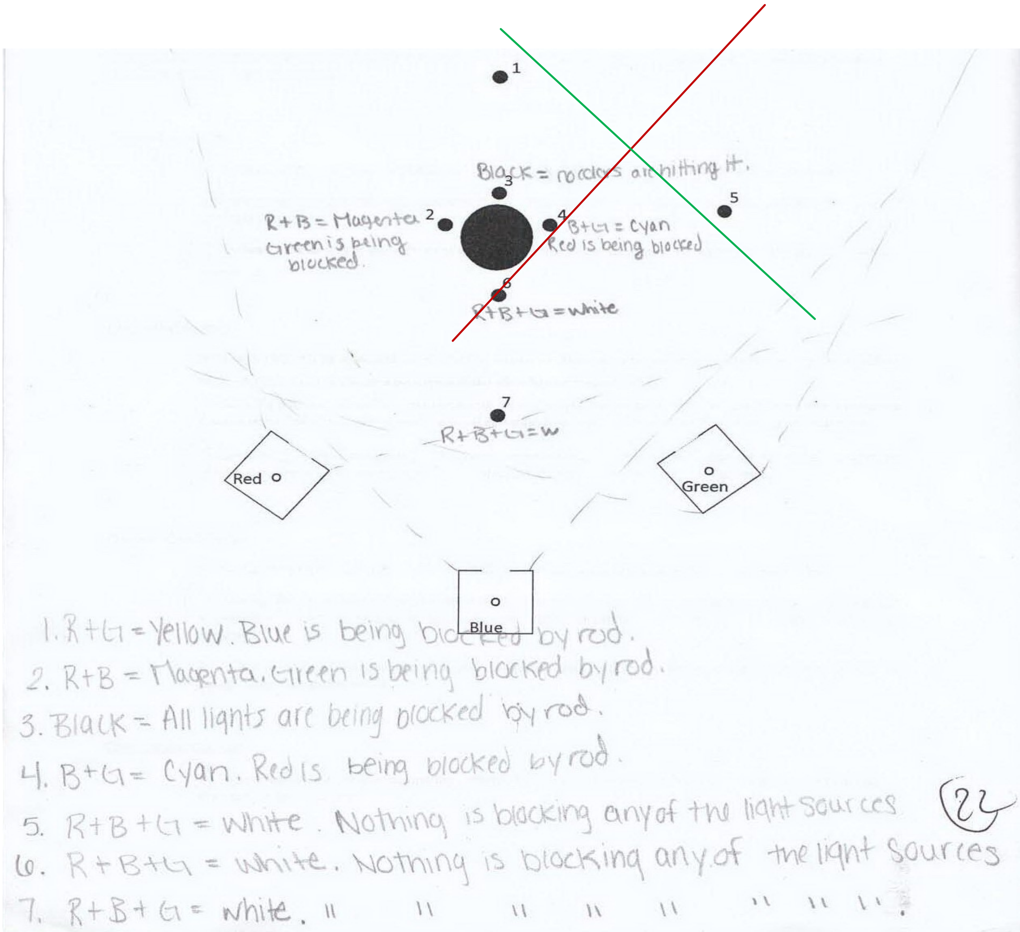
Conceptual Schema

We can see that in all three responses, students blocked one or more colored rays under the shadow of the rod. Moreover, they combined the colors of light beams landing from different sources illuminating the given point. Therefore, evidence existed that the students had a sufficient conceptual understanding for them to be able to answer this question.

Procedural Knowledge

As illustrated in their responses, the students showed diverse levels of proficiency in demonstrating procedural knowledge. The first student was able to sketch the light beams that emerged from the sources to find where the light source spread and where it was obstructed in order to distinguish the dark and illuminated regions. Therefore, we can conclude first response showed an adequate level of proficiency in applying the geometrical ray model diagram.

To find the illuminated regions, the second student traced the spread of the beams based on the assumptions the student made about the way the rays were extended from the apertures, however, the student did not provide details that would explain how some of the color beams were blocked in the indicated regions. In other words, in the second student's diagram, the beam fields that student sketched are wide and extended, so the way the rays spread did not account for the regions that were marked shadowed and those regions were not actually blocked and could be illuminated by the light beams that student sketched. In the diagram below, I added two green and red beams to the student's' response, which shows examples of contradictions.



The third student did not provide any evidence to support the student's claim that some of the regions were exposed to certain light colors and others blocked for some of the other colors.

Evidence of cognitive processing:

Compare and Contrast

Based on the discussion that we had with the professor of this class, we learned that the students in this class were not familiar with inverse square law for the light intensity. As such, we did not expect the students to compare the light intensities based on the distances of the light sources from the identified regions. However, as illustrated in the examples, in a few cases students compared the proportions of the beams reaching the identified points based on their distance from the light sources.

<u>Infer</u>

The first student derived logical conclusions based on valid premises to predict the colors of different regions. Those premises used by the first student are evident in the steps, the student performed to answer the question and include the following:

- Tracing the light beams that emerged from each individual source
- Applying the geometrical ray model diagram to distinguish the shadowed regions from the illuminated regions for each color of light rays
- Coloring each illuminated point with the combination of colors that reached that point

Each statement by itself is a conceptual schema, whereas making an inference is defined as a process of finding a meaningful link that connects one step to another. We can see a successive and logical flow from one statement to another without gaps between cause and effect. However, the second student's response lacked the connections existing in the first student's first three steps.

<u>Explain</u>

The first student demonstrated the knowledge of "What", "Why" and "How" by discussing how and why some regions were illuminated and others were shaded. The student traced the lights rays that emerged from the light sources and found the regions that were exposed to those light rays. Also, the student demonstrated skill in determining which areas would be blocked for each light source to predict the colors of the different regions. The other two students discussed reasons for the combinations of colors;

however, their discussion failed to display knowledge of how the rays from the different sources were blocked.

<u>Apply</u>

By comparing the three responses, we can see the first student linked between the scientific concepts of shadow, light, color combination, color-coding, and procedural knowledge of geometrical ray diagrams in three different orientations; however, the other two students' discussions were based on an ineffective application of a geometrical ray diagram.

Evidence of conceptual structure:

The first student linked descriptive concepts such as color combinations and shaded regions with a hypothetical concept, the geometrical ray diagram. The second and third student stated that the rod, which showed that their reasoning did not go beyond observation, blocks light rays and they only used the concept of shadows and color combinations, which are both descriptive concepts.

4.2.3 *Sample responses for the wind direction question*

Response 1)

Near Probe: According to the data given, the probe could be located on or near the coastline of a large body of water. The rotating or changing wind could be related to the Earth's land and sea breezes. Where predominate winds flow from sea to land or from land to sea. For example, on the U.S. East Coast near the Atlantic Ocean the wind is more dominate from land to sea or West to East. This is causes by the Sun heating the earth's crust and radiant heat heating the surrounding air. This heated air then travels to *colder regions (over the ocean) where the heat is transferred through conduction to the water. This process then reverses during the night when the sun has fully set. The water is now heating the air and the earth's crust has become cooler than before. The heated air then travels from sea to land or east to west. The process continues everyday on earth which could be an explanation of the probes data collection.*

Response 2)

If there is this probe on a planet and the surface measures the wind direction differs from day and night, then this implies that the movement of the planets circulation around the stars alters or changes/ evolves. Thus during the day and night since the probe measurements are different it also must mean that there was a change in temperature and speed. I say this because if the motion of wind changed during the day and night, if it differed due to the time, then there must have been another variable such as temp and speed either decreased or increases. My hypothesis is if the motion of the satellite detected on this unidentified planet changed from east to west in the day times, then it has to do with the temp being greater in the daytime in relation to how much speed that is required for this east to west motion to occur. This is vice versa/opposite for the night. In addition, this change in temperature and speed, which causes an east to west motion in the day and west to east motion in the night, is connected to the positions of the satellite on the planet and the planet in relation to the sun. While these variables are unknown, the surface features you might see would be during the day cold and barren. It would….. Such as rain, snow all of this would be features during the day and at night it would be dry and hot because of the position. Thus, during the day winds might move clockwise *east to west during the day and counter clockwise west to east during the night. Circular motion similar to coriolis* [sic] *force.*

Response 3)

The surface features on this planet would be similar to earth in that there are mountains and valleys. The probe seems to have landed on a mountainside. On this planet, the wind blows downhill into valleys during the day and uphill back toward the mountains at night. This probe has landed on the western side of the mountain and is measuring wind going east to west downhill during the day and west to east uphill at night.

Evidence of knowledge types:

<u>Factual Knowledge</u>

The first student's response referred to some basic factual knowledge about heating, cooling and air currents. The second student's response considered other relevant factual knowledge concerning the planet's circulation, the speed, the temperature and the Coriolis force but failed to link these discrete notions in the context of the scenario posed by the question. The third student's response did not present any factual knowledge that could be applied to higher level cognitive processes.

<u>Conceptual Schema</u>

In the first response, student's understanding of convection had flaws and his/her statement that hot air goes to colder regions was wrong, but s/he demonstrated some understanding of heat transfer, conduction, and convection. Although the student did not explicitly mention convection, some evidence of knowing the concept of convection could be inferred from the student's association of the airflow with the air's temperature.

However, the student overlooked the concept of heat capacity, which would explain why the air over the ocean is colder than the air over the land in daytime and lower during the night. The second student presented an incomprehensible and implausible conjecture consisting of disconnected notions of speed and temperature, in which facts did not proceed to a conceptual schema.

Procedural knowledge

This question does not require any knowledge of specific skills, knowledge of criteria, or knowledge of specifically defined procedures.

Evidence of cognitive processing:

Compare and Contrast

The first student compared the differences in the heating processes during the night and during the day. However, a more insightful understanding of those differences would have compared the underlying factors that cause the temperature differences in the ocean, shore, and the air above the land and ocean.

Infer

The first student's response reflects some subtle connections between the cause and effect, which is the change of wind direction from day to night and variations in surface features that can cause this phenomenon. Below is the list of premises that the first student used:

- The Sun heats the Earth's crust.
- The Earth radiates heat to the surrounding air.
- The heated air then travels to colder regions (over the ocean), where the heat is transferred through conduction to the water.

- This process is reversed during the night when the water heats the air.
- The air and the earth's crust have become cooler than they were before.
- The heated air then travels from sea to land or from east to west.

Whether the student made an inference depends on how meaningful the connections that exist among the aforementioned statements are. Although student's third premise, about heated air that travels to colder regions is invalid, however, his/her reasoning is plausible. In this analysis our emphasis is on quality of inquiry and we did not discount students' responses for the lack of correctness. We can see that the student's descriptions provided plausible connections, between invalid premises of heated air that travels to colder regions and the predicted direction of wind. The chain of inferences that student made are all plausible except for a missing link that exists between the third and fourth statements, as is not clear why the air above the ocean is colder than the air above the land.

Explain

The first response presented a cause-and-effect model to explain the "Why" and the "How" of the direction change in the wind from day to night. Although, s/he wrongly concluded the wrong direction for the wind because s/he did not treat convection correctly, nonetheless_the argument is cohesive and leads to a reasonable conclusion, except that s/he failed to link causally the third and fourth statements.

Apply

By comparing the three responses, we can see that only the first student established connections between the scientific concepts and the context under discussion.

Evidence of conceptual structure:

The first student used concepts such as land, earth crust, water, wind direction, and heat, that all are descriptive concepts; however, the students also used hypothetical concepts such as conduction, convection, and the displacement of air from warm to cold regions. The second student's conceptual structure noted all types of concepts; however, the concepts were, disconnected and detached from one another. For example, the student used concepts such as the wind's direction, the temperature, and circular motion, which are descriptive concepts and others such as speed and Coriolis force, which can be considered hypothetical and theoretical concepts respectively. The third student referred to a selection of disconnected descriptive concepts, such as the land's surface features (in this case mountains and valleys) only.

.***Question 3 Predicting the color of the plants based on their genes***

Response 1)

A) Y=Yellow, y=white, S=Swollen Pod, s=pinched pod,

Possible genotype: S-Yy sSYy

	SY	*Sy*
SY	*Ssyy*	*SSYy*
Sy	*SsYy*	*Ssyy*

SsYY=Swollen with yellow flowers

SSYy=Swollen with yellow flowers

SsYy=Swollen with yellow flowers

Ssyy=Swollen with white flowers

My hypothesis is that swollen pods are the dominant plant type and that yellow is the dominant color. Each plant had to be carrying the recessive gene for white flowers, because white expressed in the offspring. However, it couldn't be dominant because each parent expressed yellow. If white was the dominant gene the yellow would get cancelled out, but because same white flowers occurred, we know the parents carried the recessive gene white

B) *No swollen pods were produced showing that each parent had the same genotype (ss) had either of them had dominant trait (S) the parent plant would be swollen. This supports my hypothesis. The flower color does not disprove my hypothesis but working at all, offspring would make my hypothesis more concrete. This paring of parents with my hypothesis of swollen and yellow being dominant would demonstrate a 1:1 ratio of yellow to white flowers. I would want to look at more offspring.*

	Sy	*Sy*
SY	*Ssyy*	*ssYy*
Sy	*ssyy*	*Ssyy*

Response 2)

A)

	SY	*sY*
SY	*SSYY*	*SsYY*

sY	*SsYY*	*ssYY*

Ss Yy *S-swollen* *Y-yello* *s-piched* *y-white*

The dominant trait is the swollen pods and yellow flowers. The recessive trait is the pinched pods and white flowers.

B) No the result does not support the hypothesis because now it's stating that the pinched pods and white flowers are dominant.

	SY	*Sy*
SY	*ssYY*	*ssYy*
Sy	*ssYy*	*Ssyy*

Response 3)

a) My hypothesis is that the swollen, white flowers are dominant, and the pinched pods yellow flowers are recessive, because when 2 swollen pods were crossed, they only produced swollen and when two yellow flowers were crossed they also got white.

b)Yes, this does support my hypothesis, because the pinched pods could have been homozygous, meaning there was no chance for swollen pods, and yellow/white was used for flowers, and white flowers were produced, therefore white flowers were dominant.

Evidence of knowledge types:

<u>Factual Knowledge</u>

All three responses demonstrated familiarity with terminology of allele, genes, phenotypes, genotypes, and heterozygous, homozygous, recessive, and dominant alleles.

<u>Conceptual Schema</u>

We can see also that all three students possessed knowledge of genotypes phenotypes, and alleles. They also considered the characteristics that are attributed to the alleles, such as whether they are recessive or dominant and how dominant traits mask recessive traits. However, only the first student associated the types of allele pair, with the characteristics manifested in the phenotypes. Nonetheless, all three students provided some evidence for having a conceptual schema.

Procedural Knowledge

The first and second students used the simple Punnett Square diagram to predict the joint probability for the outcomes of crossbreeding the genes. However, the third student did not find the possible gene combinations of the cross, consequently, he presented a hypothesis without a plausible supporting argument.

Evidence of cognitive processing:

Compare and Contrast

After using the Punnett Square, the first student identified all possible occurrences of the phenotypes that could be expressed from the crossing of the genotypes and alleles in the breeding experiment. Then the student compared the characteristics of the offspring with several possible expressions forecast by the Punnett Square and chose the one that matched the experimental observation given in the question. Even though the second student appeared to possess all of knowledge, that student did not perform any cognitive processes higher than recalling prior knowledge to support the proposed hypothesis in that student's response. In the same way, the third student did not use cognitive processes that lead to the construction of new knowledge based on connecting prior knowledge to knowledge gained from the question's scenario to form a hypothesis.

Infer

After comparing the phenotypes that resulted from the breeding to the predicted possibilities, the first student inferred which types of alleles had actually participated in the breeding.

Explain

The structure of the first student's response connected the knowledge of “What”, “Why” and “How” and was appropriate for an argument that was cohesive and led to a conclusion. The first response showed that the student understood the question, and what happened in the experiment. The first student’s response also reflected, the knowledge of how the events occurred as the response explained different possible ways that alleles can be paired and produce offspring with different traits. The first student also clearly understood why the events occurred by accounting for the conditions imposed by the experiment and explaining why the other possibilities should be neglected.

The two other responses provided weak evidence that students know what is happening; however, the knowledge of why and how things happened is missing in both of the responses.

Apply

Thus far, we have considered the different types of knowledge and cognitive processes that responding to this question required, which are different manifestations of the application of knowledge if that knowledge had been discussed in the context of question. However, to proceed to the higher cognitive processes involved with application, the students would have needed to interpret the process of the experiment in terms of

Mendelian principles and terminology of genes. Furthermore, the process of applying knowledge to this question required scaffolding on a mathematical model of joint probability distribution to predict outcomes in the new context. In this scaffolding, students needed to be able to construct the mathematical model as it applied to genes to find the combinations of genotypes with different types of alleles. However, conditions could be inferred from the question, the students could select among the possible outcomes and eliminate the rest.

Evidence of conceptual structure:

The first student connected the theoretical concepts of allele, genotype, and the probabilities of various phenotypes occurring to the color of the plant, which is descriptive concept. The second student referred to the theoretical concept of dominant alleles and connected that to the color of flower, which is descriptive. Using Punnet square, that student presented the theoretical concept of the probabilities for the different phenotypes; however, the student failed to connect that concept to the others.

4.3 Summary

Using the samples of the open-ended questions discussed in this chapter, I attempted to highlight the steps that we used in developing the content questions. Although, the questions were designed for different disciplines, in each instance we applied the same principles in structuring the questions. In other words, by using same structures even in different disciplines, we were able to design questions that would elicit responses with certain characteristics regarding the students' capabilities for constructing new knowledge, their thinking skills, and their abilities to construct explanations and defend scientific arguments. I identified examples of questions from physics, biology, geology, and astronomy to propose an assessment protocol that uses the same scoring scale across different disciplines. We believed the question design was more effective if it did not lead to a single correct answer but allowed different acceptable answers Therefore, we used an alternative type of assessment that instead of valuing correctness valued other characteristics, such as whether arguments are justified with plausible explanations or with plausible explanations well-defined. For each question, I selected three sample responses that illustrated the distinctive characteristics of typical categories of thinking that existed for the question.

Chapter 5 - Data Analysis

In this study, we adopted both qualitative and quantitative approaches to answer our research questions. Our main research objective is to explore the relation between the quality of students' reasoning as displayed on written content examination questions and the degree to which course is considered to be reformed. In attempt to answer our research question, we followed two steps as described below:

I. We classified students' reasoning based on their responses to written content questions.

II. We found the relation between classified responses and the degree to which science instruction was measured to be reformed.

In the previous chapters, we addressed step (I) and classified students' reasoning based on their responses to written content questions. We focused on qualitative methods when we discussed our methodology for analyzing student reasoning using responses to content questions and comparing the results across disciplines. First, we created a protocol to develop content questions with same level of thought processes in different disciplines. Then, we developed a rubric to classify students' reasoning based on the analytical scoring of the responses.

In this chapter, we describe the use of quantitative methodology to address step (II) by exploring the relation between the quality of students' thought processes and the degree to which a course is measured to be reformed.

5.1 Site visits and measure of reform

We took this study into the field by making site visits to each institution. The university faculty members were participants whose classes we observed. We used the Reformed Teaching Observation Protocol (RTOP) as our observational instrument to measure the degree to which a science classroom is reformed and measured the degree of reform for the science courses that were designed for pre-service elementary teachers.

