Assessing Students' Reasoning across Disciplines in Entry-Level Science Courses

Mojgan Matloob Haghanikar, Sytil Murphy, and Dean Zollman

Kansas State University

Abstract

As a part of a study of the science preparation of elementary school teachers, we studied students' reasoning skills in the context of applying particular scientific concepts. We have devised open ended content questions which apply recently learned concepts in a new context. This requires that students recognize and generalize the relevant facts or concepts and their interrelationships to suggest an applicable or plausible explanation. To evaluate students' answers, we developed a rubric based on Bloom's taxonomy as revised and expanded by Anderson (Anderson et al., 2001). To classify the quality of the students' performance, we described distinguishing characteristics of the responses as indicators of *Poor*, *Developed* and *In-depth* performance levels for each type of knowledge and cognitive processing. This method fulfils our primary objective of constructing a method for comparing students' reasoning across different disciplines. In this paper, we will present an example of a content question and the method of analysis for this case.

Background and Introduction

The National Study of Education in Undergraduate Science (NSEUS)[1] is studying the effect of active engagement in undergraduate science courses taken by future elementary teachers on their teaching of science. Comparing students’ learning of content knowledge in science courses from interactive engagement teaching-learning strategies to those with traditional teaching methods is a part of this project.

On most campuses, which have science courses specifically for elementary education majors, a traditional course and an interactive engagement course at the same level and in the same subject area often do not exist. And, even if such a pair of courses exists at one institution, between institutions, we gain an even greater variety of subject areas. The variety of the science courses and question contexts in NSEUS include; physics, chemistry, biology and geology. Accordingly, the direct comparison on subject matter learning is impossible and we are concentrating on comparisons of reasoning skills within the content that the students have learned. Because of the size of the study, we cannot interview the students and must rely on written responses to exam questions. Therefore, we wrote questions designed to elicit reasoning skills and developed a rubric for comparing the reasoning patterns in the students’ written responses.

Research Design and Literature Review

This study puts together the findings of cognitive science, previous research, educational goals and requirements of National Science Education Standard (NRC, 1996) to classify students’ reasoning abilities. Participants were pre-service elementary teachers who were given

[1] www.nseus.org

open ended questions on their final exam. The whole approach toward developing the assessment protocol resembles 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 responses and planning the question design accordingly.

Stage one

At the first stage we used National Science Education Standards (NRC, 1996) and previous research to identify the desirable priorities for the assessment of inquiry. Courses that claim to have implemented inquiry should maintain the level of the NRC standards in their instructional activities and should provide opportunities for students to develop conceptual understanding and develop their procedural skills. It is not sufficient for students to only recall information. Instead, they should engage in higher levels of thinking such as, classifying, summarizing, inferring, comparing, explaining and applying their prior knowledge to a new context.

According to National Science Education Standards definition of a well reasoned response "demonstrates reasoning characterized by successive statements that follow one another logically without gaps from statement to statement." In other words, students' responses should represent a complete chain of "What", "Why" and "How".

However, judging correctness, evaluating the use of controlling variables, or measuring students' conceptual knowledge may not effectively assess the students' gains due to inquiry (Russ et al., 2008). Instead, Russ et al., draw attention toward the construction of cause and effect relationships that explain 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.

In agreement with researchers of schemata theory (Mayer, 2002), the level of understanding 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, reasoning can be defined in terms of the thought processes and knowledge types that students bring to a new context. There is no benefit to gaining knowledge and then not using it anywhere else. Rebello et al. (2005) and several other researchers (McKeough, Lupart, & Marini, 1995) defined transfer as the ultimate goal of learning. Dufresne et al. (2005) viewed transfer as a constructive process where the activation and application of knowledge pieces highly depends on the context features and user's prior knowledge. As knowledge pieces are brought together, new knowledge is created through mental processes such as association, classification, combination and refinement (Dufresne et al., 2005).

Given this, we need to design an assessment that values the qualities of inquiry-based learning. The students' performance were judged in terms of the sophistication of the cause and effect relationships, students' prior knowledge and the application of this knowledge to a new situation in their written responses to exam questions.

