Reliability of Using Piaget's Logic of Meanings to Analyze Pre-Service Teachers'

Understanding of Conceptual Problems in Earth Science

Michael Wavering*

University of Arkansas

wavering@uark.edu

Katherine Mangione

Middle Tennessee State University

kmangione @mtsu.edu

Craig McBride

University of Puget Sound

cmcbride@pugetsound.edu

*Corresponding Author. Tel.: 4795754283. wavering@uark.edu. 306 Peabody Hall, University of Arkansas, Fayetteville, AR 72701.

Publication Date: May 29, 2013

# Reliability of Using Piaget's Logic of Meanings to Analyze Pre-Service Teachers' Understanding of Conceptual Problems in Earth Science

Mangione's dissertation (2010) looked at the alternative conceptions in earth science of preservice teachers. Some of the data from this study was a series of eleven transcribed interviews, which explored in depth answers to the twenty questions used in the study. These interviews seemed to be candidates to use the Logic of Meanings described by Piaget and Garcia (1991) to analyze the structure of the arguments used by the interviewees in the dissertation study. A preliminary exploration using the Logic of Meanings was used in a paper by Wavering (2011). This paper proposed the use of the Logic of Meanings for the exact purpose proposed by the current investigation. A more important problem arose from the use of the Logic of Meanings and that was the development of a protocol for analyzing the interviews that would produce high reliability among researchers. This paper describes a study to develop a protocol using the Logic of Meanings to reliably analyze the transcripts of interviews of alternative conceptions in earth science from Mangione's (2010) study.

Inhelder and Piaget (1958) outlined the structures of logical reasoning for concrete and formal operations. The sixteen binary operations were defined in this text (pp. 103-104) (Table 1) and more recently Piaget and Garcia (1991) redefined the sixteen binary operations into a Logic of Meanings. This was the framework used to analyze the interviews that comprised the data for this study. Franklin (1992) developed a two-tiered instrument to identify misconceptions about physics phenomena. Mangione (2010) used the two-tiered structure to design an instrument to identify alternative conceptions of earth science content used by preservice teachers. Author 2 also conducted the eleven interviews used for this study to obtain more in depth information about preservice teachers' alternative conceptions of earth science content. Wavering (2011) demonstrated the use of Piaget's Logic of Meaning (Piaget and Garcia, 1991) to analyze student's reasoning in secondary science classrooms. The current study arose from the need to have a high level of reliability among the researchers in using Piaget's Logic of Meaning to analyze the interviews resulting from Mangione's study. The effort to establish reliability resulted in the Binary Coding Protocol for Analysis of Interviews (Table 2).

## 1. Method

### 1.1 Participants

Six elementary education and five secondary education preservice teachers participated in the study and the interviews. The participants took a written test of twenty questions about earth science alternative conceptions (Mangione, Zellers, & Wavering, 2010). Eleven participants volunteered for interviews to verify their responses and obtain more in depth data.

### 1.2 Procedure

The recorded interviews were transcribed to aid in analyzing the responses using the Sixteen Binary Operations (Table 1) to determine the Logic of Meanings of the participants as they reasoned about earth science misconceptions.

### 1.3 Development of a Scoring Protocol

When the authors started analyzing the transcriptions using the Sixteen Binary Operations, it became clear that there was a low level of inter-rater reliability among the three authors of this paper. Our first attempt yielded approximately 40% agreement, which was considered unacceptable. The goal then became to achieve at least 80% agreement. It became evident that some means of standardized coding needed to be developed. In Piaget's and Inhelder's work (1958), it was not clear how the operations would be used for the purposes which we wanted to use them. In order to use the operations with a high degree of reliability, we created the Binary Coding Protocol for Analysis of Interviews (Table 2). The authors, through phone calls and face-to-face meetings discussed the issues of using Piaget's Binary Operations to score interviews. The next section presents the protocol and examples that resulted from the work of the raters.