Reformed Teaching Observation Protocol (RTOP) uses the characteristics of reformed teaching practices based on National Science Education Standards. According to MacIsaac and Falconer (2002), the characteristics that are considered in RTOP can be organized into five categories: Lesson Design and Implementation, Propositional Pedagogical Knowledge, Procedural Knowledge, Communicative Interactions and Student/Teacher Relationships. Each category includes five questions according to which the observer ranks the occurrence of the event on a scale of 0-4. Summing the 25 item scores results in an RTOP lesson score ranging from 0–100. MacIsaac and Falconer (2002) summarized the categories of RTOP as follows:

"**Lesson Design and Implementation**

The creation of science lessons that:

1) Respect student preconceptions and knowledge

2) Foster learning communities

3) Explore before formal presentation

4) Seek and recognize alternative approaches

5) Include student ideas in classroom direction

Content (Propositional Knowledge)

Teachers knowing their science and teaching lessons that:

6) Involve fundamental concepts of science

7) Promote coherent understanding across topics and situations

8) Demonstrate teacher content knowledge (e.g. apparently “unrelated” questions)

9) Encourage appropriate abstraction

10) Explore and value interdisciplinary contexts and real-world phenomena

Content (Procedural Knowledge)

Science lessons that use scientific reasoning and teachers’ understanding of pedagogy to:

11) Use a variety of representations to characterize phenomena

12) Make and test predictions, hypotheses, estimates, or conjectures

13) Include critical assessment that are actively engaging and thought provoking

14) Demonstrate meta-cognition (critical self-reflection)

15) Show intellectual dialogue, challenge, debate negotiation interpretation, and discourse

Classroom Culture (Communicative Interactions)

The use of student discourse to modify the focus of lesson control such that:

16) Students communicate their own ideas in a variety of methods

17) Teachers’ questions foster divergent modes of thinking

18) Lots of student, particularly inter student, talk is present

19) Student questions and comments shape discourse — the “teachable moment” is pursued

20) Climate of respect and expectation for student contributions

Classroom Culture (Student-Teacher Relationships)

Lesson interactions where:

21) Students actively participate (minds-on, hands-on) and set agendas

22) Students take primary and active responsibility for their own learning

23) The teacher is patient (plays out student initiatives and is silent when appropriate)

24) The teacher acts as a resource and students supply initiative

25) The teacher is a listener".

5.1.1 *Analyzing the scores*

We collected data from 18 courses located at 13 universities to measure the level of reform of science courses. We compared classes based on RTOP scores. In the 18 courses that we observed we obtained a range of RTOP overall scores that varied from 36.6 to 90. To each course, we assigned five scores as discussed above. Each category included five questions and the score assigned to each category was based on the average score of those five questions. Three raters who simultaneously observed the course ranked the occurrence of the event on a scale of 0-4. They rated each question individually and the score assigned to each question was averaged among the raters' scores. Summing the 25 item scores, results in an RTOP lesson overall score ranging from 0–100.

5.2 Statistical model

To complete step II, we need to find the relation between the ranks of classified responses to the degree to which science instruction is reformed. Any data analysis concerned with describing the relationships between dependent and independent variables involves regression model. Oftentimes the dependent outcome is a discrete variable (Hosmer & Lemeshow, 2000). For example, the risk of heart attack can be predicted based on person's age, blood cholesterol or diet. According to Hosmer & Lemeshow (2000), when the dependent variable is binary or dichotomous, the most common way of modeling is logistic regression. Logistic regression follows the general principles of ordinary regression that is the basic of using model-building techniques. In both cases, we find the best-fit and reasonable model that describes the trend of relationship. However, the major difference between linear regression and logistic regression accounts for the type of outcome variable. The outcome in logistic regression model is the probability of an event occurring. Hence, by increase and decrease in the predictor, the probability of an event occurring increases/decreases. However, in ordinary regression, the outcome has linear relationship with the predictors and the outcome variable of (Y) can fit the model of straight line:

$Y= b_0+ b_1X_1+\epsilon$ Equation 5-1

in which b_0 is the constant that defines Y intercept, b_1 is the slope of the line and ϵ is the residual term.

Multiple linear regressions are used to model the linear relationship between a dependent variable and one or more independent variables. The dependent variable is sometimes also called the predicted outcome, and the independent variables, the predictors. In simple multiple regressions, several predictors exist and a similar equation can be derived in which each predictor is multiplied by its relevant regression coefficient (Field, 2009).

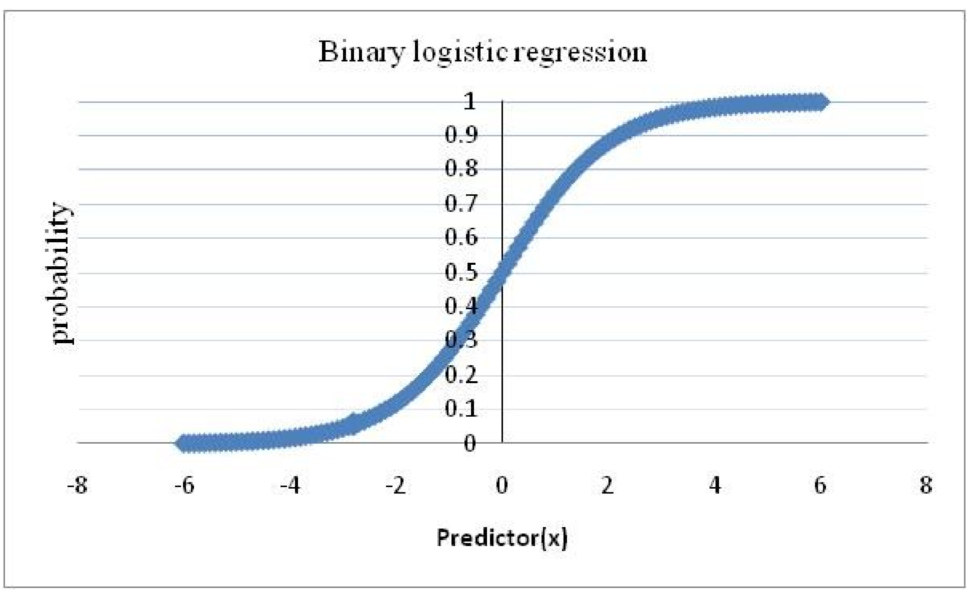
$Z= b_0+ b_1X_1+ b_2X_{2+} b_3X_{3+}\ldots\ldots\ldots b_nX_{n+}\epsilon$ Equation 5-2

5.2.1 *Logistic regression*

Logistic regression is a type of regression in which outcome variable is binary or dichotomous and the predictor variable is continuous or categorical (Field, 2009). The simplest type of logistic regression is binary logistic regression when the outcome variable can be classified in two categories such as success and failure, or evidence and no-evidence. However, the principles applied in binary logistic regression can be easily extended to the cases that more than two categories of outcome variable exist (Field, 2009). In logistic regression, instead of estimating variable Y in terms of x predictors, we predict the probability of Y in terms of the x predicators. Using the logistic equation the probability of Y can be predicated as below:

$$P(Y) = \frac{1}{1+e^{-(b_0+b_1x_1)}} \qquad \text{Equation 5-3}$$

Figure 5-1 Binary logistic regression model-graphical representation



To make use of more than one predictor the equation 5-3 can be modified as below:

$$P(Y) = \frac{1}{1+e^{-(b_0+b_1 x_1+ b_2x_2 \ldots\ldots\ldots b_nx_n)}} \qquad \text{Equation 5-4}$$

Equation 5-4 is the same as 5-3, but instead of one predicator of x, it has been extended to include many predictors. In both cases, like the normal regression model, we evaluate whether our collected data fit the statistical model. We run the analysis to estimate the coefficients of b_n in a way that probability of occurrences of the observed data fit the logistic regression probability equation with the best likelihood. Statistical packages are designed to use several methods to evaluate the likelihood. To evaluate the significance of the b_n coefficients, statistical packages are designed to perform Wald statistics on our data. Wald statistic evaluates whether the b_n coefficients are significantly different from zero, which means that the corresponding independent variable has significant contribution to the probability of outcomes (Field, 2009).

5.2.2 *Assumptions of logistic regression*

In performing logistic regression, we assume that data are case specific and each independent variable has a single value for each sample under examination. For example, the cases under study should not be related; in essence, one should not measure the same samples at different times. In addition, the assumption of multi-collinearity is violated when the independent variables are highly correlated. It is difficult to differentiate between the impacts of several variables if they are highly correlated. To resolve multicollinearity violation statisticians reduce the number of collinear variables until there is only one remaining out of the set. Alternatively, it may be possible to combine two or more highly correlated variables into a single new independent variable.

5.3 Why Logistic regression is appropriate for our study?

In this study, we are looking for the relation between the quality of students' thought processes and the degree to which course is considered to be reformed in terms of RTOP measure. To explore the relationship between the quality of the students' reasoning and the measure of reform, we collected data from 904 students at the 18 universities. For every student's response to the content questions, we assigned binary codes to seven traits of our rubric. The binary codes indicated whether there was evidence that the particular trait was observed as described in our rubric. The independent variables are the RTOP scores that can accept any value between 0 and 4, and the outcome variables are the rubric traits that can accept just 0 and 1. Totally, we should have seven logistic equations between each trait and RTOP scores.

5.3.1 *Odds Ratio-An estimation for logistic regression*

As a first step in our analysis, we estimated the strength of relationships between evidence of cognitive processes appearing in the students' responses and the level of inquiry as measured by RTOP. To accomplish this estimation we used a simplified logistic regression analysis based on two dichotomous variables. A common way to create a dichotomous variable is to divide a continuous one into two groups – high and low. To obtain a dichotomous variable for the RTOP score, we used the average RTOP score for all of the classes observed as the dividing point that average (65.5) was considered the boundary between high and low RTOP scores. Fortunately, we had no classes with a score of 65.5. First we calculated the odds that students showed evidence for each trait of rubric if they were in a class with a higher than average RTOP. (i.e.

RTOP > 65.5). Then we calculated the odds for students in a class with lower than average RTOP scores. The odds are given as follows:

$$\textbf{Odds} = \frac{\textbf{Number of (evidence)}}{\textbf{Number of (no evidence)}} \qquad \text{Equation 5-5}$$

Then we calculate the odds ratio as

$$\textbf{Odds ratio} = \frac{\textbf{High RTOP odds}}{\textbf{Low RTOP odds}} \qquad \text{Equation 5-6}$$

Table 5-1 shows the number of students who showed evidence or no evidence for cognitive process of "Apply" for the two groups of RTOP < 65.5 and RTOP > 65.5.

$$\text{Odds ratio for Apply} = \frac{254/195}{191/190} = 1.3$$

Table 5-1 Odds ratio for the "Apply" cognitive process.

	Below Average RTOP	Above Average RTOP	Total
Evidence of process in written answer	191	254	445
No-Evidence of process	190	195	385
Total	449	381	730

The odds ratio of 1.3, implies that a student in a higher than average RTOP class is 1.3 times more likely to show evidence of using "Apply" than one in a low RTOP class. However, for the cognitive process of "Explain" the odds ratio was 1, indicating that there was an equal chance of showing cognitive process of "Explain" for both low and high RTOP courses.

Table 5-2 gives the odds ratios for each of the cognitive processes we investigated.

Table 5-2 Odds ratios for cognitive processes

Cognitive process	Odds ratio
Understand /Compare	1.84
Understand /Infer	1.42
Understand/Explain	1
Apply	1.3

The simplified version of logistic regression indicated that evidence for cognitive processes depend on RTOP scores in the favor of higher RTOP scores. Because RTOP scores for a class are a measure of the level of interactive engagement, this preliminary result indicates that students in interactive engagement classes are more likely to show evidence of cognitive processes on their written exam questions. While the simplified model and preliminary results of Table 5-2 provided insight into the relationship between RTOP scores and the evidence displayed on content exams, the functional form of this relationship was still needed to be described. The results of using simplified model of logistic regression were promising for the use of the full version of binary logistic regression. In the following section, we describe the use of the full version of this statistical model.

5.4 Logistic Regression Sample output- Interpretation

In this section, I use a sample output to describe how we interpreted the steps of logistic regression analysis. In this example analysis, we commanded SAS to run a backward elimination process that includes several steps. In our sample analysis, we considered the effect of intervention1, intervention2, intervention3, intervention4 and intervention5 on some kind of binary variable. Categorical variables that have only two mutual exclusive categories (e.g. being dead or alive, pass or fail) are called binary variables. The binary logistic model evaluates if the interventions are quantitative predictors of the categories of outcome variables to occur.

To find the best fit, SAS examined different models. At the first step, SAS verified the simplest model including all the five interventions were put together with one coefficient in front of them. Therefore, the model was summarized as an intercept and the effect of other variables considered as a coefficient beta. The SAS package used chi-squared statistics to compare two existing models. In the first model, the coefficient beta was assumed different from zero and the second model consisted just intercept.

In the syntax below, SAS compared the probability of observing the Chi-Square statistic for the two suggested models with respect to the frequencies of the observed value. Under the null hypothesis, all of the regression coefficients in the model should be equal to zero. Often the null hypothesis is rejected when the p-value (pr>ChiSq) is less than the significance level α which is often 0.05 or 0.01. Therefore, the small p-value (pr>ChiSq) would lead us to conclude that the regression coefficients in the model are not equal to zero. When the null hypothesis is rejected, the coefficient beta is said to be statistically significant.

In the distribution of the Chi-Square test statistics, "DF" is the degrees of freedom and is determined by the number of predictors in the model.

We run Wald chi-square to evaluate the null hypothesis that assumed the coefficient beta in regression equations was equal to zero. The null hypothesis was rejected because we obtained a p-value that was smaller than the alpha value of .05 (or .01). Hence, the coefficient Beta is significant and should not be set to zero. This means that at least the effect of one of the interventions should be accounted in the model. Below is the logistic output for the first step.

```
Intercept Intervention1 Intervention2 Intervention3 Intervention4 Intervention5


                          The LOGISTIC Procedure

                 Testing Global Null Hypothesis: BETA=0

        Test                 Chi-Square       DF     Pr > ChiSq

        Likelihood Ratio        78.7879        5         <.0001
        Score                   86.7932        5         <.0001
        Wald                    68.1766        5         <.0001
```

5.4.1 *Backward Elimination Procedure–Step2*

To identify which intervention had significant effect we performed backward elimination procedure and removed the interventions one at a time and evaluated if the removal had significant effect on the likelihood model. Therefore, in the 2nd step, SAS removed intervention1 from the model and summarized as an intercept and the effect of other variables considered as a coefficient beta. The Wald chi-square evaluates the null hypothesis whether the constant equals to zero. The null hypothesis was rejected because the p-value (pr>ChiSq) is smaller than the alpha value of .05 (or .01). Then, the corresponding effect to Beta is significant. Hence, the constant that includes the effect of intervention2, intervention3, intervention4 and intervention5 is not zero. On the other hand, SAS used Residual Chi-Square Test to determine whether intervention1 was a significant contributor to the model. The Residual Chi-Square Test measured the difference between the observed value and the predicted value and the results were not significant for intervetion1, implied that removing intervention1, did not have significant effect.

```
Effect removed Intervention1    Testing Global Null Hypothesis: BETA=0

              Test                 Chi-Square       DF     Pr > ChiSq

              Likelihood Ratio        77.2576        4         <.0001
              Score                   78.2806        4         <.0001
              Wald                    64.1114        4         <.0001

                              Residual Chi-Square Test

                         Chi-Square       DF     Pr > ChiSq

                             1.5917        1         0.2071
```

5.4.2 *Backward Elimination Procedure–Step3*

In the step 3, the effects of intervention 1 and 2 are removed and the effects of intervention3, intervention4 and intervention5 considered as a coefficient beta. The Wald chi-square evaluated the null hypothesis whether the constant equals zero. This hypothesis is rejected because the p-value is smaller than the alpha value of .05 (or .01). Hence, the constant is not zero and other variables should be included. For measuring the difference between the observed value and the predicted value, SAS used the Residual Chi-Square Test. The residual was not significant; consequently removing the effect of intervention2 would not be significant.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

print of data 09:43 Tuesday, June 28,

The LOGISTIC Procedure

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	529.475	461.864
SC	534.177	480.672
-2 Log L	527.475	453.864

Effect removed intervention2 Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	73.6106	3	<.0001
Score	73.0023	3	<.0001
Wald	64.5052	3	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
5.3682	2	0.0683

NOTE: No (additional) effects met the 0.05 significance level for removal from the model.

5.4.3 *Summary of backward elimination*

In the same way, SAS evaluated the significances of the intervention3, intervention4 and intervention5. After performing the analysis of likelihood estimates, only the contribution of two effects of interventions 1 and 2, were removed from the model as the effects met the 0.05 significance level of removal from the model. Therefore, the variables that were included in the model were intervention3, intervention4 and intervention5.

Step	Effect Removed	DF	Number In	Wald Chi-Square	Pr > ChiSq
1	Predictor1	1	4	1.5790	0.2089
2	Predictor2	1	3	3.5839	0.0583

5.4.4 *Estimates for regression coefficients*

This section is the major part of logistic regression analysis as it informs about the coefficients of the predictors included in model. In the sample analysis, we predict the probability of the occurrence of the outcome variable in terms of the predictors. Within the framework of inferential statistics, we found the regression coefficients (b_n) to fit the probability equation (5-4) discussed in section 5-2-1.

$$\boldsymbol{P(Y)} = \frac{\mathbf{1}}{\mathbf{1+e^{-(b_0+b_1\,x_1+\,b_2x_2\,\ldots\ldots\ldots\,b_nx_n)}}}$$

We considered *X, Y* to be continuous, dichotomous respectively. According to Equation (5-4), the relationship between the probability of *Y* and *X* is nonlinear and the value of the coefficient b_n determines the direction of the relationship between *X* and the *Y.* When b_n is greater than zero, larger (or smaller) *X* values are associated with smaller (or larger) values of *Y.* Conversely, if b_n is less than zero, larger (or smaller) *X* values are associated with larger (or smaller) values of *Y.* In the case where the regression coefficient is zero, there is no linear relationship in the population. By inserting the coefficient that we obtained from the analysis of maximum likelihood, into equation (5-4), we can express the probability of outcome variable as below:

$$\mathbf{P_{outcome-variable}} = \frac{\mathbf{1}}{\mathbf{1+e^{-(3.48-0.21\,x_1+\,4.47x_4-4.04x_3)}}} \qquad \text{Equation 5-7}$$

Summary of Backward Elimination

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	3.4752	0.3914	78.8457	<.0001
Intervention3	1	-0.2096	0.0355	34.8234	<.0001
Intervention4	1	4.4740	0.5940	56.7385	<.0001
Intervention5	1	-4.0421	0.5766	49.1408	<.0001

5.4.5 *Odds Ratio Estimates*

Finally, I need to connect the odds ratio to the logistic regression. In the section 5.3.1, we calculated a rough estimate of odds ratio to learn more about logistic regression analysis and obtain more insight about our data. To simplify the analysis, we sought the relationship between outcome variable and overall RTOP score. The average RTOP score was considered the boundary between high and low RTOP. In this section, I explain how we can calculate odds ratios from the output to interpret the data. Odds ratios can be found from the estimate point Exp (Z), in which:

$$Z= b_0+ b_1X_1+ b_2X_{2+}\ b_3X_{3+\ldots\ldots\ldots}\ b_nX_{n+}\epsilon \qquad \text{Equation 5-8}$$

Equation (5-4) describes the probability of occurring an event as;

$$\boldsymbol{P(Y)} = \frac{\mathbf{1}}{\mathbf{1+e^{-(b_0+b_1\,x_1+\,b_2x_2\ldots\ldots\ldots\,b_nx_n)}}}$$

If the probability of an event happening is P(Y), then the probability of the event not happening would be [1-P(Y)] and we can calculate the odds as:

$$\textbf{Odds} = \frac{\boldsymbol{P(Y)}}{\mathbf{1}-\boldsymbol{P(Y)}} \qquad \text{Equation 5-9}$$

Therefore, if we know the odds before and after a unit change in the predictor variable, we can calculate the proportionate change in the odds as:

$$\textbf{Odds ratio} = \frac{\boldsymbol{Odds\ after\ one\ unit\ change\ in\ the\ predictor}}{\boldsymbol{Original\ odds}} \qquad \text{Equation 5-10}$$

The odds can be expressed as Exp (Z), and the ratio of the odds is called the point estimate in SAS outputs for logistic regression. If the odds ratio is greater than one, then as predictor increases the odds of observing outcome variable is more likely. Conversely, for the ratio's less than one as the predictor increases the odds of happening outcome variable become less likely (Field, 2009). The excerpt below which is taken from our SAS output sample analysis shows the point estimates for the confidence interval of 95%. Based on the given confidence interval we are 95% confident that the true value of odd ratio is between the two limits of the interval.