Stage two

In the second stage, we characterized the indications and evidences for understanding. As such, when we evaluate responses, we typically look for qualities and attributes that reflect the organization of knowledge pieces or the implementation of thought processes in the students' responses. Based on a review of knowledge construction (Mayer, 2002; Bransford et al.,1999), Anderson et al. (2001) expanded on Bloom's taxonomy (Bloom, 1956), including adding another dimension to it. Anderson and colleagues classified the dynamic process of both knowledge construction and cognitive processes and then organized the thought processes and knowledge types into a hierarchical two dimensional taxonomy. The complexity of the cognitive dimension increases from "Remember" to "Apply." The hierarchies of knowledge proceed from the lowest level of factual knowledge toward more complex and abstract levels of conceptual, procedural and meta cognitive knowledge. Sometimes conceptual knowledge develops out of procedural proficiency and vice versa. Thus, procedural knowledge may not be more abstract than conceptual knowledge in all cases. To provide a better visualization of the objectives and organization of our classification scheme, we organized our thoughts according to the thought processes that we selected from the revised Bloom's taxonomy (Anderson et al., 2001).

Stage three

In an effective assessment there is an alignment between the goals and the assessment design. In the first and second stages, we suggested a definition for understanding and what indications serve as evidence of understanding. Sometimes, rote learners, use memorization to exhibit thorough and complete answers. Yet, if we go beyond rote questioning, and design content questions that require application of newly learnt concepts in a new context, the rote learners would no longer be able to answer. Our goal is to assess how students proceed from the initial step of knowledge construction toward applying that knowledge to a new context. As such, we need to ask questions that leads students through the dynamic process of knowledge construction.

Methodology

The constructive process of transfer depends on many prerequisite steps as specified in Anderson’s taxonomy. Wiggins and McTighe (2006) distinguished between analytic-trait scoring and holistic scoring. Analytic-trait scoring is a type of scoring that evaluates students’ achievement with several distinct criteria. As a result, the performance is examined several times through the lenses of separate criteria. In contrastin holistic scoring, the assessors report their overall impression about a performance.

In effect, our data analysis approach is an analytic-trait type of scoring with the traits we selected from Anderson’s taxonomy as shown in Table 1. Inspired by the revised Bloom’s taxonomy, we developed a two dimensional framework (Table 1) with each large color-coded group of 9 cells displaying the intersection of rows and columns that belong to a certain type of knowledge and cognitive process. In other words, the color coded regions represent the types of knowledge and cognitive processes that students brought to the situation. For example, if students recall only facts, their level of reasoning is placed in the pink region at the top left of the taxonomy. However, if there is evidence of the application of the facts relevant to the features of the new situation, then the level of reasoning is indicated by the green region on the right corner.

Based on Anderson’s revision of Bloom's taxonomy, every type of knowledge and cognitive processes is classified in terms of more specific subtypes. For instance, ‘Conceptual Knowledge’ has four subtypes: the knowledge of interrelationships between facts (*conceptual schema*), ‘*Knowledge of Classifications*’, ‘*Knowledge of Principles’* and ‘*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 need to be more specific about the subcategories we choose to incorporate in the question.

Given the restrictions of the research, it was impossible to design questions that required many types of knowledge structures and cognitive processes. Our content questions were placed on the final exams with only 10 to 15 minutes allowed for answering each question. In addition, to compare reasoning 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. In our question design, from the category of 'Conceptual Knowledge' we mostly considered the subcategory of '*conceptual schema'* and for the cognitive process of 'Understand' we considered '*Infer'*, '*Compare'* and '*Explain*'. For the category of 'Apply', a higher level of cognitive processing, we selected the subcategory of '*Implement'* which applies knowledge types to an unfamiliar task. In sum, we selected seven traits from Anderson's taxonomy for our analytical-trait scoring technique (Wiggins and McTighe, 2006). From the knowledge dimension we selected factual knowledge, conceptual schema and procedural knowledge while from the cognitive dimension we chose compare, infer, explain and apply.

Rubric

After identifying the traits for our analytical-trait scoring (Wiggins and McTighe, 2006), we classified students' responses in terms of the quality of the traits that they exhibited in their responses. We described distinguishing characteristics that are indicators of *Poor, Developed* and *In-depth* levels of accomplishments for each type of knowledge and cognitive process. Fundamentally, we followed a procedure similar to that of Wiggins and McTighe (1998), who defined six traits for understanding (*Explanation, Interpretation, Application, Perspective,* *Empathy, and Self-Knowledge*) and developed an analytic scoring rubric to rate the performance level each trait *(Sophisticated, In-depth, Developed, Intuitive, and Naive*).