## 2.0 Binary Coding Protocol

The Binary Coding Protocol in its entirety is presented in Table 2. This section presents each of the protocol rules with examples from the interviews, which illustrate how the protocol is used.

### 2.1 Protocol

**Rule 1**

First word or phrase in an argument label p, next word or phrase label q.

Example

Question 1: Desmond measures his shadow at noon for several months. He notices as the year progresses his shadow lengthens and then shortens. Desmond realizes he might be able to correlate the season based on the length of his shadow. What season is it in figure B? (longer shadow than compared to figure A) [These questions are from the set of twenty questions used in the interviews with participants.]

Subject 030 Response:

R (Researcher): Okay, winter for picture B. And why?

S (Subject): Because <u>his shadow is longer</u>, (**p**) that means <u>the sun is lower in the horizon</u>. (**q**)

**(p + q)** Binary (3) (Conjunction)

**Rule 2**

Use –p (not p) or –q (not q) for phrases with the word not, not directly precedes word or phrase.

Question 18: Danny, an Eagle Scout of 20 years, claims he can predict the occurrence of a very cold winter. Is this possible?

Subject 030 Response:

S: Uh, like before – I can't really give you an example for cold winters – but before a thunderstorm animals head for shelter. Like you don't see birds flying in a thunderstorm, (**-p**) because they've already taken shelter 'cause the change in barometric pressure is an indication of a coming storm. (**q**) **(-p + q) (8)** (Inverse of Converse Implication)

**Rule 2a**

Code negative words or prefixes as –p and –q, for example: less, non-, dis-, mal-. (See rule 2 for not use)

Question 20: George lives in Phoenix, Arizona. Martha lives in Tampa, Florida. One afternoon both washed their laundry and hung it out to dry. It was a sunny day of 75 degrees Fahrenheit in both places. Whose laundry will dry first?

Subject 030 Response:

S: Phoenix, Arizona is less humid (**-p**) than Tampa, Florida. So the dry air helps increase the rate of evaporation. (**q**) **(-p + q) (8)** Inverse of converse implication

**Rule 3**

Use subscripts ($p_2$, $q_2$, etc.), when subject changes variables during an answer to a question.

Question 13: Emilio finds himself standing on the bank of the Amazon River. His compass indicates that water is flowing west to east. Is his compass malfunctioning?

Subject 030 Response:

S has already responded to compass malfunctioning (the response was coded using **p** and **q**) and now goes further:

Ah, you can actually look at some water levels in lakes and that sort of thing and it…may not be the deepest part of the lake but water coming to ($p_2$) or from it ($-p_2$) so it doesn't necessarily have to do with the elevation. ($-q_2$) **($p_2 + -q_2$) v ($-p_2 + -q_2$) (13)** (Independence of q in relation to p)

Note: v is used to chain binary operations together and means or.

**Rule 4**

For multiple codes using the same p's and q's (i.e., the same variables) for a given response add statements together to get an overall binary operation representing the argument. This represents the complexity of the argument.

Question 8: Stephen attempted to break into the library by cutting through the glass window using information he learned in geology class. He finally gave up and ran away after the alarms were triggered. The university police found scratches on the window and traces of a mineral on the ground by the window. Which of the following minerals could it have been? (A Mohs scale chart listing 10 minerals is provided.)

Subject 015 Response:

S: …I don't think it would be calcite, (-p) because it isn't hard enough to do it [scratch glass] (-q) **(-p + -q) (2)** (Conjunctive negation)

After some encouragement by the researcher to continue, the subject responds.

S: I think that number 10, the hardest one [diamond] can cut glass (q). I think that is what it means. I mean I remember something about glass. So, yeah, I say topaz (p). **(p + q) (3)** (Conjunction)

Summing these two arguments provides the full argument. **(p + q) v (-p + -q) (9)** (Equivalence)

**Rule 5**

In some responses subject will restate or summarize before making an argument. This part of the response is not coded.