The LOGISTIC Procedure

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
Intervention2	0.811	0.756	0.869
Intervention3	87.703	27.380	280.922
Intervention4	0.018	0.006	0.054

5.5 The Output Report

In this section, I report the output results our statistical analysis. The data were collected from 18 science courses at 13 universities. Below, I present a snap shot of our data. The column "number" represents the number of students in the class, and numbers entered under the traits, represents number of students who showed evidence, in their written response, for the occurrence of the trait. For example, class # 810 has 56 students, of whom 41 students showed evidence of having factual knowledge, 37 conceptual, 12 procedural, 24 compare, 30 infer, 30 explain, and 25 apply.

Table 5-3 The snapshot of excel spreadsheet for our collected data

CLASS#	number	FACTUAL	CONCEPTUAL	PROCEDURAL	COMPARE	INFER	EXPLAIN	APPLY
810	56	41.0	37.0	12.0	24.0	30.0	30.0	25.0
1110	50	37.0	29.0	11.0	18.0	12.0	13.0	13.0
1510	38	38.0	38.0	9.0	18.0	38.0	25.0	27.0
1610	22	10.0	6.0	3.0	2.0	1.0	1.0	1.0
620	138	134.0	132.0	77.0	108.0	124.0	104.0	74.0
1020	42	41.0	40.0	21.0	30.0	25.0	22.0	17.0
1010	21	21.0	21.0	15.0	14.0	16.0	15.0	10.0
610	48	47.0	46.0	.	46.0	45.0	.	.
520	85	68.0	62.0	.	30.0	52.0	48.0	59.0
510	19	18.0	17.0	.	9.0	11.0	11.0	17.0
910	96	93.0	93.0	.	82.0	80.0	81.0	78.0
920	115	106.0	94.0	.	75.0	72.0	72.0	68.0
1320	16	15.0	13.0	.	7.0	8.0	8.0	6.0
1810	19	19.0	19.0	.	11.0	19.0	19.0	19.0
310	11	10.0	9.0	.	8.0	9.0	8.0	8.0
1910	19	17.0	16.0	.	16.0	16.0	16.0	14.0
1909	19	18.0	12.0	7.0	6.0	6.0	4.0	4.0
100	14	9.0	9.0	.	2.0	2.0	3.0	3.0

The spreadsheet shows total students performance per class, so the number for each trait shows number of the students showed evidence for the particular type of trait.

Table 5-4 Excel spreadsheet showing the RTOP scores of the classes

CLASS#	number	LESSON-DESIGN	PROPOSITIONAL-KNOWLEDGE	PROCEDURAL-KNOWLEDGE	COMMUNICATIVE-INTERVENTION	STUDENT-TEACHER-RELATIONSHIPS	OVERALL-SCORE
810	56	0.75	2.84	0.8	1.05	1.85	1.46
1110	50	1.4	3.26	13	1.93	2.07	2.04
1510	38	3.2	3.4	1.8	3.5	3.6	3.1
1610	22	2.64	3.32	2.84	2.88	3.64	3.06
620	138	0.47	2	0.33	0.53	0.67	0.8
1020	42	3.1	3.25	2.7	2.65	3.05	2.95
1010	21	2.29	3.02	2.18	2.39	2.63	2.37
610	48	3.27	3.6	3.67	3.8	3.93	3.6
520	85	1	3.6	0.8	0.6	1	1.2
510	19	2.8	3.4	3.4	3.6	3.4	3.32
910	96	3.6	3.2	3.6	3.6	4	3.6
920	115	2.6	2.6	2.4	3	3.4	2.8
1320	16	0.4	2.6	0.8	1.2	1.4	1.28
1810	19	2.4	3.4	2.4	3	3	2.84
310	11	3	3.4	3	3.4	3.8	3.32
1910	19	3.2	3.8	3	3.2	3.8	3.4
1909	19	3.2	3.2	3.2	3.2	3.2	3.2
100	14	2.38	2.38	2.33	2.73	3.2	2.59

We used SAS version 9.2 to run the logistic regression analysis. We reported seven outputs for the outcome variables of factual knowledge, conceptual schema, procedural knowledge, compare and contrast, infer, explain and apply. The outcome variables have two values, including whether or not the evidence existed that those traits occurred.

For this analysis, we commanded SAS to perform backward stepwise method. SAS allows you to have different steps in your logistic regression model. The difference between the steps is the predictors that are included in the process of iteration. The question that to what extent the model matches the outcome variable can be assessed using chi squared statistics that examines the difference between model with predictors included and model only consisting one constant (Field, 2009).

At the start, all the predictors were included and in several iteration steps, one or more predictors were eliminated. Then the analysis was programmed to evaluate if the elimination was significant to the model or not.

5.5.1 *Logistic regression: checking the assumptions*

In the previous sections, we focused on logistic regression analysis and interpreting the results of our logistic model. In order for our analysis to be valid, our model has to satisfy the assumptions of logistic regression. If we violate the assumptions of logistic regression, then our statistical inferences would not be reliable. One of the assumptions of logistic regression is the absence of multi-collinearity, which is a state of high correlations among the independent variables. To check the multi-collinearity, we performed Pearson correlation analysis. We found the correlation coefficients for the components of RTOP, which the results are shown in Table (5-5). The highly correlated variables are lesson design, procedural knowledge, communicative-interactions and student/teacher relationships. One approach to solve the problem would be to combine the parameters that are affected by multi-collinearity. As such, we combined lesson design, procedural knowledge, communicative-interactions and student/teacher relationships and defined a new variable called combination score obtained by calculating the average of the four highly correlated variables. After combining the variables our independent variables reduced to propositional knowledge and combination score.

Table 5-5 Pearson correlation coefficients for the components of RTOP scores

Pearson Correlations Coefficients	Lesson Design	Propositional Knowledge	Procedural Knowledge	Class Interactions	Student/Teacher relationships
Lesson Design	1.00	0.55	0.92	0.94	0.94
Propositional Knowledge	0.55	1.00	0.53	0.50	0.50
Procedural Knowledge	0.92	0.53	1.00	0.92	0.91
Class Interactions	0.94	0.49	0.92	1.00	0.97
Student Teacher relationships	0.94	0.49	0.91	0.97	1.00

In performing logistic regression, we also assume that data are case specific and each independent variable has a single value for each sample under examination. Therefore, to meet the assumption of independence of errors, we changed the unit of analysis from student to class by aggregating the data.

5.6 Logistic model with several predictors

We ran our analysis using the SAS package version 9.2, and followed the same procedure that was described in the Section 5.4 to interpret the SAS output results. In this section, I report the nature of the relationships for each trait of rubric derived from the SAS output.

These estimated models describe the relationships between the independent variables and the outcome variables. Using the coefficients we can also infer if the effect is positive or negative, moreover we can compare the amount of increase or decrease that can be caused by independent variable. If the sign of the coefficient is negative, the predictor has negative effect and vice versa. If the effect of independent variable is not significant, the coefficients are not significantly different from 0, which should be taken into account when interpreting the coefficients.

The syntax below corresponds to the logistic model of factual knowledge. By inserting the intercept and regression coefficients of propositional knowledge and combination score in the equation 5-4, we derived equations 5-11 and 5-12. These equations depict the likelihood of students' performance on factual knowledge in relation to propositional knowledge score and combination score. These two scores represent the degree to which the RTOP measures the class to be reformed.

Z=3.80 – 0.82(Propositional Score) + 0.39(Combination Score)   Equation 5-11

$$P_{Factual} = \frac{1}{1+e^{-(3.80-0.82\text{Prop}+0.39\text{Combination})}}$$ Equation 5-12

Analysis for Factual

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	3.8023	0.6570	33.4930	<.0001
Prop_Knowledge	1	-0.8159	0.2257	13.0672	0.0003
Combinationscore	1	0.3906	0.1056	13.6873	0.0002

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
Prop_Knowledge	0.442	0.284	0.688
Combinationscore	1.478	1.202	1.818

According to equation 5-11 and equation 5-12 as the propositional knowledge score increases, the likelihood of the evidence of factual knowledge decreases. In contrast, the combination score positively affects the likelihood of factual knowledge occurring in the students' responses.

5.6.1 *Finding regression coefficients for each trait*

We followed the same approach, and selected the syntax corresponding to the logistic model of each trait. In the Tables 5-6 and 5-7, I report regression coefficients and P-values of propositional knowledge and combination score of each trait that were derived from the SAS output. In addition, I report odds ratios and confidence intervals of odds ratios of propositional knowledge and combination score for each trait.

By inserting the intercept and regression coefficients of propositional knowledge and combination score of each trait in the equation 5-4, we derived equations that depicted the likelihood of occurrences the corresponding traits. Totally, we derived seven equations that are summarized in Tables 5-8 and 5-9 for knowledge type and cognitive processes respectively. These models describe the relationship between the probability of occurring a certain trait and RTOP measures.

Table 5-6 Analysis of Maximum Likelihood Estimates for the knowledge dimension

Parameter	Factual Knowledge	Conceptual schema	Procedural Knowledge
Intercept coefficient	3.80	3.16	4.15
Intercept P-value	P <0.0001	P <0.0001	<0.0001
Propositional knowledge coefficient	-0.82	-0.73	-2.14
Propositional knowledge P-value	P=0.0003	P= 0.0002	<0.0001
Odds ratio	0.44	0.48	0.12
Confidence intervals	0.28-0.69	0.33-0.71	0.05--0.29
Combination score coefficient	0.39	0.31	0.80
Combination score P-value	P= 0.0002	P= 0.0005	P=0.0013
Odds ratio	1.48	1.36	2.22
Confidence intervals	1.20-1.81	1.14-1.62	1.37-3.62

Table 5-7 Analysis of Maximum Likelihood Estimates for the cognitive dimension

Parameter	Compare	Infer	Explain	Apply
Intercept coefficient	2.08	1.81	1.47	-0.66
Intercept P-value	P <0.0001	P <0.0001	P=0.0004	P=0.0968
Propositional Knowledge coefficient	-0.84	-0.45	-0.44	0.23
Propositional knowledge P-value	P <0.0001	P= 0.0043	P= 0.0041	P=0.1280
Odds ratio	0.43	0.64	0.65	1.26
Confidence intervals	0.32-0.58	0.47-0.87	0.48-0.87	0.94-1.69
Combination-score coefficient	0.41	0.13	0.14	0.13
Combination score P-value	P <0.0001	P= 0.065	P=0.059	P=0.07
Odds ratio	1.50	1.14	1.14	1.14
Confidence interval	1.30-1.73	0.99-1.32	0.99-1.32	0.99-1.30

Table 5-8 Logistic regression model for the knowledge dimension

Type of trait	Logistic regression model	Describing the trend
Factual Knowledge	$P_{Factual} = \frac{1}{1 + e^{-(3.80 - 0.82Prop + 0.39Combination)}}$ Z=3.80 – 0.82(Propositional) + 0.39(Combination score)	As the propositional knowledge score increases, the likelihood of the evidence of factual knowledge decreases. In contrast, the combination score positively affects the likelihood of factual knowledge as it appears in the students' responses.
Conceptual Schema	$P_{Conceptual} = \frac{1}{1 + e^{-(3.16 - 0.73Prop + 0.31Combine)}}$ Z=3.16 – 0.73(Prop) + 0.31(Combination score)	As the propositional knowledge score increases, the likelihood of conceptual knowledge occurring in student responses decreases. In contrast, as the combination score increases the likelihood of conceptual knowledge occurring increases
Procedural Knowledge	$P_{Procedural} = \frac{1}{1 + e^{-(4.15 - 2.14Prop + 0.80Combination)}}$ Z=4.15 – 2.14(Propositional) + 0.80(Combination score)	For the procedural knowledge, the same trend as factual knowledge and conceptual knowledge holds; however, the propositional knowledge shows a stronger negative effect on procedural knowledge rather than factual and conceptual knowledge

Table 5-9 Logistic regression model for the cognitive dimension

Type of trait	Logistic regression model	Describing the trend
Compare	$P_{Compare} = \frac{1}{1+e^{-(2.08-0.84\text{Prop}+0.41\text{Combination})}}$ Z=2.08 – 0.84(Propositional) + 0.41(Combination score)	As the propositional knowledge score increases, the likelihood of the evidence of the cognitive process of *Compare* decreases. In contrast, the combination score has a positive effect on the likelihood of students cognitively showing compare.
Infer	$P_{Infer} = \frac{1}{1+e^{-(1.82-0.45\text{Prop})}}$ Z=1.82 – 0.45(Propositional)	Propositional knowledge has a negative effect on students' inference skills. The effect of the combination score is negligible as the p value for this coefficient was insignificant
Explain	$P_{Explain} = \frac{1}{1+e^{-(1.47-0.44\text{Prop})}}$ Z=1.47– 0.44(Propositional)	The propositional knowledge again has a negative effect on students' explanation skills. The effect of the combination score is negligible as the p value for this coefficient was insignificant.
Apply	$P_{Apply} = \frac{1}{1+e^{-(0.23\text{Prop})}}$ Z=- 0.23(Propositional)	In contrast to other traits, propositional knowledge again has a positive effect on students' cognitive process of *Apply*. The effect of combination score is negligible as the p value for this coefficient is insignificant

5.7 Logistic Regression Model for Overall RTOP Score

While the models with several predictors provide more information in terms of showing the characteristics of the courses and the contributions of each individual characteristics to the students' reasoning skills, still there are some benefits to describe to model in terms of one predictor, which is the overall RTOP score. This score is the average of all the RTOP sub scores. The models with several predictors are more complicated and science education community is more familiar with RTOP overall score; therefore, it would be easier to communicate the results in terms of RTOP overall score. Moreover, it provides a summarized overview of our data analysis. In this section, I present seven logistic regression models that were obtained from statistical analysis of each trait in terms of RTOP overall score.

As we can see from the syntax below, the Wald chi-square test rejected the null hypothesis because the p value (Pr>ChiSq) was smaller than the critical alpha value of .05.

```
                          Analysis for Factual

                         The LOGISTIC Procedure

                  Analysis of Maximum Likelihood Estimates

                                     Standard          Wald
Parameter       DF     Estimate        Error    Chi-Square    Pr > ChiSq

Intercept        1       1.6573       0.2613       40.2184        <.0001
Avg_Score100     1      0.00889      0.00439        4.1130        0.0426


                          Odds Ratio Estimates

                             Point          95% Wald
           Effect          Estimate      Confidence Limits

           Avg_Score100       1.009       1.000       1.018
```

By inserting the intercept and regression coefficient of overall score in the equation 5-3, we derived equations 5-13 and 5-14:

$$P_{factual} = \frac{1}{1+e^{-(1.66+0.009*avescore)}} \qquad \text{Equation 5-13}$$

$$Z=1.66+ 0.009(\text{Overall score}) \qquad \text{Equation 5-14}$$

These equations depict the likelihood of students' performance on factual knowledge in relation to RTOP overall score.

5.7.1 *Finding regression coefficients for each trait*

We followed the same approach, and selected the syntax corresponding to the logistic model of each trait. In the Tables 5-10 and 5-11, I report regression coefficients and P-values of the overall score for each. The numerical values listed in the Tables 5-10 and 5-11 were derived from the SAS output. In addition, I report odds ratios and confidence intervals of odds ratios of overall score for each trait.

By inserting the intercept and regression coefficients of overall score of each trait in the equation 5-4, we derived equations that depicted the likelihood of occurrences the corresponding traits. Totally, we derived seven equations that are summarized in Tables 5-12 and 5-13 for knowledge type and cognitive processes respectively. These models describe the relationship between the probability of a certain trait occurring and the RTOP measures. The null hypothesis was rejected in most of the cases; hence, we conclude that the regression coefficient significantly affects the response variables. However, for the traits 'Infer' and 'Explain', the null hypotheses were not rejected.

Table 5-10 Likelihood estimates -knowledge dimension in terms of overall score

Parameter	Factual Knowledge	Conceptual schema	Procedural Knowledge
Intercept coefficient	1.66	1.30	0.23
Intercept P-value	P <0.0001	P <0.0001	P=0.32
Overall score coefficient	0.009	0.006	-0.01
Overall score P-value	P=0.04	P= 0.11	P=0.002
Odd ratio	1.00	1.01	0.99
Confidence level 95%	1.00-1.02	0.99-1.01	0.98-1.00

Table 5-11 Likelihood estimates -cognitive dimension in terms of overall score

Parameter	Compare	Infer	Explain	Apply
Intercept coefficient	-0.03	0.73	0.40	-0.18
Intercept P-value	P =0.84	P <0.0001	P=0.0249	P=0.32
Overall score coefficient	0.008	0.001	0.001	0.008
P-value	P =0.0024	P= 0.81	P= 0.6837	P=0.005
Odds ratio	1.01	1.00	1.00	1.01
	1.00-1.01	1.00-1.01	1.00-1.01	1.00-1.01

Confidence interval 95%				

Table 5-12 Logistic regression model for the knowledge dimension

Type of trait	**Logistic regression model**	**Describing the trend**
Factual Knowledge	$P_{factual} = \frac{1}{1+e^{-(1.66+0.009)}}$ Z=1.66+ 0.009x	As the RTOP score increases, the likelihood of the evidence of factual knowledge in student responses increases.
Conceptual Schema	$P_{Conceptual} = \frac{1}{1+e^{-(1.30+0.006x)}}$ Z=1.30+ 0.006(overall score)	As the RTOP score increases, the likelihood of the evidence of conceptual knowledge in students' responses increases
Procedural Knowledge	$P_{procedural} = \frac{1}{1+e^{-(0.23-0.01x)}}$ Z=0.23- 0.01(x)	The regression model for procedural knowledge is slightly different from factual and conceptual knowledge as the likelihood of the evidence of procedural knowledge decreases, as the RTOP score increases.

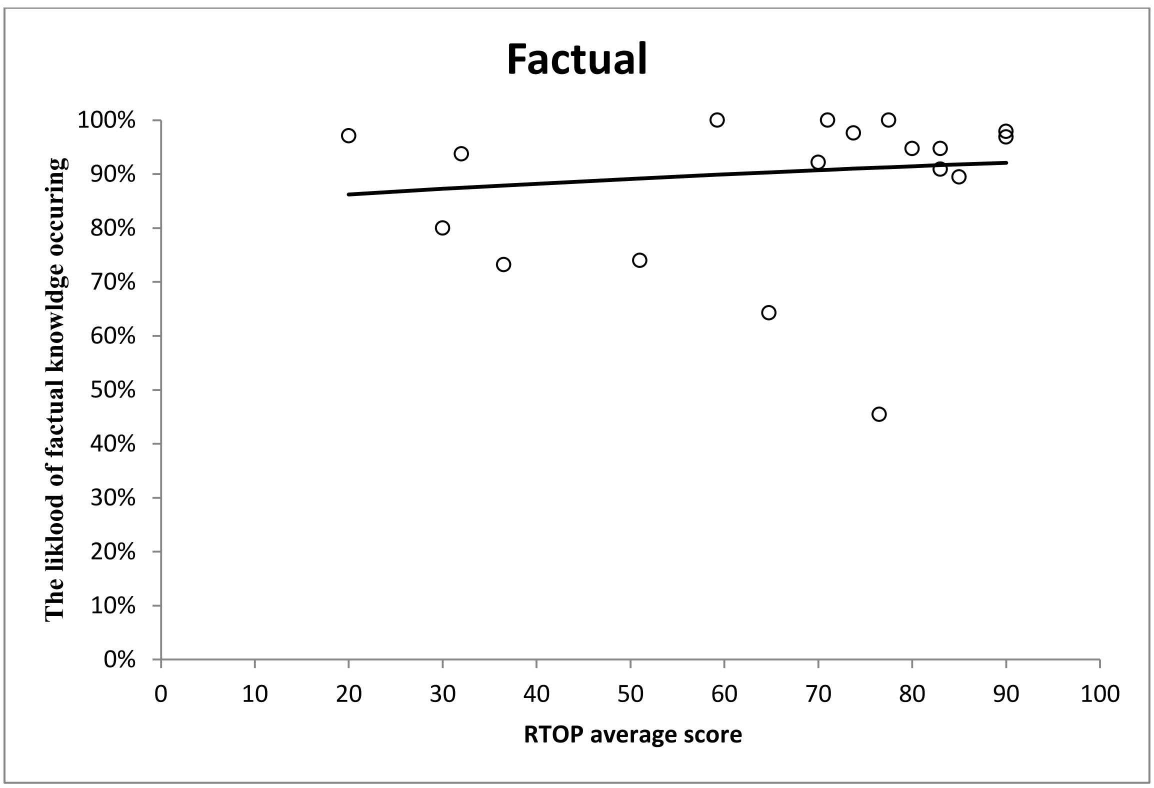
Table 5-13 Logistic regression model for the cognitive dimension

Type of trait	Logistic regression model	Describing the trend
Compare	$P_{Compare} = \frac{1}{1+e^{-(-0.035+0.01x)}}$ Z=-0.035+0.01(x)	As the RTOP score increases, the likelihood of the evidence for compare in student responses increases.
Infer	$P_{Infer} = \frac{1}{1+e^{-(0.73+0.001x)}}$ Z=-0.73+0.001(x)	There is no relationship between the RTOP average score and evidence in student responses for inference
Explain	$P_{Explain} = \frac{1}{1+e^{-(0.40-0.001x)}}$ Z=-0.40+0.001(x)	There is no relationship between evidence of students' ability to explain and the increase in RTOP average score
Apply	$P_{Apply} = \frac{1}{1+e^{-(-0.18+0..008x)}}$ Z=--0.18+0.008(x)	Likelihood of evidence in their responses of students' ability to apply slightly increases as the RTOP average score increases

5.7.2 *Graphical representation of logistic regression models*

In this section, I show the logistic fit charts that are the graphical representation of the likelihood functions of the traits in terms of RTOP overall score. Graphical representation provides more visualization about the shape and steepness of the likelihood function and accuracy of the model. Moreover, it shows how the model outputs are distributed all over the model's range.