To provide a better visualization of classified levels of the traits, we modified Bloom's revised taxonomy and divided the regions into sub regions (Table 1). Within each region, we formed a 3x3 matrix to track students' level of reasoning across knowledge types in relation to cognitive processing levels. The rows of each 3x3 matrix represent the three levels of accomplishment we defined (*Poor*, *Developed*, and *In-depth*) for the knowledge types while the columns show the same for each type of cognitive dimension. Therefore, the embedded cells inside the matrices not only show relevant knowledge and cognitive type, but also the level of accomplishment. For example, in Table 1, the cell marked with the number 1 indicates a comparison of some aspects of concepts while the cell marked with the number 2 shows that the student has demonstrated procedural knowledge in the context of new situation.

The distinguishing characteristics that are indicators of the three levels of accomplishment for each trait are provided in the rubric (Appendix 1). It was not convenient to embed all the descriptions and definitions in the original taxonomy (Table 1) and we needed to display the rubric as a content list.

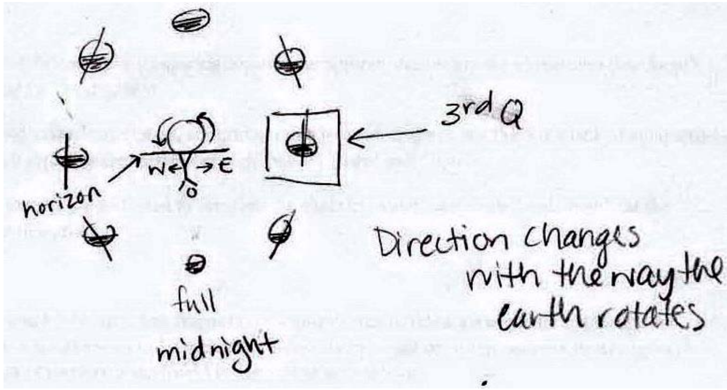
Analysis

The following question was given to a traditionally taught astronomy course for non-scientists majors at a small Midwestern university. There were 78 students who completed this question at their final exam.

As mentioned earlier, our goal is to assess how students proceed from the initial step of knowledge construction toward applying that knowledge to a new context. We were supposed to evaluate the students' performance in the context of moon phases. Consequently, we envisioned a question requiring students to bring their knowledge of facts, concepts and procedural skills to a new context and through the cognitive processes of compare , infer, Explain and apply predict the outcome. The question stated below, taken from Physics by Inquiry (McDermott et al., 1996), meets our requirements. The question statement is followed by three example responses showing a progression in student reasoning. For each example, we discuss how we used our rubric (Appendix 1) to judge the distinguishing features of the responses. We evaluated the responses in terms of the evidences that occurred in the responses. For our purposes it is not a question of whether the student actually has access to that type of knowledge or skill but whether he/she displayed it in the response. The abbreviations "P," "D" and "I" stand for *Poor, Developed,* and *In-depth* levels of performance.

Question: 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.

Example 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. Direction changes with the way earth rotates.*



- <u>Factual Knowledge: I</u>

Student pointed to various moon stages, earth's spin and moon cycle are basic facts for constructing other types of knowledge.

- <u>Conceptual Schema: I</u>

The response represents a mental model consisting of several interconnected concepts such as the changing sunlit portions in a lunar cycle; midnight is located opposite to the sun, and the moon's rise and set in relation to the earth's spin.

- <u>Procedural Knowledge I</u>

There is evidence for using geometrical procedures including; the configuration of the sun, earth and moon during a moon cycle, finding east and west with east being in the direction of rotation and west being directly opposite.

- <u>Compare and Contrast: I</u>

This student fully compared the relative alignments of the moon, sun and earth and how that affects the moon's sunlit portion. Moreover, there is an indication that the student is relating the change of direction to the way the earth rotates

- <u>Infer: I</u>

There is a plausible connection between the location of the observer at midnight, the rising moon, looking at the east and seeing the 3rd quarter.

- Explain I

The explanation cohesively described what is happening and why and how the 3rd quarter happened to be due east of the observer.

- Apply: I

The student constructed the moon phase's model and the location of the observer according to the information given in the problem. Both procedural and conceptual knowledge are associated with features of the new context

Example 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 have been a third quarter moon, because the 1st quarter moon is setting.*

- Factual Knowledge: D

Besides stating that two moon stages appear highest at a certain time, no other facts justify why two moon phases appear highest overhead at certain times.