Figure 5-2 Factual knowledge-Logistic regression model



The line in Figure 5-2 is the graphical representation of equation $P_{factual} = \frac{1}{1+e^{-(1.66+0..009x)}}$ that shows a gradual increase of outcome variable with respect to the RTOP measures. The open circles represent the data point for the 18 classes.

Figure 5-3 Conceptual knowledge-Logistic regression model

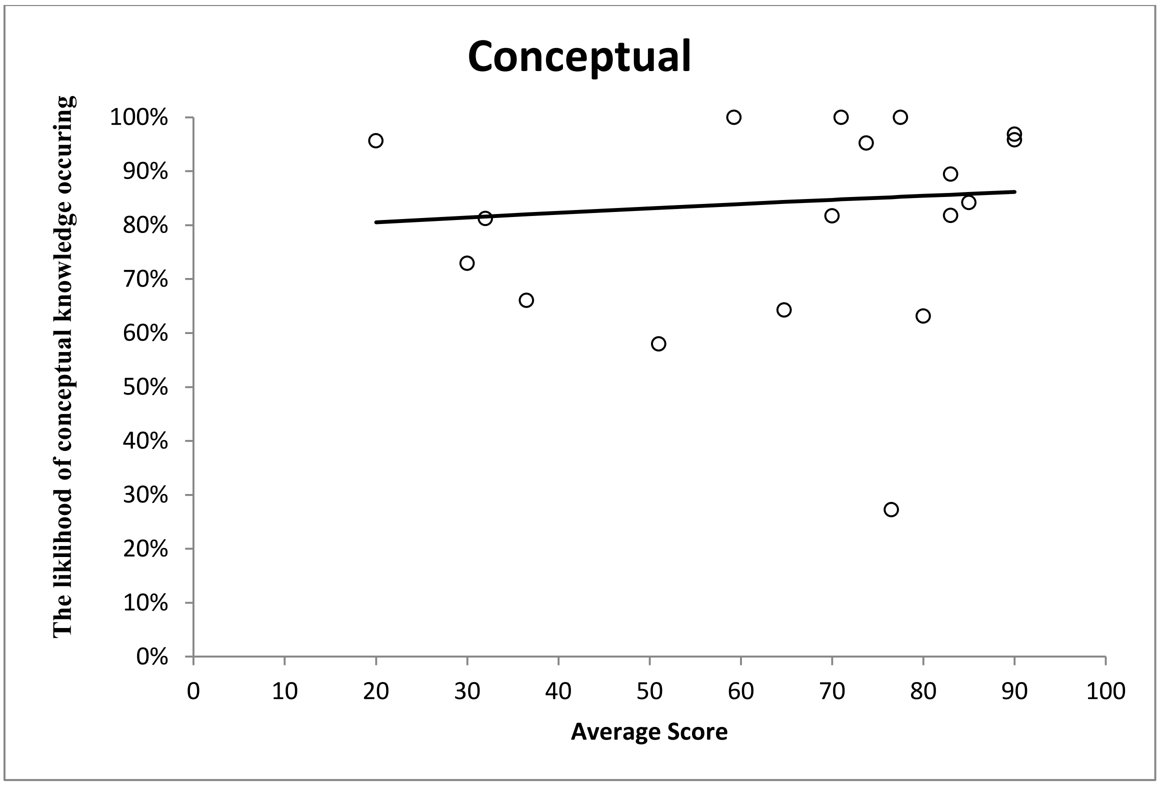


Figure 5-3 represents a graphical representation of equation $P_{Conceptual} = \frac{1}{1+e^{-(1.30+\ 0.006x)}}$ that shows a slight increase of conceptual schema with respect to the RTOP measures.

Figure 5-4 Procedural knowledge-Logistic regression model

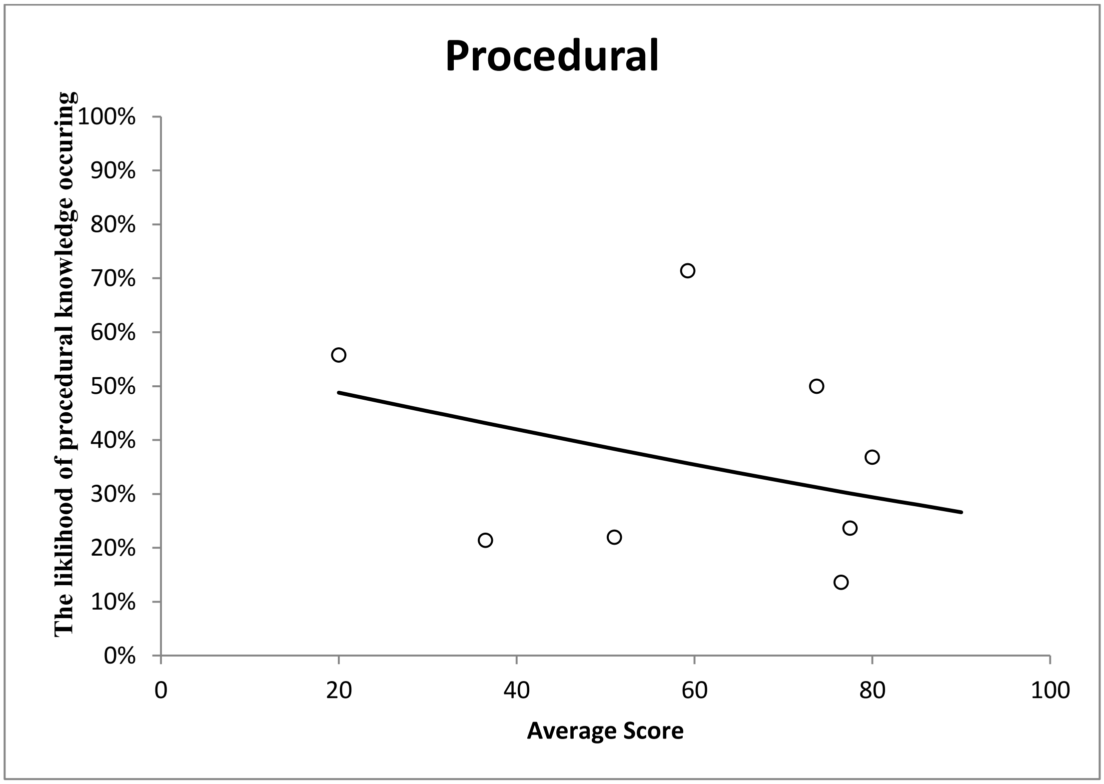


Figure 5-4 represents the graphical representation of equation $P_{procedural} = \frac{1}{1+e^{-(0.23-0.01x)}}$ that shows a steep decrease of procedural knowledge as the RTOP scores increases.

Figure 5-5 Cognitive process of compare-Logistic regression model

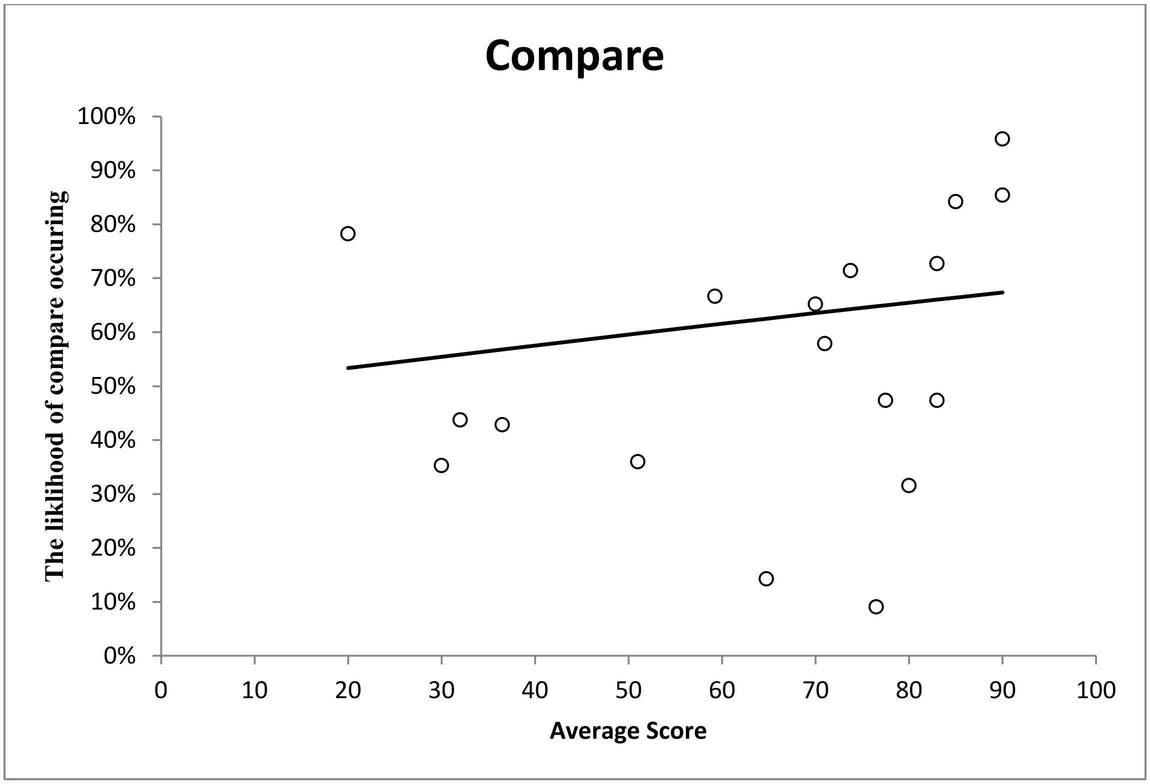


Figure 5-5 is the graphical representation of equation $P_{Compare} = \frac{1}{1+e^{-(-0..035+0.008x)}}$ that shows a gradual increase of cognitive process of compare as the RTOP scores increases.

Figure 5-6 Cognitive process of infer-Logistic regression model

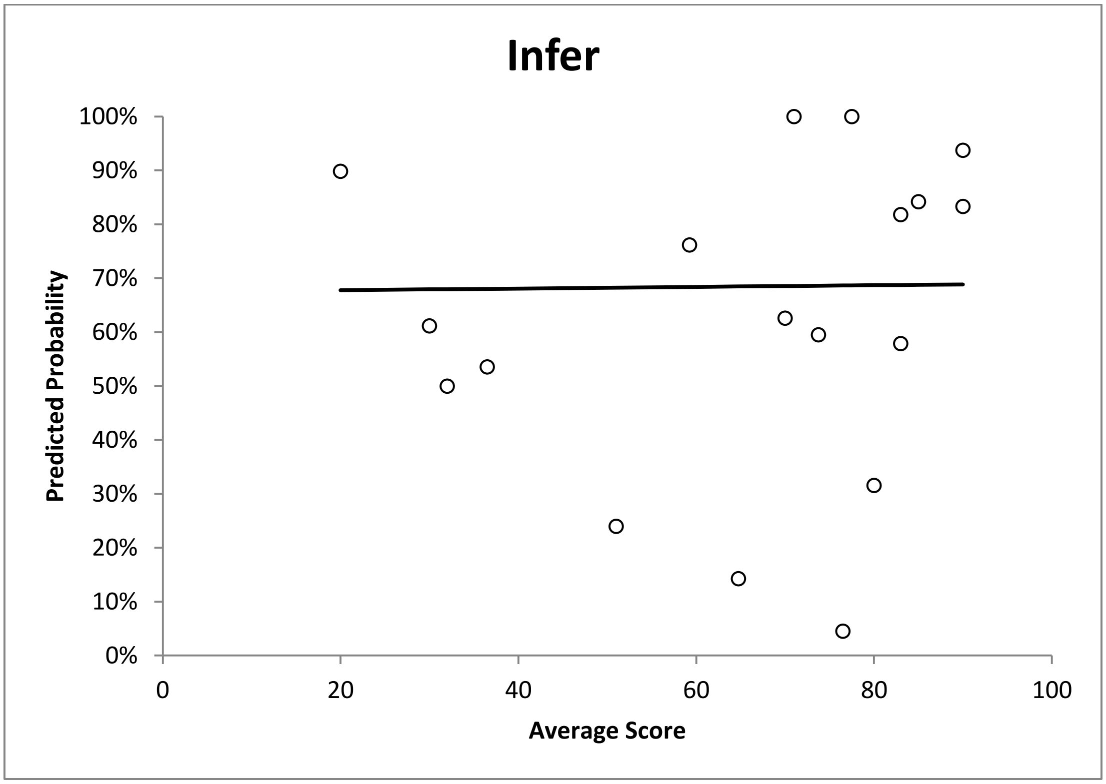


Figure 5-6 is the graphical representation of equation below:

$$P_{Infer} = \frac{1}{1 + e^{-(0.73+0.001x)}}$$

A completely flat regression line would imply that there is no relationship between the RTOP average score and students' inference.

Figure 5-7 Cognitive process of explain-Logistic regression model

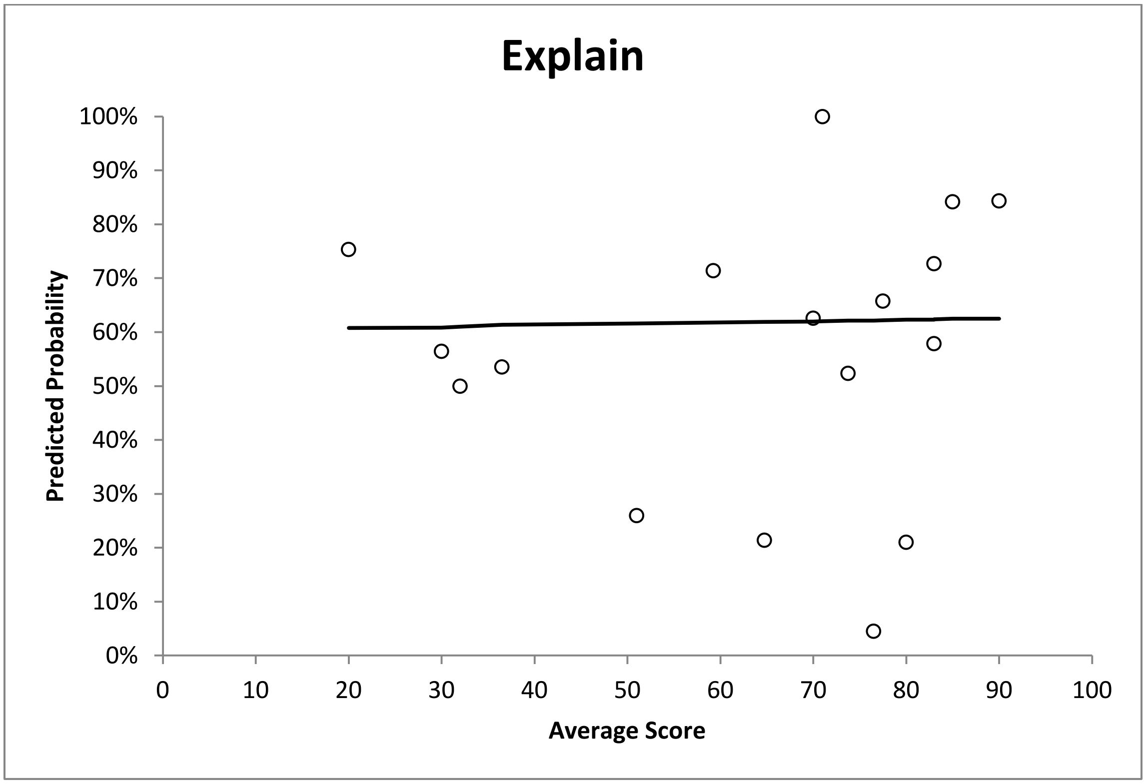


Figure 5-7 is the graphical representation of likelihood function of explain that is described by equation below:

$$P_{Explain} = \frac{1}{1 + \mathrm{e}^{-(0.40 - 0.001\mathrm{x})}}$$

A completely flat regression line would imply that there is no relationship between students' ability to explain and the increase in RTOP average score.

Figure 5-8 Cognitive process of apply-Logistic regression model

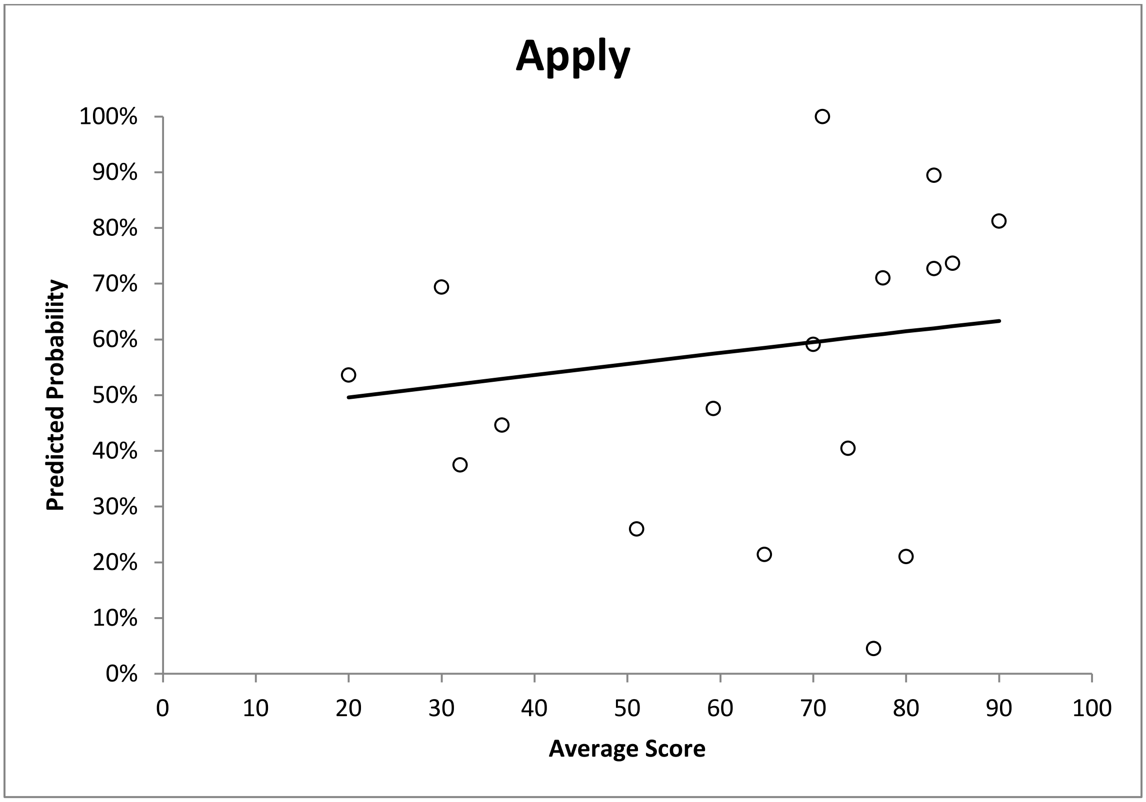


Figure 5-8 shows the graphical representation of likelihood function of apply (Table 5-13)

$$P_{Apply} = \frac{1}{1 + \mathrm{e}^{-(-0.18+0.008\mathrm{x})}}$$

The regression line shows a gradual increase, which implies that likelihood of students' ability to apply slightly increases as the RTOP average score increases.