- Conceptual Schema: D

The response reflects a few interconnected aspects including the sequential order of the moon phases.

- Procedural Knowledge: P

There is no evidence for the geometrical configuration of the sun, moon, and earth.

- <u>Compare and contrast: D</u>

The student compares some of the properties of the 1st and 3rd quarters and infers the observer should have seen first quarter, however, the student did not justify why and how he attributed high overhead locations of the moon to certain times.

- <u>Infer: D</u>

The student's inference according to the sequential pattern of lunar events is acceptable; however, there is no justification for why moon phases appear high in the sky at specific times.

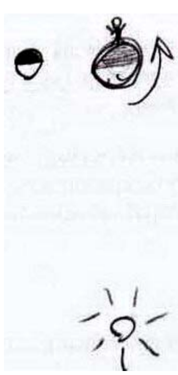
- <u>Explain: D</u>

The statement of relating moon phases to their positions in the sky at a specific time is not an indication of understanding and it shows no more than recall.

- <u>Apply: D</u>

In the model the student used the timing of moon phases and midnight as clues to guess the answer. The student did not show concepts and procedures that justify the why the 3rd appears at the west of observer.

Example 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>Factual Knowledge: D</u>

A few facts are exhibited to start the discussion such as 1st and 3rd quarters and their sunlit portions. Yet other facts are required to justify the predicted type of moon phase third quarter appeared on the left.

- Conceptual Schema: D

The sunlit portions are clearly facing the sun and the location of the observer at midnight is located opposite to the sun. In addition, the student associated the moonrise/set times to the different moon phases.

- Compare and contrast: D

The student compares the times of moonrise for different quarters but does not compare enough aspects to explain how and why they are different. However, the student did not justify why and how he associated the first and third quarters to certain times.

- Procedural Knowledge P

Negligible evidence exists to indicate understanding of the geometrical configuration of moon phases in relation to the earth and sun.

- Infer: P

The response includes few informative statements about moon phases appearing at certain times, and there is no indication of any event that in effect caused another event.

- Explain: P

There is no justification of why and how things happened.

- Apply: P

No association of concepts, procedures and context features occurred.

Findings

After coding the answers for each trait, we counted the frequency distribution of the three levels of *Poor*, *Developed* and *In-depth* for each of the seven traits. Having higher frequencies at the in-depth level is an indicator of better performance for that trait. The seven figures below display the histograms that are associated with the above-mentioned sample question. In the first figure, we can see that the histogram is skewed to the right, which actually implies most of the students have access to the factual knowledge. However, the second figure shows more flatten distribution of the relevant categories. The smaller percentage of the in-depth level for the conceptual schema shows the significant number of students who displayed factual knowledge, did not associate those facts to exhibit conceptual schema. By comparing Figure 2 and Figure 3 we can deduce most of the students who exhibited conceptual schema, used diagrams to show the relative alignment of the sun, earth and moon, yet, there were students who exhibited a developed level of conceptual schema without using procedural skills. Mostly, in these types of responses, the moon phases are associated with their positions in the sky at a specific time which is not an indication of an in-depth understanding.

The rest of the figures belong to the traits of cognitive dimension. The in-depth percentages due to the traits of cognitive dimension show a dramatic decrease compared to the counterparts in the first three figures. The significant decrease implies that large number of student, who displayed knowledge of facts or concepts, did not further created new knowledge through comparing, inferring, explaining and connecting the prior knowledge to the features of the question.

The inter-rater reliability for this application of the rubric was 72%. Further investigation revealed there are some resolvable discrepancies that can raise the reliability to ~82%.

Summary

In this paper, we present a method for analyzing student reasoning based upon responses to final exam questions. Three sample responses show hierarchies in sophistication of student reasoning to a single question. The distinction is in terms of students' abilities to present a plausible argument rather than the correctness of their canonical knowledge. In other words, we defined the progression of the thought processes in the line between recalling and applying knowledge and analyzed students responses as they proceed through this line.

Since the characteristics determining a well-reasoned answer and the categories used to address those characteristics are independent of the question's context, we can follow this same method of analysis to classify the sophistication of reasoning in different contexts and further compare students reasoning across disciplines.

References

Anderson, L.W. and Krathwohl, D.R. (2001). *Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, Longman, New York.

Bloom, B.S. (ed.) (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals*, Handbook I: Cognitive Domain, New York: McKay.

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

Dufresne, R., et al. (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.

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

McDermott L.C., *et al.* (1996). *Physics by Inquiry*, John Wiley & Sons, New York

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

National Research Council (1996). *National Science Education Standard,* Washington, DC, National Academy Press.

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

Rebello, N. S., Zollman, D. A., Allbaugh, A. R., Engelhardt, P. V., Gray, K. E., Hrepic, Z., et al. (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.

Wiggins, G and McTighe, J. (1998). *Understanding by design,* Association for Supervision and Curriculum Development, Virginia.

Wiggins, G and McTighe, J. 2nd ed(2006). *Understanding by design,* Pearson Education, Inc., Upper Saddle River, New Jersey.

Table 1- Modified version of 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					
			1			
Procedural Knowledge						
						2

Figure 1-Distribution of students' performance on factual knowledge

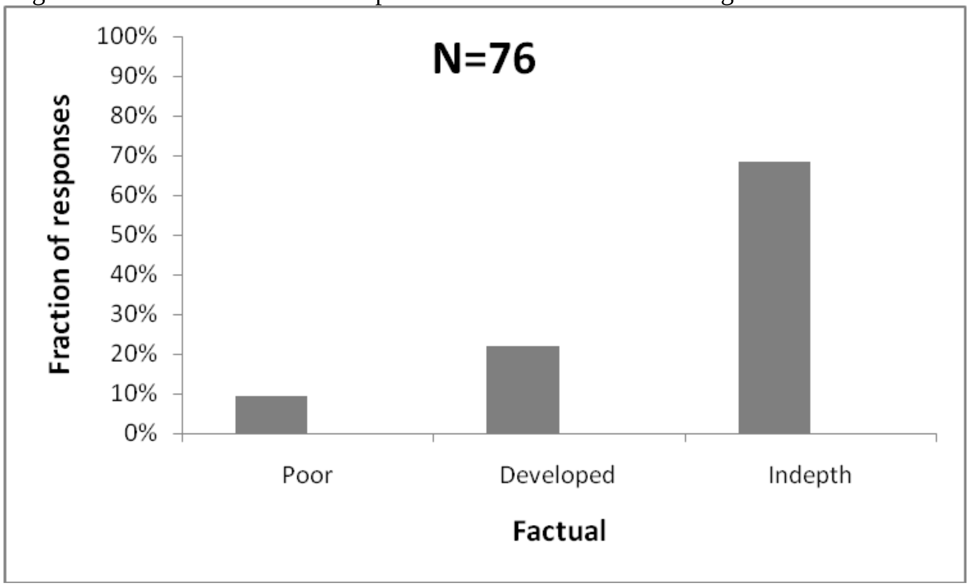


Figure 2- Distribution of students' performance on conceptual schema

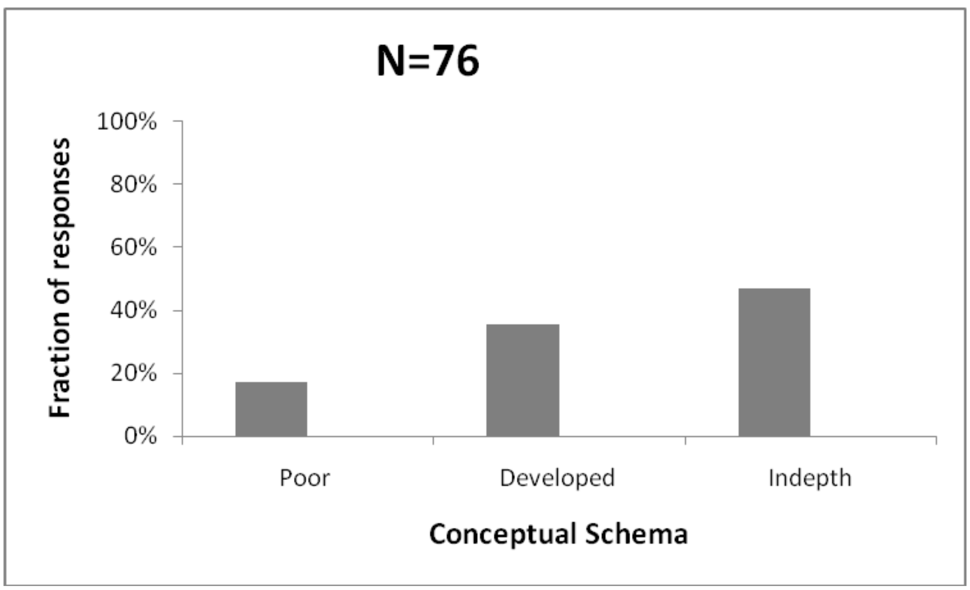


Figure 3 -Distribution of students' performance on Procedural knowledge

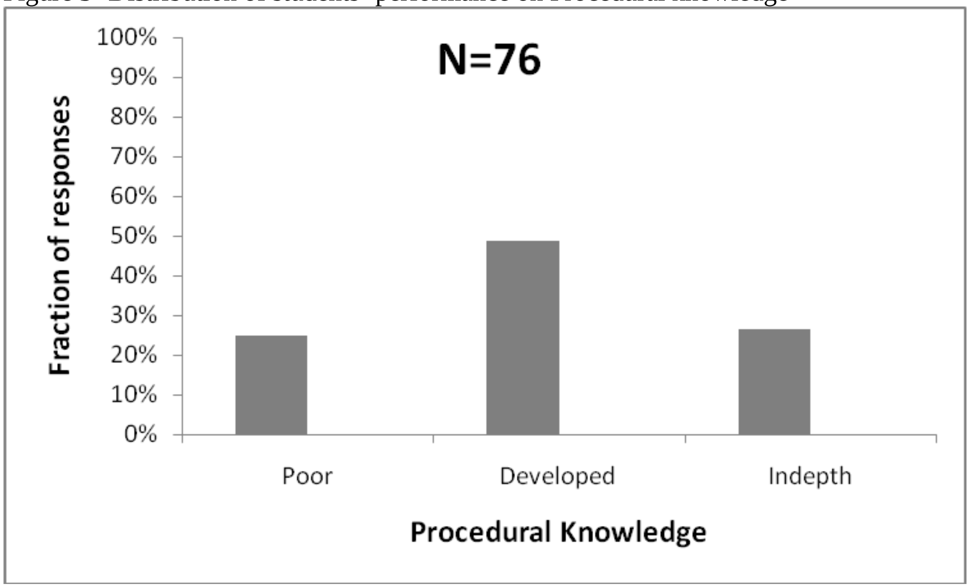


Figure 4 -Distribution of students' performance on cognitive process of "Compare"

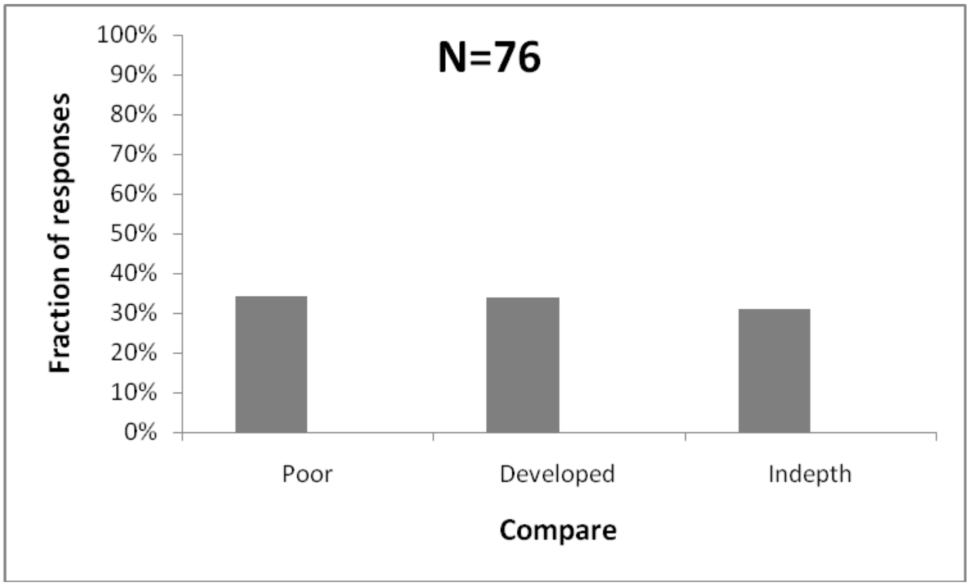


Figure 5- Distribution of students' performance on cognitive process of "Infer"