5.8 Clustering student's conceptual structure

As I mentioned in the previous chapters, we classified students' responses in terms of conceptual structure. As a part of our protocol, we used Lawson's et al. classification scheme and categorized concepts into three types of Descriptive, Hypothetical, and Theoretical in order to explore how students relate those concepts. First, we classified concepts into three types: Descriptive, Hypothetical, and Theoretical, and categorized the level of abstraction of the responses in terms of the types of concepts and the links between them. This procedure followed an approach by Nieswandt and Bellomo (2009), for analyzing the conceptual structure that student utilized in responding to written questions. According to Nieswandt and Bellomo, at the least sophisticated level, students made connections between two same level descriptive concepts (one-concept-level). Then, they examined students' abilities to make connections between two different levels of concepts (cross-concept-level). The criteria for meaningful understanding in this type of assessment was applying all three levels of concepts and finding connections between them (Multi-level concepts).

Inspired by Nieswandt and Bellomo study (2009), and with a slight modification we defined following three categories for the conceptual structure that students exhibited:

- **High-level-links** = Types of conceptual structures that consist of at least three meaningful links among higher level concepts such T or H, including one possible connection to D, the concept structure link for this groups includes structures such as; D-H-T, T-T-T, H-H-H, T-H-T, T-H-H, T-T-T-T, etc.
- **Middle-level-links** = Types of conceptual structures that consist of one meaningful link between H,T,D, excluding the case D-D. This group includes structures such as T-T, T-H and H-H or D-H.
- **Low-level-links** = Types of conceptual structure that includes only meaningful links between descriptive concepts.

We classified students' responses in terms of the represented conceptual structure. Most of answers were not sophisticated with only descriptive concepts or some connections between theoretical and descriptive concepts. The occurrence of more sophisticated responses with multi-level links among concepts (theoretical-descriptive-hypothetical) was rare.

5.8.1 *Clustering the data*

Sometime students' responses involved lengthy explanations, and hence included a large number of concept links. To interpret the messy data and extract the useful information, we classified cases into six groups. Each group contained answers which were relatively similar in the types of concepts that the students used and in the way that they linked the concepts. The groups were based on the following definitions

Group A= Answers that included two or more than two High-level links

Group B= Answers that included one High-level-link in combination with Middle-level-links

Group C= Answers that included two or more Middle- level- links

Group D= Answers that included one Middle-level-link

Group E = Answers that included just Low- level links

Group F= Answers with discrete concepts with no links

Once we had classified the answers, we found the relation between the conceptual structure of the responses and the degree to which the courses are considered to be reformed as measured by the RTOP.

5.9 Multinomial logistic regression

When the outcome variable consists of more than two types of categories, the multi-nominal logistic regression is an appropriate model for the analysis. In the section 5.8 I explained the classification of students' responses based on the conceptual structure of the responses into the six mutually exclusive groups. Given a set of independent RTOP variables and the frequency of the answers that are located in each group, we can measure the probability of students displaying certain types of conceptual links given a certain RTOP score. Multinomial logistic regression is an appropriate statistical method for estimating the probabilities of the six possible categories of conceptual structure in terms of the RTOP overall scores. The logic behind this method is the same as binary logistic regression; therefore, we would not need an additional equation (Field, 2009).

Multinomial logistic regression includes several regression models to describe the probability distribution of the outcome variable (Allison, 2008). The regression models determine the probabilities of the occurrence of the categories of outcome variable. For the analysis of the conceptual structure, the outcome variable includes six categories and SAS derived six regression models to describe the probabilities of students displaying different types of conceptual links for a given RTOP score. To describe the model in more detail, we define probability of each category as "Pj " $=\pi_{j}$, where j is a dummy variable that runs through the indices that we chose to represent the six distinctive groups which are letters from "A" to "F". Since, there are only six outcome categories, the summation of the probabilities ($\sum_{j=A}^{j=F} \pi_j = 1$) should be equal to one.

The process of finding probabilities of the categories starts by selecting a baseline category and calculating the observed probability of that category. This process continues by breaking the outcome variable into series of comparisons between different categories

(Field, 2009). In the following sections, I describe how we can use multinomial logistic regression to estimate the probabilities of six groups of conceptual structure.

5.9.1 *Cumulative probabilities*

For the analysis of conceptual structure with six categories of outcome variable, first SAS selects category “A” and calculates the probability that category “A” occurred. The logic behind finding the probability of the category “A” is the same as binary logistic regression in which we fit the observed data to the equation 5-15 and estimated the intercept μ_A and regression coefficient β.

$$\gamma_A = \frac{1}{1+e^{-(\mu_A + ßx\ )}} \qquad \text{Equation 5-15}$$

Instead of finding the probabilities for other categories, SAS calculates following five cumulative probabilities γ_A, γ_B, γ_C, γ_D, γ_E and γ_F, which are described below:

$$\gamma_A = Ԓ_A \qquad \text{Equation 5-16}$$

$$\gamma_B = Ԓ_A + Ԓ_B \qquad \text{Equation 5-17}$$

$$\gamma_C = Ԓ_A + Ԓ_B + Ԓ_C \qquad \text{Equation 5-18}$$

$$\gamma_F = Ԓ_A + Ԓ_B + Ԓ_C + Ԓ_D \qquad \text{Equation 5-19}$$

$$\gamma_E = Ԓ_A + Ԓ_B + Ԓ_C + Ԓ_D + Ԓ_E \qquad \text{Equation 5-20}$$

$$\gamma_F = 1 \qquad \text{Equation 5-21}$$

Equations 5-16 to 5-21 represent cumulative probabilities. In terms of our groups, γ_A is the probability that a students’ answer will display results so that it will be classified into Group A. γ_B is the probability of students’ answer being classified in either Group A or

Group B; and so forth. Thus, γ_F is the probability of the students' response being placed in *any* category and must be 1.

Again the same logic as binary logistic regression can used for estimating the probabilities of γ_A, γ_B, γ_C, γ_D and γ_E. We used equation 5-22 and estimated the best fit to find five logistic models for cumulative probabilities γ_A, γ_B, γ_C, γ_D and γ_E.

In equation 5-22, "j" is a dummy variable that runs from "A" to "F" and γ_j is the probability for each group. The independent variable x stands for RTOP overall score. The intercept μ accepts six different indices in the six different estimated models, which runs from μ_A to μ_F. The coefficient of regression that is shown as β remains constant since we estimated and compared the models based on the change of intercepts.

$$\gamma_j = \frac{1}{1+e^{-(\mu_j + ßx)}} \qquad \text{Equation 5-22}$$

For the sake of simplicity, we can also use inverse of equation 5-22, called logit function (equation 5-23) to represent the probabilities. The advantage of using logarithmic form of probability is its linearity. Logit is the logarithm of the odds; similarly, the difference between the logits of two probabilities is the logarithm of the odds ratio (*R*). For this analysis, the logits are simply parallel linear lines since the β remains constant, and we can compare logits based on their intercepts

$$\text{Logit}(\gamma_j) = \log(\gamma_j / (1-\gamma_j)) = \mu_j + ßx \qquad \text{Equation 5-23}$$

Again, "j" is a dummy variable that runs from "A" to "F" and γ_j is the probability for each group. The independent variable x stands for RTOP overall score. The intercept μ_j

accepts six different indices in the six different estimated models which runs from μ_A to μ_F and β is the coefficient of the regression.

5.9.2 *Data Analysis: Tables and graphs*

We applied SAS package version 9.2, and followed the procedure that I explained in the section 5-9-1 to run multinomial logistic regression for the six categories of the conceptual structure. Using the procedure that we described in the previous section we fit the equation 5-22 to the observed data for the six categories and derived μ_A to μ_F and β . These values are listed in Table 5-14.

Table 5-14 Intercepts and regression coefficients for cumulative probabilities

Parameter	Estimate	Pr>ChiSq
Intercept μ_a	-0.33	0.043
Intercept μ_b	0.45	0.0039
Intercept μ_c	1.70	<0.0001
Intercept μ_d	3.20	<0.0001
Intercept μ_e	4.19	<0.0001
Coefficient ß	-0.69	<0.0001

Using Table 5-14 and values we derived for μ_A to μ_F and β. By inserting these values in equation 5-22 we found the cumulative probabilities as described through equations

$$\gamma_A = \frac{\mathbf{1}}{\mathbf{1+e^{-(-0.33-0.69x)}}}$$ Equation 5-24

$$\gamma_{\mathbf{B}} = \frac{\mathbf{1}}{\mathbf{1+e^{-(0.45+0.694x)}}}$$ Equation 5-25

$$\gamma_{\mathbf{C}} = \frac{\mathbf{1}}{\mathbf{1+e^{-(1.70+0.69x)}}}$$ Equation 5-26

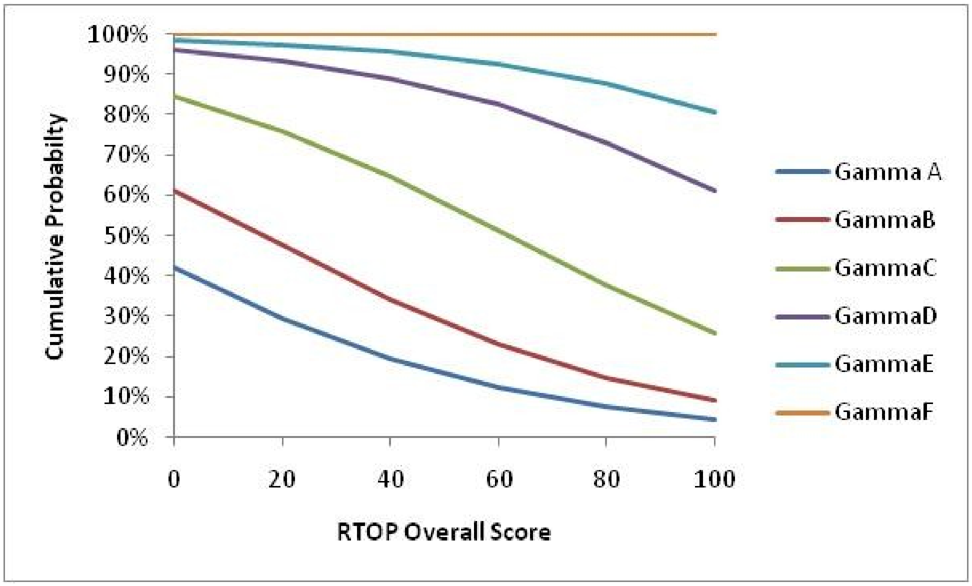
$$\gamma_D = \frac{1}{1+e^{-(3.20+0.69x)}}$$ Equation 5-27

$$\gamma_E = \frac{1}{1+e^{-(4.19-0.69x)}}$$ Equation 5-28

$$\gamma_{F=1}$$ Equation 5-29

Figure 5-9 shows the series of translational curves that are displaced from the base curve Gamma A. These series of curves are graphical representations of likelihood functions of cumulative probabilities determined by equations 5-24 to 5-29. The base curve that is indicated by dark blue colored curve, shows the probability of student displaying a group "A" link in his/her answer as a function of RTOP. The probability of student displaying either a group "A" or group "B" type answer is given by the burgundy colored curve. The probability of a student displaying a group "A", "B" or "C" answers as a function of RTOP score is given by the green curve, etc. Apparently, the cumulative probabilities are successively increasing and finally the probability that student's answer to be at least in one of the groups is equal to one that is indicated by Gamma F.

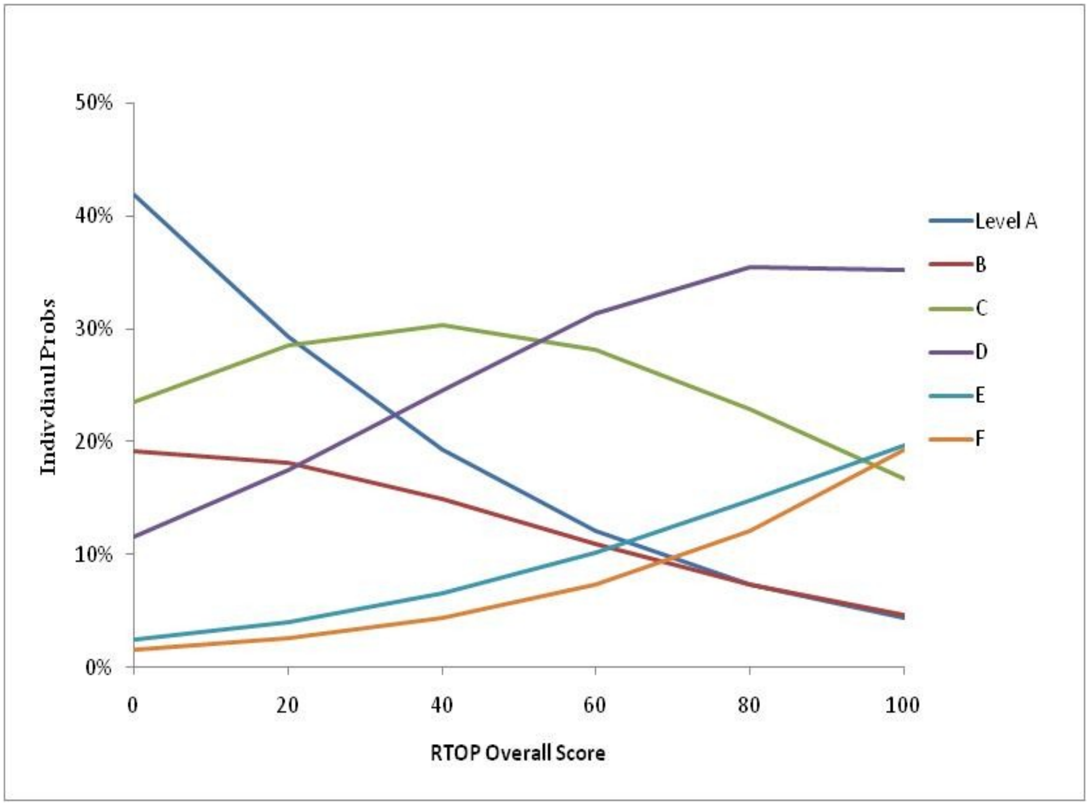
Figure 5-9 Cumulative probability versus RTOP overall score



5.9.3 *Individual probability*

After deriving the cumulative probabilities that are described by equations 5-24 to 5-29, we can use equations 5-16 to 5-21 to find individual probabilities Ԓ$_{A}$, Ԓ$_{B}$, Ԓ$_{C}$, Ԓ$_{D}$, Ԓ$_{E}$. We found the values of Ԓ$_{A}$, Ԓ$_{B}$, Ԓ$_{C}$, Ԓ$_{D}$, Ԓ$_{E}$ in an Excel spreadsheet and drew the plots below for the and individual probabilities (Figure 5-10).

Figure 5-10 Individual probabilities in terms of RTOP overall score



In Figures 5-10, we demonstrated the graphical representation of the probability distributions for individual probabilities. The curves colored dark blue, green and burgundy are belonging to the categories "A","B" and "C" , which are representing higher-level and middle-level links. As we can see, all three curves show decrease as we move toward higher RTOP scores. However, the lower-level links, which are "D", "E" and "F", show increase as we move toward higher RTOPs.

No reasonable explanation can support the causality that students in higher RTOP courses exhibited lower level links. Students demonstrated either lower level links because teachers administrated questions that required lower level links or if the questions required higher level links the student were not successful in applying higher order conceptual links. In sum, we observed on the students' written responses that the usage of higher order conceptual structures were less prevalent in higher RTOP courses. At first glance, this result may seem to be a negative statement about inquiry-based classes. That is, students in traditional classes are more likely to use and link higher-level concepts than students in reformed classes (as measured by the RTOP). However, inquiry-based classes frequently focus much of the instruction on observation, measurement and reasoning from the students own observations. Students in traditional classes are more likely to be introduced theoretical concepts and thus are more likely than the inquiry students to display higher-level concepts in their answers on written exams. Thus, it is not clear whether these results indicate that the students in the inquiry classes are not able to use the high-level concepts and links or have just not had sufficient practice in doing so. Unfortunately, answering that question is beyond the scope of the current study.

5.10 Summary

In this chapter, I explained the inferential statistical methods we used to analyze the data that we collected. I defined the parameters of the hypotheses for our collected data and explained the statistical procedures that we adopted to formalize the relationships between the parameters of hypothesis.

We applied logistic regression to a sample of 18 courses with about 900 students and reported the results for performing binary logistic regression to estimate the relationship between traits of reasoning and RTOP overall score. Multiple logistic regressions allowed us to take in to account partial RTOP scores in our model.

In addition, we studied conceptual structure of students' responses and based on conceptual classification schemes categorized students' responses into six groups. The outcome variable with six categories was multinomial instead of binomial that required more complicated data analysis. A summary of the results of these analyses is presented in Chapter 6.

Chapter 6 - Conclusion and Implications

6.1 Research summary

As a part of National Study of Education in Undergraduate Science (NSEUS) project, we studied the effect of active engagement on the reasoning abilities of pre-service teachers taking undergraduate science courses. One of the goals of NSEUS was to explore the relationship between students' learning of content knowledge and the degree that science courses were geared toward interactive engagement teaching-learning strategies. In the first chapter of this study, I provided an overview of the overarching NSEUS project, concerning the chronological development of the background research and theoretical framework behind NSEUS. I reviewed the history of the events that progressively led to the organization of the above-mentioned nationwide project, the set of overarching questions that the project seeks to address, the principal research methodology, the instruments, and surveys applied in the project. I reviewed the scientific literature in the second chapter. In the following chapters -- three, four and five -- I described how we addressed the research questions and discussed the underlying structures informing those questions. In addition, I discussed our content analysis strategy and presented the results of the data that we analyzed.

6.2 Research questions

In particular, we focused to study the following four major research questions:

Q1: How do we elicit students' reasoning using written content questions?

Q2: How do we classify students' reasoning in terms of the quality of their responses to written content questions?

Q3: What is the relation between the quality of students' thought processes as displayed on written content questions and the degree to which course is considered to be reformed (in terms of RTOP measure)?

Q4: What is the relation between the conceptual structure of the responses and the degree to which course is considered to be reformed (in terms of RTOP measure)?

6.2.1 *Addressing qualitative research questions: questions one and two*

In chapters three and four, we addressed the first two questions. We devised a protocol to design written content questions with a controlled level of cognitive load. We used a common template for different disciplines that required students to recognize and generalize the relevant facts or concepts and their interrelationships to suggest an applicable or plausible theory. By developing the question development protocol, we defined criteria that we should include in the structure of our question design as the characteristics of reasoning.

The universities that participated in our study were nationally distributed and the courses investigated at these universities covered a variety of science disciplines. Accordingly, a direct comparison on subject matter learning was impossible. Instead, we concentrated on comparisons of reasoning skills within the content that the students had learned and developed a common template to elicit reasoning skills. Using this template, we designed questions that elicited same level of thought processes regardless of the scientific content of the study. In other words, we developed content questions in physics, biology, geology, astronomy and chemistry with same level of though processing. We developed a rubric for comparing the students' level of thought processing as reflected in their responses to content written questions.

6.2.2 *Addressing quantitative research questions:*

Questions three, four

Along with analyzing students' reasoning, we visited 18 science courses at 13 different universities and observed the science classes in which the elementary education majors were enrolled. We ranked the courses with respect to characteristics that are valued for inquiry courses.

Using our written content questions, we collected large amount of data and adopted several statistical models to describe the relation between thinking processes and different measures on inquiry. In chapter five, we used the logistic regression statistics to address questions three and four. We applied binary logistic regression to describe the relation between the quality of students' thought processes and the degree to which course is considered to be reformed in terms of RTOP measure. We performed two versions for our analysis including simplified and generalized version of logistic regression. The difference between simplified and generalized version of logistic regression lies in the number of independent variables that were accounted for in the model. In the simplified version, we ran the analysis with one independent variable in the model and in the generalized version; we derived the model in terms of two independent variables.

To address question four, we performed multinomial logistic regression to find the relation between the conceptual structure of the responses and the degree to which course is considered to be reformed. The data were a good fit for the multinomial logistic regression when the outcome variable was conceptual structure with six categories including different levels of conceptual structure and the predictor was RTOP overall score.

We conclude that for research questions three and four we compared students' levels of thought processes across different RTOP scores. We collected data across disciplines and represented the relationships between students' reasoning and RTOP score using several logistic regression models. Consequently, we covered the scope of the quantitative research questions as we used our protocol and analyzed the large amount of the collected data.

6.2.3 *Data analysis results for overall RTOP score*

In the simplified version, we obtained the relation between the quality levels of students' thought processes and the RTOP overall score. We obtained three regression models for knowledge types (factual, conceptual, and procedural) and four models for cognitive levels (compare, infer, explain, and apply). In sum, the variations of RTOP overall scores were between 20/100 and 90/100. There was a correlation between the RTOP overall score with the likelihood of the traits being displayed in the students' written answers. In other words, as RTOP overall score increased factual, conceptual, compare and apply increased slightly, however, infer and explain remained constant and procedural knowledge decreased from 50% to 25%.

6.2.3.1 Interpreting results-overall score

For the simplified version of analysis, we obtained that evidence for the occurrences of the knowledge and cognitive traits were dependent on overall RTOP scores with higher RTOP scores indicating higher levels of these traits in the students' answers. While, the results showed gradual increase for the traits factual, conceptual, compare and apply, there was no significant change for infer and explain and there was a steep decrease for procedural knowledge as RTOP scores increase.

Because RTOP scores for a class are measure of the level of interactive engagement, results indicate that for the students in an interactive engagement class were more likely to show higher levels of reasoning on their written exam questions.

Accordingly, for the question three we concluded the higher RTOP overall scores were correlated with the better evidence of facts and concepts in the students' responses. Regardless of how they improved to reconstruct the knowledge of facts and concepts in connection with the features of question, they did not improve further to develop higher cognitive processes such as infer and explain to present a plausible answer.

6.2.4 *Data analysis results for several RTOP scores*

In the second version of the analysis, we made use of two predictors. The predictors were characteristics of the courses that were obtained from RTOP sub scores. Initially we considered five predictors for our analysis derived from the categories of RTOP sub scores including; the quality of lesson design, propositional knowledge, procedural knowledge, communication/interaction and student/teacher relationship. However, for the data we collected from 18 science courses, the measures for lesson design, procedural knowledge, and communication/interaction and student/teacher relationship were highly correlated and thus violated the assumption of lack of multicollinearity, which is needed for logistic regression. To remedy the issue of the multicollinearity, we removed highly correlated variables and replaced them by a combination score that was obtained by averaging the scores of lesson design, procedural knowledge, and communication/interaction and student/teacher relationship. Consequently, the RTOP categories reduced to only two categories including the combination score and propositional knowledge. We fit the data to the generalized logistic model with two predictors and found the trend of the relationships.

6.2.5 *Interpreting results- logistic model with several predictors*

The results showed as the propositional knowledge RTOP score increased, the likelihood of the evidence of factual knowledge and conceptual schema also decreased in the students' responses. In contrast, the combination score positively affects the likelihood of evidence of factual knowledge and conceptual schema. For the procedural knowledge, the same trend happened for factual knowledge and conceptual knowledge, however, the propositional knowledge shows a stronger negative effect on procedural knowledge rather than factual knowledge and conceptual knowledge.

Propositional knowledge RTOP score showed a negative effect on the likelihood of the traits including; compare, Infer, explain. In contrast, propositional knowledge had a positive effect on students' cognitive process of apply. The combination score was positively correlated to the likelihood of students showing cognitive process of compare. However, the effect of combination score was negligible on other cognitive skills.

6.3 Implications for further research

6.3.1 *Assessment tools*

The findings of this study have a number of important implications for future practice. In addition to further investigating the effectiveness of interactive-engagement approach for teaching, our research focused on alternative learning goals in learning science aside from the usual problem-solving and conceptual learning goals. Up to now, most studies in physics education research or other disciplines have based on the use of interviews to probe into students' reasoning or measuring students' pre and post understanding using different sort of concept inventories. In the current study, we have laid out an alternative research methodology that utilizes written content questions.

There are few advantages in making use of content questions as an assessment instrument. By using written type of assessment tools, more students become accessible, as the questions can be administrated to distant locations or it can be administrated online and the approach is less costly and time consuming.

In addition, content questions are comprehensive assessment instruments in comparison to contemporary standardized tests such as concept inventories and the Lawson test. Lawson's Classroom Test of Scientific Reasoning assesses students' abilities in six dimensions including conservation of matter and volume, proportional reasoning, control of variables, probability reasoning, correlation reasoning, and hypothetical-deductive reasoning. Many instructors particularly those who are interested in adopting inquiry oriented teaching strategies are seeking a more comprehensive assessment instrument that goes beyond measuring the correctness. They would like to evaluate their instructional methods through a broader assessment of their students' skills, including experimental and mathematical skills, problem solving and critical thinking skills. Our assessment tool can be considered as an additional assessment tool for measuring students' cognitive abilities in a new context related to what students have already learnt.

Furthermore, our product can serve as a standard type of assessment that could compare students' performance on the nationwide basis. The current study can be used much more extensively in further research for developing on-line content question tools for automating the process of assessing students' reasoning.

6.3.2 *Exploring hidden influential parameters*

Earlier we explained the results of our logistic regression analysis for several predictors. We conclude RTOP-"combination score" had a positive effect on students' reasoning and the effect of RTOP on" propositional knowledge" was negative in most of the cases. However, these effects do not necessarily imply causation, but they do imply causation if we perform a controlled experiment and keep other variables under control. Our study initially was designed for exploring the trend of relationships and correlations. As a result, we gained insight to suggest possible causes that require further investigation; in other words, correlation can be a hint to a direct cause and effect relationship. Since propositional knowledge involves teacher's competency about knowledge, promoting coherent understanding, encouraging appropriate abstraction, and exploring real-world phenomena, we expected that the higher scores for the propositional knowledge should improve the likelihood[2] of the traits. However, we obtained opposite results to what we expected. This may suggest that we have not performed a controlled experiment and another hidden variable in higher RTOP classes affected the projection of the propositional knowledge.

---

[2] Likelihood of a trait measures the probability of observing a certain trait

For example, refer to our conceptual structure analysis; the high-level concept links were less prevalent in students' responses when RTOP scores were higher. Concurrently, the negative coefficient of propositional knowledge in logistic regression model means that propositional knowledge, as measured by the content sub-scale of the RTOP, negatively influenced the probability of the traits.

We speculate the absence of using High-Level concept links in the courses with higher RTOP courses can be a hidden factor that influences the coefficient less effectiveness of propositional knowledge.

6.3.3 *Pattern we observed in the RTOP scores*

As I mentioned earlier, for the 18 courses that we collected data some of the items of RTOP were highly correlated. We ran Pearson correlation test and the highly correlated variables were lesson design, procedural knowledge, communicative-interactions and student/teacher relationships. These findings have important implications for future use of RTOP. Moreover, the fact that four measurements of lesson design, procedural knowledge, communicative-interactions and student/teacher relationships were measuring the same things was an intriguing problem that could be usefully explored in further research.

6.4 Implications for classroom teaching

6.4.1 *Lesson planning and assessment*

While it is hugely valuable to emphasize reform teaching, the main implication of this study is the protocol that we developed to design questions and the rubric for analyzing students' responses. The protocol is particularly designed for elementary education majors, however, teachers of other disciplines or courses can follow the same approach to design lessons, questions and rubrics. The whole approach toward developing the assessment protocol resembled the *"Backward design"* strategy (Wiggins & McTighe, 1998). We defined three stages for developing our assessment design namely; identifying desired results, determining what are the acceptable evidences for a well-reasoned response and planning the question design accordingly.

Learning taxonomies can serve to define teacher's learning goals for his/her class and thereby shape the lesson design learning experience. After identifying learning goals, learning taxonomies can be used to help design rubrics for assessment and research purposes. By using this strategy, teachers will target higher levels of thought processes and levels of abstraction rather than introducing certain content, concept or procedure.

6.4.2 *Question bank*

When we were at the stage of administering content questions, we shared the question development protocol with our collaborators who are science experts in different disciplines. Based on their understanding of the protocol, they suggested a collection of content questions that is presented in Appendix (E).

6.5 Future research

Our study has provided some understanding about students reasoning skills and recognizing how the type of instruction may affect students' reasoning. At the same time, our research opened avenue for further unanswered questions. Logistic regression is an appropriate method for situations where we have binary or categorical outcomes. In the current study we collapsed, our analysis into binary systems, however, the procedure can be expanded for several categorical responses together with categorical or continuous predictors. Therefore, our research design can be improved if we develop more sensitive rubrics that take into account several categorical possibilities for each trait.

Moreover, logistic regression has flexibility to consider and compare the effect of many variables simultaneously. For example, it would be valuable to take into account the students' prior knowledge in the logistic model. If we measure students' reasoning skills for example using the Lawson test (Lawson, 1978) scores, we can consider students' level of intellectual development as another independent variable in our logistic model. We can also take into account the effect of other variables such as gender, age, type of school, students' background etc.

Our study seems to teach us much about parameters that were correlated with students' reasoning; however, more controlled experiments would address our doubts. Further work needs to be done with controlled experiment conditions to reveal causes that are hidden. For example, we can modify our research design to perform experiments in the courses that apply same level of conceptual structure.

A further study also could assess multivariate analysis to determine the variables that change in respond to the change of others.

References

Alison, P. (2008). *Logistic regression using SAS, theory and application* (7th ed.). North Carolina: SAS Institute Inc.

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

Anderson, D.L., Fisher, K.M., Norman G.J. (2002). Development and evaluation of the conceptual inventory of natural selection. *Journal of Research in Science Teaching, 39*, 952–978.

Anderson, L. & Krathwohl, D. A. (2001). *Taxonomy for learning, teaching and assessing: A revision of bloom's taxonomy of educational objectives*. New York: Longman.

Atkinson, R. L., Atkinson, R. C., Smith, E. E., & Bem, D. J. (1993). *Introduction to psychology* (11th ed.). Fort Worth, Texas: Harcourt Brace Jovanovich.

Aufschnaiter, C. V., Euduran, S., Osborne, J., & Simon, S. (2008). Arguing to learn and learning to argue: Case studies of how students' argumentation relates to their scientific knowledge. *Journal of Research in Science Teaching, 45*(1), 101-103.

.

Ausubel, D. P. (1977). The facilitation of meaningful verbal learning in the classroom. *Educational Psychologist, 12*, 162-178.

Ausubel, D. P., Novak, J. D., & Hanesian, H. (1978). *Educational psychology: A cognitive view* (2nd ed.). New York: Holt, Rinehart and Winston.

Bandura, A. (1986). *Social foundations of thought and action*. Englewood Cliffs, NJ: Prentice-Hall.

Bao, L., Cai, T., Koenig, K., Fang, K., Han, J., Wang, J., Liu, Q., Ding, L., Cui, L., Luo, Y., Wang,Y., Li, L., Wu, N. (2009). Learning and scientific reasoning. *Science, 323*, 586–587.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*, 612–637.

Beach, K. (1999). Consequential transitions: A socio cultural expedition beyond transfer in education. *Review of Research in Education, 24*, 101-139.

Beatty, I.D., & Gerace, W.J. (2002). Probing physics students' conceptual knowledge structures through term association. *American Journal of Physics, 70*(7), P.750.

Bennett, C. and Dewar, J. (2007). Developing and applying a taxonomy for mathematical knowledge-expertise. *Electronic Proceedings of the Tenth SIGMAA Conference on*

Research in Undergraduate Mathematics Education, http://cresmet.asu.edu/crume2007/eproc.html.

Bloom, B. S., & Krathwohl, D.R. (1956). *Taxonomy of educational objectives: Handbook 1, the cognitive domain* . New York: D. McKay Co., Inc.

Bransford, J. D., & Schwartz, D. L. (1999).Rethinking transfer: A simple proposal with multiple Implications . *Review of Research in Education, 24*, 61–100.

Bransford, J., Brown, A.L. & Cocking R.R. (1999). *How people learn: Brain, mind, experience, and school*. Washington, D.C.: National Academy Press for National Research Council.

Bruner, J. (1960). *The process of education*. Cambridge, Mass: Harvard University Press.

Bruner, J. S. (1966). *Toward a theory of instruction*. Cambridge Mass: Harvard University Press.

Bybee, R. W. (2000). Teaching science as inquiry. In *Inquiry to inquiry, learning and teaching in science*. Washington, DC: American Association for the Advancement of Science.

Chambers, D. W. (1983). Stereotypic images of the scientist: The draw- a-scientist test. *Science Education, 67*(2), 255-265.

Chi,M., Slotta, J. & Deleeuw, N. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction, 4*, 27-43.

Cobern, W. W. (2000). *The thinking about science survey instrument (TSSI) – SLCSP 151.* Kalamazoo, MI: Scientific Literacy and Cultural Studies Project.

Coletta, V.P., & Philips,J.A. (2005). Interpreting FCI scores: Normalized gain, pre instruction scores, and scientific reasoning ability. *American Journal of Physics, 73*(12)

Costenson, K. & Lawson, A.E. (1986). Why isn't inquiry Used in more classrooms? *The American Biology Teacher, 48*(3), 150-158.

Cresswell, J. W. (1997). *Qualitative inquiry and research design: Choosing among the five traditions* Thousand Oaks: Sage Publication.

Cummings, K., Marx, J., Thornton, R., & Kuhl, D. (1999). Evaluating innovations in studio. *American Journal of Physics, 67*(7), S38-S44.

Dawson, C. J. (1978). Pupils difficulties: What can the teacher do? *Education in Chemistry, 133*(15), 120-121.

Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process*. Boston: D.C. Heath.

Dewey, J. (1938). *Experience and education*. New York: Collier Books.

Dewey, J. (1982). (Originally published in 1910). *How we think*. Lexington, MA: Heath.

Druckman, D., & Bjork, R.A. (1994).*Learning, remembering and believing*. Washington, D.C.: National Academy press.

Dufresne, R., Mestre, J., Thaden-Koch, T., Gerace, W., Leonard, W. (2005). Knowledge representation and coordination in the transfer process. In J. P. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (). Greenwich, CT: Information Age.

Dyson, A. H. (1999). Transforming transfer: Unruly children, contrary texts, and the persistence of the pedagogical order. *Review of Research in Education, 24*, 141-172.

Ebenezer, J., Kaya, O. N., & Ebenezer, D. L. (2011). Engaging students in environmental research projects: Perceptions of fluency with innovative technologies and levels of scientific inquiry abilities. *Journal of Research in Science Teaching, 48*(1), 94-116.

Ed, Y. (2009). Teaching scientific knowledge doesn't improve scientific reasoning. Message posted to http://scienceblogs.com/notrocketscience/2009/01/teaching_scientific_knowledge_doesnt_improve_scientific_reas.php

Ensook, K., & Sung-Jae,P. (2002). Students do not overcome conceptual difficulties after solving 1000 traditional problems. *American Journal of Physics, 70*(7), P.759.

Field, A. (2009). *Discovering statistics unsing spss* (3rd ed.). London, New Dehli: Sage Publications Ltd.

Finely, F.N., & Pocovi, M.C. (2000). Considering scientific method of inquiry. In *Inquiry to inquiry, learning and teaching in science.* Washington, DC: American Association for the Advancement of Science.

Finkelstein, N., et al. (2005). When learning about the real world is better done virtually: A study of substituting computer simulations for laboratory equipment. *Physical Review, Special Topics: Physics Education Research, 1*(010103)

Flavell, J. H. (1979). Meta-cognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 3*(10), 906-11.

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

Greeno, J., Smith, D., & Moore, J. (1993). Transfer of situated learning. In D. K. Detterman & Sternberg R. J. (Ed.), *Transfer on trial: Intelligence, cognition, and instruction* (pp. 99-167). Norwood, NJ: Albex.

Hake, R. R. (1992). Socratic pedagogy in the introductory physics lab. *Phys. Teach, 30*, 546-552.

Harrow, A. J. (1972). *A taxonomy of the psychomotor domain*. New York: David McKay Co.

Herron, J. D. (1975). Piaget for chemists. *Journal of Chemical Education, 52*(3), 146-150.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher, 30*, 141-151.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression*. New York: John Wiley & Sons Inc.

Hurd, P.D., Bybee, R.W., Kahle, J.B., & Yager,R. (1980). Biology education in secondary schools of the united states. *The American Biology, 42*(7), 388-410.

Hyde, D. M. G. (1970). *Piaget and conceptual development*. London, New York: Holt, Rinehart and Wilson.

Inhelder & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence; an essay on the construction of formal operational structures*. New York: Basic books.

Jaakkola, T., Nurmi, S. & Veermans, K. (2011). A comparison of students' conceptual understanding of electric circuits in simulation only and simulation-laboratory contexts *Journal of Research in Science Teaching, 48*(1), 71-93.

Jenkins, E. W. (1978). Piaget and school chemistry: A critique. *Education in Chemistry, 15*, 85-86.

Karplus, R., & Lavatelli,C. (1969). *The development theory of Piaget: Conservation*. San Fransisco: Davidson Film Procedures.

Karplus, R., Pulos, S., & Stage, E. K. (1983). Early adolescents' proportional reasoning on "rate" problems. *Educational Studies in Mathematics, 14*, 219-234.

Keselman, A. (2003). Supporting inquiry learning by promoting normative understanding of multivariable causality. *Journal of Research in Science Teaching, 40*(9), 898-921.

Klapper, M. H. (1995). The scienc teacher. *Research Library, 62*(6), p.36.

Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of educational objectives; the classification of educational goals. handbook II: The affective domain*. New York: Longman, Green.

Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life*. New York: Cambridge University Press.

Lave,J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation (learning in doing: Social, cognitive and computational perspectives)*. Cambridge: University Press.

Lawson, A. E. (1978). The development and validation of a classroom test of formal reasoning. *Journal of Research in Science Teaching, 15*(1), 11-24.

Lawson, A.E., Alkhoury, S., Benford, R., Clark, B.R. & Falconer, K.A.(2000). What kinds of scientific concepts exist? concept construction and intellectual development in college biology. *Journal of Research in Science Teaching, 37*(9), 996-1018.

Lawson, A.E., & Thompson, L.D. (1988). Formal reasoning ability and misconceptions concerning genetics and natural selection. *Journal of Research in Science Teaching, 25*(9), 733-746.

Lederman, N. G. (1992). Students' and teachers' conceptions of the nature of science: A review of the research. *Journal of Research in Science Teaching, 29*(4), 331-359.

Lederman, N. G. (2004). Syntax of nature of science within inquiry and science instruction. In Flick, L. B., Lederman, N. G. (Ed.), *Scientific inquiry and nature of science* (pp. 301-317). Netherlands: Kluwer Academic Publishers.

Lederman, N. G., & Lederman, J. S. (2004). Revising instruction to teach nature of science. *The Science Teacher, 71*(9), 36-39.

Libarkin, J.C., & Anderson, S.W. (2005). Assessment of learning in entry-level geoscience courses: Results from the geoscience concept inventory. *Journal of Geoscience Education, 53*, 394-401.

Likert, R. A. (1932). *Technique for the measurement of attitudes*. New York: Colombia University Press.