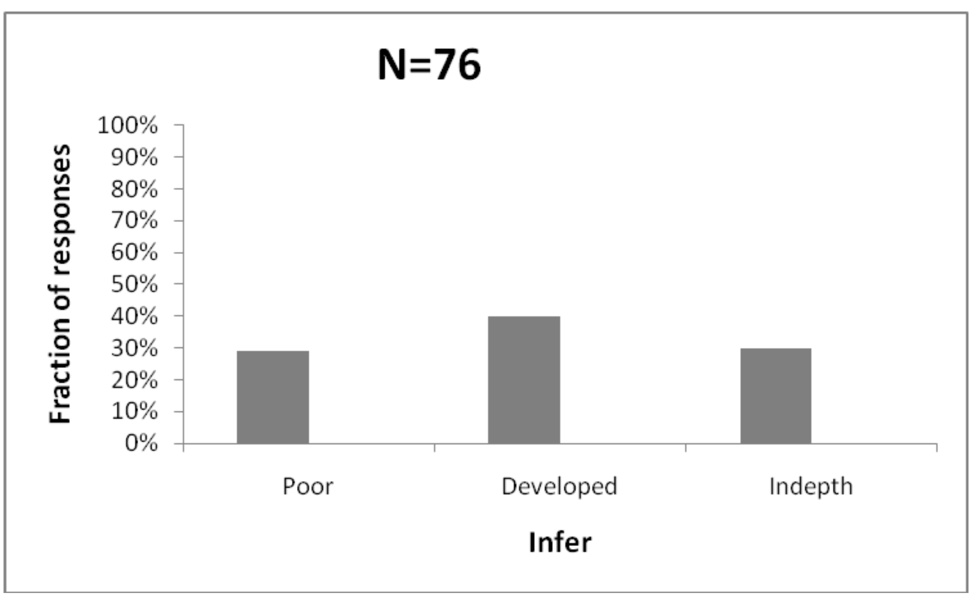


Figure 6- Distribution of students' performance on cognitive process of "Explain"

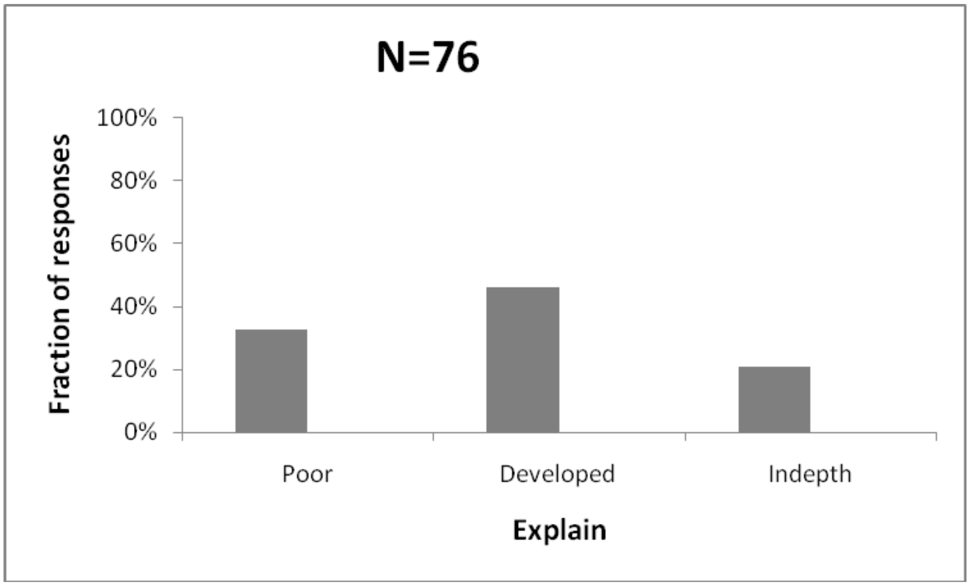
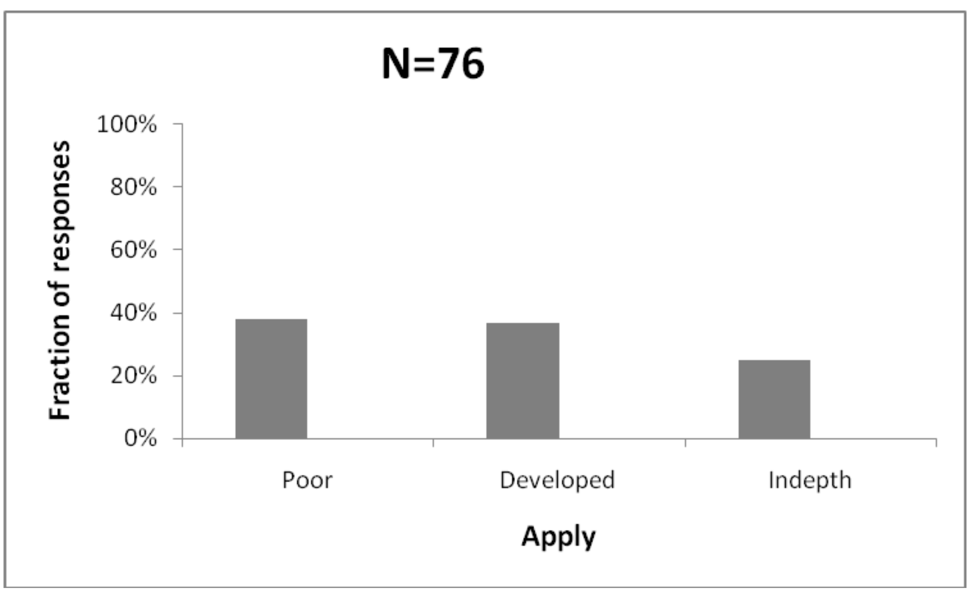