Lobato, J. (2003). How design experiments can inform a rethinking of transfer and vice versa. *Educational Researcher, 32*(1), 17-20.

Lunar phases. (2001,). Message posted to http://en.wikipedia.org/wiki/Lunar_phase

MacIsaac, D.L. & Falconer, K.A. (2002). Reforming physics education via RTOP. *The Physics Teacher, 40*(8), 479-485.

MacIsaac, D.L., & Falconer, K.A. (2002). Reforming physics education via RTOP. *The Physics Teacher, 40* (8), 479.

Maloney, D., O'Kuma, T.L., Hieggelke, C.J., Van Heuvelen, A.(2001). Surveying students' conceptual knowledge of electricity and magnetism . *American Journal of Physics, 69*(S1), S12.

Martina Nieswandt & Kathrine Bellomo. (2009).Written extended-response questions as classroom assessment tools for meaningful understanding of evolutionary theory. *Journal of Research in Science Teaching, 46*(3), 333-356.

Marx, R.W., Blumenfeld, P.C., Krajcik, J.S., Fishman, B., Soloway, E., Geier, R., & Revital, T.T. (2004). Inquiry-based science in the middle grades: Assessment of learning in urban systemic reform. *Journal of Research in Science Teaching, 41*(10), 1063-1080.

Marzano, R. J., & Kendall, J. S. (Ed.). (2007). *The new taxonomy of educational objectives* (2nd ed.). Thousand Oaks, CA: Corwin Press.

Mayer, R. E. (1999). *The promise of educational psychology: Learning in the content areas* . Upper Saddle River NJ: Prentice Hall.

Mayer, R. E. (2002). Understanding conceptual change: A commentary Reconsidering conceptual change: Issues in theory and practice. In Limon, M., & Mason, L. (Ed.), *Reconsidering conceptual change: Issues in theory and practice* () Corwin Press,Thousands Oaks.

Mayer, R. E. (1975). Information processing variables in learning to solve problems. *Review of Educational Research, 45*, 525-541.

McBride, D.L., Zollman, D., & Rebello, N.S. (2010). Method for analyzing students' utilization of prior physics learning in new contexts. *Physical Review, Special Topics: Physics Education Research, 6*(6)

McDermott, L. C. (1984). Research on conceptual understanding in mechanics. *Physics Today, 37*(7), 24.

McDermott, L. C. and the Physics Education Group at the University of Washington (1996). *Physics by inquiry*. New York: John Wiley & Sons.

McKeough, A., Lupart, J., & Marini, A. (Eds.).(1995). *Teaching for transfer: Fostering generalization in learning*. NJ: Erlbaum: Mahawah.

McNally, D. W. (1974). *Piaget, education and teaching*. Edinburgh: New Educational Press.

Mulford, D.R., & Robinson, W.R. (2002). An inventory for alternate conceptions among first-semester general chemistry students, *Journal of Chemical Education, 79*(6), 739-744.

National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.

National Science Foundation. (2011). *NSEUS.* Retrieved 07/12, 2011, from http://nseus.org/ .

Nieswandt,M., & Bellomo,K. (2009). Written extended-response questions as classroom assessment tools for meaningful understanding of evolutionary theory. *Journal of Research in Science Teaching, 46*(3), 333-356.

Novak, J. (1978). An alternative to Piagetian psychology for science and mathematics education. *Studies in Science Education, 5*, 1-30.

Novak, J. D. (1998). *Learning, creating and using knowledge: Concept maps as facilitative tools in schools and corporations.* Mahwah, NJ: ,Lawrence Erlbaum.

Novak, J. D. (2006). The development of the concept mapping tool and the evolution of a new model for education: Implications for mathematics education *Focus on Learning Problems in Mathematics, 28* (3 & 4), 1-32.

Piaget, J. (1961). The genetic approach to the psychology of thought. *Journal of Educational Psychology, 52*(6), 275-281.

Piaget, J. (1967). *Biology and knowledge*. Chicago: Chicago University Press.

Piaget, J., & Inhelder, B. (1969). *The psychology of the child*. New York: Basic Books.

Piburn, M, Sawada, D, Turley, J, Falconer, K, Benford, R, Bloom, I, & Judson, E. (2000). *Reformed teaching observation protocol (RTOP) reference manual.* (Technical Report No. IN00-3). Tempe, AZ: Arizona Collaborative for Excellence in the Preparation of Teachers (ACEPT).

Porter, A. C., & Smithson, J. L. (2001). Are content standards being implemented in the classroom? A methodology and some tentative answers. In S. H. Fuhrman (Ed.), *From the capitol to the classroom: Standards-based reform in the states—One hundredth yearbook of the national society for the study of education, part II* (pp. 60-80). Chicago: University of Chicago Press.

Rebello, N. S., Zollman, D. A., Allbaugh, A. R., Engelhardt, P. V., Gray, K. E., Hrepic, Z. (2005). Dynamic transfer: A perspective from physics education research. In J. P. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (). Greenwich, CT: Information Age Publishing Inc.

Renner and Lawson. (1973). Promoting intellectual development through science teaching. *The Physics Teacher, 11*(265)

Resnick, L. (1987). *Education and learning to think*. Washington, D.C.: National Academy Press.

Rice, D. C., et al. (1998). Using concept maps to assess student learning in the science classroom: Must different methods compete. *Journal of Research in Science Teaching, 35*(10), 1103–1127.

Riggs, I. M., & Enochs, L. G. (1990). Toward the development of an elementary teacher's science teaching efficacy belief instrument. *Science Education, 74*(6), 625-637.

Royer, J., Mestre, J. & Dufresne, R. (2005). In J. Mestre (Ed.), *Introduction: Framing the transfer problem* [Transfer of learning from a modern multidisciplinary perspective]. Greenwich, CT: Information Age Publishing.

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

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

Russ, R.S., & Hutchison, P. (2006). It's okay to be wrong: Recognizing mechanistic reasoning during student inquiry. *7th International Conference of the Learning Sciences,*

Schwab, J. J. (1958). The teaching of science as inquiry. *Bulletin of the Atomic Scientists, 14*, 374-379.

Shavelson, R. J., Lang, H., & Lewin, B. (1993). On concept maps as potential "authentic" assessments in science. Indirect approaches to knowledge representation of high school science. *Educational Resources Information Center(ERIC),* , ED367691.

Shayer, M., & Adey, P. S. (Eds.). (2002). *Learning intelligence: Cognitive acceleration across the curriculum from 5 to 15 years*. Milton Keynes, UK: Open University Press.

Shymansky, J. (1984). BSCS programs: Just how effective were they? *The American Biology Teacher, 46*(1), 54-57.

Southerland, S. A., Gess-Newsome, J., & Johnston, A. (2003). Portraying science in the classroom: The manifestation of scientists' beliefs in classroom practice. *Journal of Research in Science Teaching, 40*, 669-691.

Sternberg, R. J. (1998). Principals of teaching for successful intelligence. *Educational Psychologist, 33*, 65-72.

Strauss A, C. J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques* Sage.

Sunal, C., Sunal, D.W., Sundberg, C., Mason, C., Zollman,, D., Lardy, C. & Matloob-Haghanikar, M. (2009). Point of view: How are we reforming teaching in undergraduate science sources? *Journal of College Science Teaching,* (November)

Sunal, D. W., Sunal,C. Sundberg, C., Mason, C., Zollman,, D., Lardy, C. (2008). Background research paper no. 2, what do we know about undergraduate science course reform? Message posted to www.nseus.org

Sawada, D. Piburn, M., Turley, J., Falconer, K., Benford, R., Bloom, I., & Judson, E. (2000). *Reformed teaching observation protocol (RTOP):Training guide.* (Technical

Report No. IN00-2). Tempe, Arizona: Arizona Collaborative for Excellence in the Preparation of Teachers (ACEPT).

Taylor, P. C., Fraser, B. J., & Fisher, D. L. (1997). Monitoring constructivist classroom learning environments. *International Journal of Educational Research, 27*(4)

Taylor, P. C., Fraser, B. J., & Fisher, D. L. (1997). Monitoring constructivist classroom learning environments. *International Journal of Educational Research, 27*, 293-302.

Teodorescu, R. (2009). *Cognitive development in introductory physics: A research-based approach to curriculum reform.* (PhD, The George Washington University). . (3359606)

Teodorescu, R., Bennhold,C. & Feldman, G. (2008 July 23-24,). Enhancing cognitive development through physics problem solving: A taxonomy of introductory physics problems. *Physics Education Research Conference Part of the PER Conference Series,* Edmonton, Canada. *, 1064*

Thornton, R. K. & Sokoloff D. R. (1998). *Assessing student learning of newton's laws, the force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula*. *American Journal of Physics, 66*, 338–352.

Thornton, R. K., Kuhl, D., Cummings,K., Marx, J. (2009). Comparing the force and motion conceptual evaluation and the force concept inventory. *Physical review special topics- Physics Education Research , 5*(1)

Vygotsky, L. S. (1978). In M. Cole M., John-Steiner V., Scribner S. and Souberman E. (Eds.), *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wenning, C. J. (2011). Experimental inquiry in introductory physics courses. *Journal of Physics Teacher Education, 6*(2), 10/11/2011.

Wiggins, G., & Mc Tighe, J. (1998). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.

Wiggins, G., & Mc Tighe, J. (2006). *Understanding by design*. New Jersey: Pearson Prentice Hall.

Zeilik, M., Schau, C., & Mattern, N. (1999). Conceptual astronomy II. replicating conceptual gains, probing attitude changes across three semesters. *American Journal of Physics, 67*(10), p.923.

Appendix A - Reformed Teaching Observation Protocol (RTOP)

National Study of
Education in Undergraduate Science
www.nseus.org

Reformed Teaching Observation Protocol (RTOP)

Daiyo Sawada Michael Piburn

External Evaluator Internal Evaluator

and

Kathleen Falconer, Jeff Turley, Russell Benford and Irene Bloom

Evaluation Facilitation Group (EFG)

Technical Report No. IN00-1

Arizona Collaborative for Excellence in the Preparation of Teachers

Arizona State University

I. BACKGROUND INFORMATION

Name of Teacher ________________________ Announced Observation?_________________

(yes or no, or explain)

Location of class ____

(district, school, room)

Years of Teaching _________________ Teaching Certification ______________

(K-8 or 7-12)

Subject Observed _____________________________________ Grade Level _______________

Observer _______________________________ Date of Observation _______________

Start time ______________________________ End time ________________________

II. CONTEXTUAL BACKGROUND ACTIVITIES

In the space provided below please give a brief description of the lesson observed, the classroom setting in which the lesson took place (spaces, seating arrangements, etc), and any relevant details about the students (number, gender, ethnicity) and teacher that you think are important. Use diagrams if they seem appropriate.

III. LESSON DESIGN AND IMPLEMENTATION

		Never Occurred				Very Descriptive
1)	The instructional strategies and activities respected students' prior knowledge and the preconceptions inherent therein.	0	1	2	3	4
2)	The lesson was designed to engage students as members of a learning community.	0	1	2	3	4
3)	In this lesson, student exploration preceded formal presentation.	0	1	2	3	4
4)	This lesson encouraged students to seek and value alternative modes of investigation or of problem solving.	0	1	2	3	4
5)	The focus and direction of the lesson was often determined by ideas originating with students.	0	1	2	3	4

IV. CONTENT

Propositional Knowledge

6)	The lesson involved fundamental concepts of the subject.	0	1	2	3	4
7)	The lesson promoted strongly coherent conceptual understanding.	0	1	2	3	4
8)	The teacher had a solid grasp of the subject matter content inherent in the lesson.	0	1	2	3	4
9)	Elements of abstraction (i.e., symbolic representation, theory building) were encouraged when it was important to do so.	0	1	2	3	4
10)	Connections with other content disciplines and/or real world phenomena were explored and valued.	0	1	2	3	4

Procedural Knowledge

11)	Students used a variety of means (models, drawings, graphs, concrete materials, manipulatives, etc.) to represent phenomena.	0	1	2	3	4
12)	Students made predictions, estimations and/or hypotheses and devised means for testing them.	0	1	2	3	4
13)	Students were actively engaged in thought-provoking activity that often involved the critical assessment of procedures.	0	1	2	3	4
14)	Students were reflective about their learning.	0	1	2	3	4
15)	Intellectual rigor, constructive criticism, and the challenging of ideas were valued.	0	1	2	3	4

V. CLASSROOM CULTURE

Communicative Interactions

16)	Students were involved in the communication of their ideas to others using a variety of means and media.	0	1	2	3	4
17)	The teacher's questions triggered divergent modes of thinking.	0	1	2	3	4
18)	There was a high proportion of student talk and a significant amount of it occurred between and among students.	0	1	2	3	4
19)	Student questions and comments often determined the focus and direction of classroom discourse.	0	1	2	3	4
20)	`There was a climate of respect for what others had to say	0	1	2	3	4

Student/Teacher Relationships

21)	Active participation of students was encouraged and valued.	0	1	2	3	4
22)	Students were encouraged to generate conjectures, alternative solutions strategies, and ways of interpreting evidence.	0	1	2	3	4

23)	In general the teacher was patient with students.	0	1	2	3	4
24)	The teacher acted as a resource person, working to support and enhance student negotiations.	0	1	2	3	4
25)	The metaphor "teacher as listener" was very characteristic of this classroom.	0	1	2	3	4

Additional comments you may wish to make about this lesson.

Continue recording salient events here.

Time	Description of Events

Adapted from Turley, J., Piburn, M., & Sawada, D. (2001).

Turley, J., Piburn, M., & Sawada, D. (2001, March). *Using the RTOP for feedback to student teachers: A metamorphosis of method.* Arizona Collaborative for Excellence in the Preparation of Teachers (ACEPT). A paper presented at the annual meeting of the National Association for Research in Science Teaching. St. Louis, Missouri.

Appendix B - Faculty Interview Questions

National Study of
Education in Undergraduate Science

Faculty Interview Questions

Code Number:

Interview Site:

Interviewer:

Notetaker:

Date:

Background: (CoRe)

1) How long have you been teaching science at the undergraduate level?

2) How long have you been teaching this "identified NOVA" or comparison course?

3) What other courses do you teach over a normal one-year period of time?

4) Have you taught at any other levels such as high school, community college, or graduate school? If so, for how long?

5) Have you participated in any university professional development for improving teaching? Please describe the extent of this experience.

6) Have you taken university level education courses such as teaching methods? If so, please elaborate (certification, education degree, etc.).

Course:

7) Describe your students' interest in this course and science in general.

8) What are the main goals that you wish your students to learn from this course? What should your students take away about science in general after taking this course?

9) What were the important knowledge and skills you needed to develop and teach this course?

10) Does the type of teaching (science instruction) relate to student interest and/or achievement in this course (e.g. lecture, hands-on, labs)? In what ways?

11) What were the significant barriers you overcame in planning and teaching this course? Compare this course to other courses you have taught at this academic level.

12) What advice would you give future faculty members when they start teaching about effective science instruction and/or strive to teach science effectively themselves?

Class Session: (CoRe) (Note to the interviewer: These questions should be based on the lesson observed, but if the lesson has been observed prior to the interview, adjust the questions accordingly.)

13) What will be the main ideas or concepts addressed during this class session or lesson?

14) Describe how you will teach these main ideas or concepts, and explain the rationale behind doing so.

15) How typical is this lesson for this class? If this is not typical, please describe a typical class session in this course.

16) Why is it important for students to know the aforementioned main ideas or concepts you taught during this class session?

17) What do you anticipate will be some difficulties and/or limitations connected with teaching these ideas or concepts?

18) What knowledge about students' thinking and/or learning influences your teaching of these ideas or concepts?

19) How will you assess students' understanding of, or confusion about, these ideas?

Appendix C - Elementary In-service Teacher Interview Questions

Code Number:

Interview Site:

Interviewer:

Note taker:

Date:

Background: (CoRe)

1) How long have you been teaching? What grade levels and number of years at each level have you taught? Have you been involved in any specialized teaching (i.e. as a departmentalized science teacher, etc.)?

2) Have you participated in professional development for improving your science teaching? Describe the extent of your professional development.

3) What university level science courses have you taken?

4) Have you taken any university level science education courses (i.e. teaching methods or content courses for education majors)? How many? What courses?

Science Courses taken at the University:

5) How would you define science or the nature of science? Has your definition of science and the scientific process changed over time due to a single university course or set of courses? If so, in what ways?

6) Has your understanding of science content (i.e. the main ideas or concepts) changed as a result of a single university course or set of courses? If so, in what ways?

7) Has your understanding of science teaching (i.e. pedagogy, methods, implementing curriculum) and the ways in which you teach science changed as a result of a single university course or set of courses? If so, in what ways?

8) Which instructional strategies (activities, assignments, etc.) did you experience as most beneficial to your learning science at the university?

9) What science content areas do you feel most (least) prepared to teach? Why?

Teaching Science at the Elementary School Level:

10) What science content areas are most/least important in your teaching at the elementary level? Why?

11) What do you feel is the best way to teach science in elementary classrooms? Why?

12) What barriers have you had to overcome in planning and teaching science?

13) How interested do your students seem to be in science?

14) What should your students take away from science in your class this year?

15) What is some of the important information that you would advise future teachers to take from their university science courses? What is the least important information to take away?

Science Lesson: (CoRe) (Note to the interviewer: These questions should be based on the lesson observed, but if the lesson has been observed prior to the interview, adjust the following questions accordingly.)

16) What will be the main ideas or concepts of this class session or lesson?

17) Describe how you will teach these main ideas or concepts, and explain the rationale behind doing so.

18) How typical is this lesson for this class? If this is not typical, please describe a typical class session in this course.

19) Why is it important for students to know the aforementioned main ideas or concepts taught during this class session?

20) What do you anticipate will be some difficulties and/or limitations connected with teaching these ideas or concepts?

21) What knowledge about students' thinking and/or learning influences how you teach the main ideas or concepts?

22) How will you assess students' understanding of, or confusion about, these ideas?

Appendix D - Undergraduate Student Focus Group Interview Questions

National Study of
Education in Undergraduate Science

Undergraduate Student Focus Group Interview Questions

Code Number:

Site Name:

Interviewer:

Notetaker(s):

Date:

College Science Experiences

1) Describe your interest in science.

2) What university level science courses have you taken?

3) How would you define science or the nature of science?

4) How has your definition of science changed due to the science courses you have taken in college? Which course(s) had the most influence? The least?

5) How has your attitude toward science changed as a result of the course(s) you have taken in college? Why did these course(s) change your view of science?

6) Describe how has your understanding of science content changed as a result of taking this course? (What have in general have you learned about science in this course?)

7) What specific activities or assignments enabled you to change your understanding of an issue in science or science content in this course? In other science courses?

8) Which instructional strategies and activities used in science courses so far did you feel were most beneficial for your learning?

Course Experience (Note to the interviewer: These questions should be based on the lesson observed, but if the lesson has been observed prior to the interview, adjust the following questions accordingly.)

9) What is a typical lesson like for this course; i.e., what normally happens during your classes?

10) What were the main ideas or concepts for <u>this</u> class session? What science concepts did you learn? Why is it important for you to understand these concepts?

11) What about these concepts did you find confusing before the lesson? What about these concepts do you, or do you not, find confusing after the lesson?

12) How did (will) the instructor assess student understanding of these concepts?

13) Did you feel that the teaching strategies used in today's lesson were effective for student understanding of the concepts covered in this lesson? Why or why not?

14) What would you have done to make the lesson more effective for your learning?

Science Teaching (education majors or adjusted questions for groups that only have non-education majors)

15) Have your ideas of science teaching changed as a result of taking this class or others at the college level? (How do you think that science should be taught?)

16) Do you think that you can become an effective science teacher? (Do you think that you could be an effective science teacher? Why or why not?)

17) What do you feel is the best way to teach science in elementary schools? Why?

18) What science content or courses do you feel most prepared to teach? (What science content do you feel that you would be prepared to teach if the moment arose?)

Appendix E - Suggested Questions experts involved in the NSEUS project

Note: The following are sample questions suggested by experts involved in the NSEUS project. However not all the question, were used in our study.

Question 1: Physics – Density

A lava lamp is a decorative lamp containing two liquids that do not mix. When cool, the heavier liquid, the lava, sits at the bottom of the lamp. However, after the light is turned on, the lava floats to the surface in blobs. Eventually, these blobs drop back down. Explain why this happens.

Question 2: Physics – Sound

The situation involves passengers who are riding on a Supersonic Concorde Airplane that is traveling at the speed of sound. There is a thunderstorm outside.

A. Explain why or why not the passengers can hear the event.

B. Compare the cases where the thunder happens behind or in front of the airplane.

Question 3: Botany - Genetics

You are given four plants. They are all from the same species but have different types of leaves (variegated or non-variegated) and flower color (yellow or white). You wish to determine which of the traits are dominant. You begin by breeding the plant with the variegated leaves and white flowers to the plant with non-variegated leaves and white flowers. The result is a plant with variegated leaves and yellow flowers.

a)What hypothesis can you form regarding which traits are dominant and which are recessive from this first cross? Justify your hypothesis based on the evidence.

b)Next, you breed a plant with non-variegated leaves and white flowers to a plant with non-variegated leaves and yellow flowers. The result is a plant with non-variegated leaves and yellow flowers.

Explain whether or not this result supports your hypothesis and why.

Question4 : Biology – Botany

Plants, are we define them today, have evolved over time.

a)Explain the evolutionary adaptations in angiosperms that resulted in their becoming the dominant group of plants in most environments today.

b)Describe physical factors on Earth that may have driven the domination of the angiosperms.

c)Discuss why the less complex plants have still survived.

Question 5: Chemistry – Chemical Bonding

If you live in a place that has a lot of snow and ice in the winter, then you have probably seen the highway department spreading salt on the road to melt the ice.

a) Explain how the chemical structure of salt affects the properties of the solution.
b) Describe why sugar, another common household item, does not have the same effect as salt when it is placed on ice.

c) Compare what happens to the chemical structure of salt when combined with water to the chemical structure of sugar when combined with water.

Question 6: Chemistry – Chemical Bonding

You are building a high temperature oven. Imagine that you are located in an area where all you have for building materials are blocks of salt and blocks of sugar.

a) Identify which block you will use, and explain why its properties are better for building the oven.

b) Considering the possibility of inclement weather, explain what happens to the chemical structure when the block of salt melts and compare that to what happens when the block of sugar melts.

Question 7: Biology – Food Webs

Use the freshwater food web below to answer the questions below.

A. Describe the flow of energy and matter through this food web.

B. Explain what might happen to this food web if the minnows became extinct. Why do you think this would happen?

Question 8: Physics – Phases of the Moon

Assume that during one day in the lunar calendar a resident of the northern hemisphere watches directly up into the sky and finds a crescent phase of the moon.

a) Draw the relative geometrical positions of the moon, Earth and sun.

b) Can you guess the time of observation?

c) Explain how you can guess the time of observation from the information presented above?

Question 9: Physics – Phases of the Moon

The moon rises and sets every day, like the sun. The sun always rises in the morning and sets in the evening. The moon, on the other hand, rises and sets at different times everyday.

a) In which phases in the lunar calendar are the moonrise and moonset at approximately the same time as sunrise and sunset? Draw the relative positions of the moon, sun and Earth and explain why that is true.

b) In what phases does the moon rise at midnight and set in the morning? Draw the relative position of the sun, Earth and moon for these phases. How do the geometrical positions of the moon, Earth and sun contribute to the times of moonrises and moonsets?

Question 10: Physics – Phases of the Moon

During the middle of the night, a student notices a quarter moon rising due east. Remember that the Earth rotates counterclockwise. Is this the first quarter or third quarter of the moon? Explain how you can tell.

Question 11: Physics – Phases of the Moon

Two students, one on the northern hemisphere and one on the southern hemisphere start observing the phases of the moon. The moon exhibits different phases as the relative geometry of the sun, Earth, and moon changes. The northern resident stands due south and the southern observer stands due to north to watch the moon. Over the calendar of a month they draw their observations.

a) Replicate and draw the sequence of phases for the northern hemisphere resident.

b) Replicate and draw the sequence of phases for the southern hemisphere resident.

c) Compare the sequence of phases for the northern and southern residents. Are they similar or different? Why? Use the relative geometrical position of the Earth, sun and moon to support your reasoning.

d) Search for the pattern between phases; which part of the moon is becoming more visible for northern phases? (Right or left?) Discuss your rationale.

e) Search for the pattern between phases; which part of the moon is becoming more visible for southern phases? (Right or left?) Discuss your rationale.

Question 12: Physics – Phases of the Moon

Is the length of time that the moon is above the horizon greater than, less than, or equal to the length of the time that the sun is above the horizon? Explain your reasoning. (Hint: is the phase of the moon exactly the same when the moon rises and when it sets?)

Question 13: Physics – Phases of the Moon

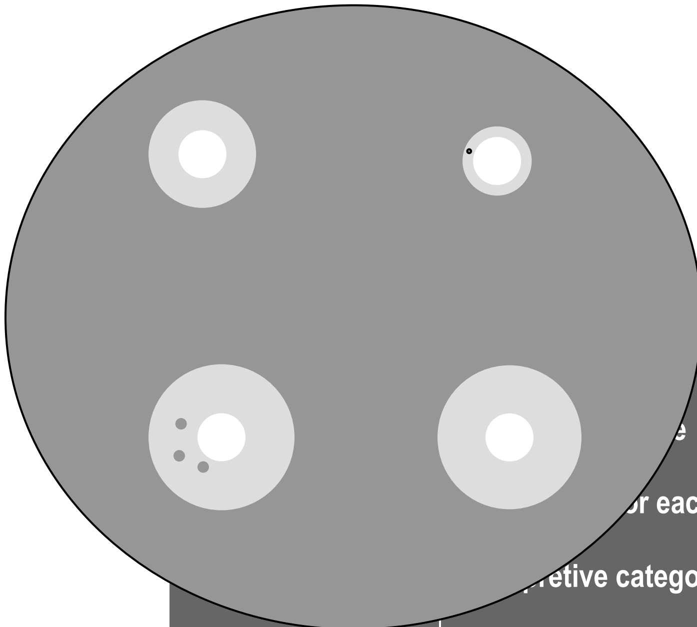
Is the length of time that the moon is above the horizon the same for different phases of the moon? Compare the length of time that the moon is above the horizon for three different phases - waxing crescent, first quarter and full moon. Explain why?

Question 14: Biology – Microbiology

You streak a plate of Tryptic Soy Agar with *Staphylococcus aureus* and add 4 antibiotic disks. The plates are incubated overnight at 37°C. Disk A contains tetracycline, you measure the zone of inhibition to be 20 mm. Disk B contains Streptomycin, you measure the zone of inhibition to be 12 mm, but note the colony growing within the zone that is about 1 mm from the disk. Disk C contains Methicillin/Oxacillin, you measure the zone to be 25 mm, but note several colonies growing within the zone between 5-10mm from the disk. Disk D contains Ampicillin, you measure the zone of inhibition to be 27 mm.

a) You are looking for a spontaneous mutant containing a multi-drug resistant plasmid. Which quadrant are you likely to find it in and why do you think so?

b) What steps do you need to take to see if the hypothesis you formed in question 1 is true? Draw your expected results.



	e r each retive category		
and microorganism	R*	I*	S*
Enterobacter aerogenes	≤11	12-13	≥14
Staphylococcus aureus	≤28		≥29
Streptococcus mitis	≤21		≥30
Ampicillin (10 microgram)			
Erythromycin (15 microgram)	≤13	14-17	≥18
Streptomycin (10	≤11	12-	≥15

microgram)		14	
Tetracycline (30 microgram)	<u><</u>14	15-18	<u>></u>19
Chloramphenicol (30 micro gm)	<u><</u>12	13-17	<u>></u>18
Methicillin/Oxacillin (1 microgram)	<u><</u>10	11-15	<u>></u>16

*Note: R=resistant; I=intermediary; S=susceptible

Quadrant C most likely contains the spontaneous multi-drug resistant mutants. 25mm is much greater than the16 mm zone of inhibition if the S. aureus culture is susceptible, but the presence of colonies 5-10 mm from the disk indicates that these colonies are resistant to Methicillin. The same logic can be applied to Quadrant B even though the culture is intermediary resistant to Streptomycin.

Hypothesis Testing

a. Test for contamination. The culture should be identified as Staphylococcus aureus. Use gram staining, etc. to verify.

b. Streak two plates -1 with original culture and 2 with the resistant colony(ies) and add the antibiotic disks ABCD to each plate.

c. The picture should indicate that the new colony is resistant to Methicillin.

d. Streak new plate with the original colony and Methicllin resistant plate colonies to test for resistance to other antibiotics. Methicillin resistant bacteria should be resistant to additional antibiotics.

Appendix F - Characteristics of different levels of answers

Table F-1 Steps of reasoning with corresponding thought processes

Steps of reasoning in answering a content question	Type of knowledge required	Type of cognitive process required
1. Retrieving and recognizing the pertinent information described and in the question	Factual Knowledge	Recall, Retrieve, Recognize
2. Paraphrasing or retrieving concepts and extracting information from the format of question (verbal, graphical, symbolic, equation mathematical) by changing to proper representation	Factual Knowledge, Procedural Knowledge of working with different representations	Understand /Interpret
3. Associating meaning to the concepts	Conceptual Schema	Recognize, Classify, Categorize, Compare and Contrast
4. Finding interrelation between the facts	Conceptual Schema	Compare
5. Generalizing the interrelated concepts to the categories	Knowledge of Classification, Categories	Classify, Compare,
6. Relating the generalized category to a relevant principal or theory	Knowledge of Theories, Knowledge of Principals	Classify, Compare,
7. Explaining the steps of applying a principle or theory, or skills for applying a method	Procedural Knowledge	Recall, Explain, Interpret
8. Finding the relation between cause and effect and generate a hypothesis, exploring connections between the knowledge and the features of a new context	Conceptual Schema, Knowledge of Classification, Theories and Principles	Infer, Compare Apply

Table F-2 Characteristics of seven levels of responses

Different types of data: levels of possible answers to content questions	Type of the Knowledge employed (Bloom's revised taxonomy)	Type of the possible cognitive process employed (Bloom's revised taxonomy)	Conceptual link structure possibilities
1. A set of disparate, disconnected and contradictory facts or including some relevant facts interpreted and discerned form the given representation	Factual knowledge	Recall	D,H, or T
2. All the facts that required are present but not showing what they imply, where they connect and why things happened	Factual knowledge Conceptual Schema	Recall and Recognize	D,H or T
3. Facts are partially or entirely retrieved but the through connections and relations of the facts are not provided,	Factual Knowledge	Recall	D-D, H-H,T-T, T-H, D-H, T-D, H-H-T,H-T-T, D-D-D, D-D-T,D-D-H, D-H-H,D-H-T,D-T-T
	Conceptual Schema	Understand/Compare and contrast	
4. Subtle interconnection of concepts that investigated cause and effects with plausible inferences and coherent explanation	Conceptual Schema	Compare and Contrast Infer	D-H-H,D-H-T,D-T-T, H-H-T, H-T-T

5. Facts and concepts are linked with larger theories and principals, but why and how the phenomena works is not thoroughly justified, account may shows difficulties and errors using procedures and rules	Factual, conceptual schema, Knowledge of theories and principles, Procedural Knowledge	Compare and Contrast, Infer, Classify,	D-H-H,D-H-T,D-T-T, H-H-T, H-T-T
6. Subject specific skills rules, principles are implemented, concepts are linked with larger theories and principles, the connections of the theories and new context is explored and why and how the phenomena works is thoroughly justified	Factual, conceptual schema, Knowledge of theories and principles Knowledge of theories and principles Procedural Knowledge	Recall Compare and Contrast, Infer, Classify Apply	H-H-T, D-T-H, T-T-H

Appendix G - Written Extended Exam Questions

Earth Sciences

On a large scale, water moves around the world's oceans in a distinctive pattern. This flow of water is sometimes referred to as the Global Ocean Conveyor because one can make an analogy to the motion of a rotating conveyor belt. Surface waters, shown in red above, circulate through the Pacific to the Indian Ocean and finally to the North Atlantic. Deep ocean waters, shown in blue, flow from the North Atlantic back through the Indian Ocean or to the Pacific.

Consider the following facts:

- Cold water is denser than warm water
- Salt water is denser than less salty water
- At high latitudes, water gets colder as it loses heat to the atmosphere
- At low latitudes, water gets warmer as it is passes through low latitudes because the rate of evaporation is high.
- Surface waters are primarily moved by the wind.

Questions:

- Why do you think surface waters descend to great depths in the North Atlantic near Greenland?
- How might the temperature, salinity and density of deep waters between Brazil and West Africa differ from surface waters in the same region?

- To the best of your ability, describe in a step-by-step fashion what drives the motion of the Global Ocean Conveyor.
- Some climate scientists have expressed concern that warming temperatures will melt ice in the northern hemisphere introducing large amounts of fresh water to the surface waters of the North Atlantic. How might this impact the Global Ocean Conveyor belt?

Biology

Your research company, New Age Human Biology, has just received a million dollars from the National Science Foundation (NSF)!!! Due to your extensive experience with other organ systems, the NSF gives the money to your lab to investigate the organ of your choice. You decide to focus on the urinary bladder to learn more (especially since infections of the urinary system are the second most common cause of infection in the human body!!!). How would you proceed with this investigation and what are some of the key questions you would ask?

Earth Sciences

Recently an extra solar planet was detected. The planet was most similar to the earth yet discovered. A (hypothetical – not actually done) probe on the surface measures the wind direction from east to west during the planet's day and west to east during the planet's night. Using your knowledge of earth's winds, "describe and explain" what surface features you might expect to find near the probe?

Physics

Now imagine that rocks with parachutes are thrown off a cliff and fall as shown in the following picture. The rock on the left is twice as heavy as the rock on the right. Both of the rocks are at terminal velocity.

a) Knowing that these two rocks are falling at terminal velocity, draw in the force of air resistance acting on both of the parachutes.

b) Are each of these rocks in equilibrium? How can you tell?

c) Using the notion of air resistance, construct an explanation for why the larger rock will hit the ground first.

Physics

A few weeks ago, a new proton accelerator began testing in Geneva, Switzerland. As the name implies, the device applies a force to protons and causes them to accelerate to very high speeds. Protons have a positive electrical charge, so electrical forces are involved in the process. In this question, ignore any forces on the proton, like gravity, that are not caused by electrical interactions.

a) One way to apply a force to a proton is to put it between two large plates as shown below. Charges are then placed on the plates. Show what charges should be on the plates to accelerate the proton to the left. Explain your answer.

b) One possible motion for the proton would be to go through the small hole in the left plate. Describe the motion of the proton after it goes through the hole. Use at least one of Newton's Laws to explain this motion.

c) As a proton moves very fast its mass increases. When the mass is greater will the force applied need to be greater to obtain the same acceleration? Use at least one of Newton's Laws to explain your answer.

Chemistry

If you live in a place that has a lot of snow and ice in the winter, then you have probably seen the highway department spreading salt on the road to melt the ice.

- Explain how the chemical structure of salt affects the properties of the solution.
- Describe why sugar, another common household item, does have the same effect as salt when it is placed on ice.
- Compare what happens to the chemical structure of salt when combined with water to the chemical structure of sugar when combined with water.

Astronomy

In the middle of the night, a student notices a quarter moon rising due east. Remember earth rotates counterclockwise.

- Is this the first quarter or third quarter of the moon?
- Explain how you can tell. Your explanation may include a diagram.

Appendix H - Rubric

Factual Knowledge

For factual knowledge we look for basic elements, single entities and separate pieces of information if they are mentioned verbally or symbolically or it can be inferred from other statements

NE= Negligible evidence for having access to basic premises and discrete entities that are required for construction of conceptual schema

E= Showing some evidence of having access to the basic facts and discrete entities that are required for construction of conceptual schema, including cases that lacks some facts in other words account that is identified by mixture of relevant and irrelevant facts

Conceptual Schema

Conceptual schema is one of the subcategories of conceptual knowledge in Anderson's table. For this type of knowledge, we look if the appropriate concepts are employed, the definitions of concepts are clear and correct meanings are attributed to the concepts. Moreover, we evaluate the understanding of relations between the concepts and concept links and we look whether special attributes, specific features are associated with appropriate categories or classes

NE= Negligible evidence of knowing the meaning of the concepts or knowing the relations between the concepts, employing wrong concepts and attributing wrong meaning to the concepts without understanding of the relations between them, or introducing the concepts by recall without showing their meanings

E= Showing some evidence of having access to relevant concepts, understanding the meaning of the concepts in relation with other concepts, constructing partially appropriate schemas in which student clarifies the relationship between the concepts or account that includes relevant concepts that are mixed with some irrelevant concepts

Concept Level Links

Based on Nieswandth and Bellemo's multi stages analysis, we draw a symbolic representation of concept links by categorizing the level of the concepts students employed in their answers in terms of descriptive, hypothetical and theoretical. We can show the links with abbreviations T (for theoretical), H (for hypothetical) and D (for descriptive) and show the links students employed as H-H, H-D-T and so on.

d) Descriptive concepts: Concepts that can be inferred or observed with direct senses e.g. magnets, organisms, food chain

e) Hypothetical concepts: the concepts that cannot be observed directly but indirectly with employing a model or if the observational time period were extended e.g. magnetic fields or fossils

f) Theoretical concepts: The concepts that cannot be observed and the meanings come from the theories which ideas originate e.g. atoms and genes.

Procedural Knowledge

For this knowledge, we look if the students have the skills, knowing the techniques and algorithms or knowing rules and steps of applying a principle. Procedural knowledge manifests in different faces, such as algorithm, trigonometry, geometry, physics formula, vectors and so on. In addition, procedural knowledge is attributed to the knowledge of performing a chemical reaction, knowledge of knowing how to find probabilities or using combination rules of probabilities, knowledge of the prescribed steps of solving particular problems, or knowledge of steps to execute a process, or series of steps that are required for verifying a principal

NE= Negligible awareness of subject specific skills and techniques to implement the procedures or rules

E= Showing some evidence of being skillful or having some knowledge in using subject specific skills and techniques and knowing the series of steps to execute the procedures,

<u>Compare and Contrast</u>

We assess if the response reflect the correspondence between elements and entities of the problem

NE= Comparing by recall, comparing irrelevant aspects which can't establish a reasonable connection between cause and effect or showing concrete comparison with comparing the apparent features of subject

E= Going beyond superficial aspects and comparing more in-depth features inferred from given information. Comparing those aspects and features that are fundamental for justifying cause and effect changes, or comparing variables that provide plausible evidence for why and how and what changes occurred, also including cases that some meaningful comparisons exists but there is a lack of compared entities for a plausible connection for what and why and how things happened.

Understand (Infer)

We assess if the answer recognize the patterns between series of the events and instances

NE= A nonsense conclusion including fragmentary segments, fail to relate assumptions and conclusion, or the links between assumptions and conclusions are either by recall or has been constructed concretely

E= Recognizing a pattern and finding a reasonable and plausible connection and developing some insightful relations between cause and effect with some evidence of plausible relationship between what and why and how

Explain

We look for a cohesive and convergent argument leads to reasonable conclusion

NE= More descriptive and superficial or borrowed idea; providing an answer without supporting, based on personal assumptions or concrete idea, non cohesive and a fragmentary and sketchy argument

E= Showing some evidence of well supported by argument that shows explicit thought, subtle connections between assumptions, theories and types of knowledge required in the problem, also showing some justification and good sketch of "What" to "How" and "Why, judged from comparing specific and in-depth features of the subject. An argument that has segments supported by another and cohesively leads to a reasonable conclusion, including cases that showed incomplete internalizations or contradictory statements, that coexisted with meaningful connections.

Apply

We assess if students can recognize the information, the relevant concepts, principles and the relations between the facts, concepts and principles in the new context.

NE=Association of facts, concepts, procedures are not explored in the context of question's scenario.

E= Association of facts, concepts, procedures and features of questions' new context are partially reconstructed or association of facts, concepts, procedures reconstructed in connection with the features of question scenario to present a plausible answer.