Figure 7- Distribution of students' performance on cognitive process of "Apply"



Appendix 1-Rubric

The abbreviations "P," "D" and "I" stand for *Poor*, *Developed*, and *In-depth* levels of performance.

- Factual knowledge

  P= Unaware of the basic premises and main concepts that one must know to be acquainted with the problem.

  D= Lacking the facts and concepts needed to justify *"what happened"* or a response that includes irrelevant facts and concepts beside relevant concepts.

  I= Showing the basic facts and concepts that one needs to provide a plausible discussion of *"what happened."*

- Conceptual schema

  P= Employing irrelevant concepts, introducing a concept without showing the meaning or attributing a wrong meaning, establishing nonsensical relations between them.

  D= Showing partially relevant concepts, inadequate evidence of demonstrating basic concepts to provide a plausible discussion of *"why"* things happened, general and superficial concrete concepts mixed with specific concepts.

  I= Employing relevant concepts and displaying the meaning of the concepts in relation with other concepts, sufficient concepts to present a plausible discussion of *"why"* things happened.

- Procedural Knowledge

P= Negligible awareness of subject specific skills and techniques to implement the procedures or rules.

D=Showing the knowledge of procedures and rules or subject specific techniques but having difficulties and displaying errors while using them.

I=Skillful in using subject specific skills and techniques and competent to implement the procedures or rules to justify *"how"* things happened.

- <u>Compare and contrast</u>

P= Comparison did not occur or comparing occurred for superficial or irrelevant features of the phenomena.

D=Comparing more in-depth features, but with a lack of compared entities to have a plausible connection between *"what,"* *"why"* and *"how."*

I= Comparing entities are sufficient to complete the chain of *"what,"* *"why"* and *"how"* things happened.

- <u>Infer</u>

P=A nonsensical conclusion including fragmentary segments, links between assumptions and conclusions are either by recall or using a concrete mental model as a result of common experiences where no linkage exists between the cause and effect.

D=Recognizing either correct or incorrect mental models with some insightful and reasonable connections between cause and effects.

I=Recognizing a plausible pattern or mental model with reasonable connections between causes and effects along with a credible deduction.

- <u>Explain</u>

P=Fragmentary and sketchy argument based on only common experiences and concrete assumptions.

D= Unable to connect all types of knowledge to present a connected, cohesive discussion to complete the chain of *“what,” “why”* and *“how*.”

I=A cohesive argument with a plausible link between *“what,” “why”* and *“how*.”

- Apply

P=Association of facts and concepts where procedures are not explored in the context of a question’s scenario.

D=Association of facts, concepts, procedures and features of questions’ new context are partially reconstructed.

I=Association of facts and concepts where procedures are reconstructed in connection with the features of question scenario to present a plausible answer.