Question 3: Bob notices that regardless of what phase the moon seems to be in, he always seems to see the same side of the moon. He wonders if this could possibly be true.

Subject 029 Response:

S: That's true. We never see the far side of the moon. (Additional response to the question was coded.)

**Rule 6**

Sometimes the researcher restates the argument, analyze only what the subject says.

Question 18: Danny, an Eagle Scout of 20 years, claims he can predict the occurrence of a very cold winter. Is this possible?

Subject 030 Response:

R: Alright, so you're saying we can look at patterns, we can look at historical trends. You mentioned local life and how they're preparing. Can you give me some specific examples? (S makes the argument in a response to this.)

**Rule 7**

Make sure to use only p and q, not r or other symbols. This enables the coder to relate to binary codings only.

**Rule 8**

Some phrases imply either/or.

Question 13: Emilio finds himself standing on the bank of the Amazon River. His compass indicates that water is flowing west to east. Is his compass malfunctioning?

Subject 030 Response:

S:  Because the way rivers are from the point of highest elevation to points of lowest elevation (**p**) *irregardless* [either north to south or any other direction] of the compass direction. (**q**) (**-q**)

**(p + q) v (p + -q) (11)** (Independence of p in relation to q)

**Rule 9**

Put answer to the question in the argument when it is stated in the response and used as a part of the argument, code as p or q.

Question 4:  The moon seems to change shape as it orbits the Earth.  To an observer on Earth, what phase would she see if the moon was in position C?

Subject 030 Response:

S:  Full moon [choice 2 on test] **(p)**

R:  Why?

S:  …Ah, like the observer sees the light being reflect off the entire face of the moon that is facing the earth.  (**q**)  The other positions (**-p**) you only see some or none of the light being reflected off the surface of the moon.  (**-q**)  **(p + q) v ((-p + -q) (9)** (Equivalence)

**Rule 10**

Keep it simple, go for the simplest form of coding.  If the coding becomes too complex, apply this principle.

Question 13:  Emilio finds himself standing on the bank of the Amazon River.  His compass indicates that water is flowing west to east.  Is his compass malfunctioning?

Subject 030 Response:

S:  No, his compass is not malfunctioning.

R:  Why not?

S: Because the way rivers are from the point of highest elevation to points of lowest elevation (p) irregardless of the compass direction. (q) (-q) **(p + q) v (p + -q) (11)** (Independence of p in relation to q)

R: Are there any of factors involved in how a river or water flow?

S: There is also the concept of path of least resistance which is in addition from the highest to lowest.

R: Which one do you think has more of an impact?

S: Probably the path of least resistance has a little bit more.

R: Why do you think that?

S: A, you can actually look at some water levels in lakes and that sort of thing and it may not be the deepest part of the lake but water could be coming to ($p_2$) or from it ($-p_2$) so it doesn't necessarily have to do with elevation. ($-q_2$) **($p_2$ + $-q_2$) v ($-p_2$ + $-q_2$) (14)** (Inverse of independence of q in relation to p)

**Rule 11**

When determining reliability between two raters, use a score of one for complete agreement of the binary operations for a question. Use a half score for partial agreement. For example, if a rater scores an argument (p + q) and another rater scores an argument (p + q) v (-p + -q), the agreement is scored as 0.5.

**Rule 12**

If a response to a question is determined not codable by both raters, this constitutes agreement and is scored 1.0.

2.2 Inter-rater Reliability

Inter-rater reliability was determined using the developed protocol to calculate the percent agreement on two of the interviews. An agreement of 80% or better was the target to demonstrate adequate reliability among the raters. Once 80% or better agreement was reached the development of the protocol was ended. There were three raters. For Subject 029 Rater A and Rater B achieved 82.5% agreement and for Subject 030 Rater A and Rater C achieved 85% agreement. Cohen's Kappa (1960) was calculated for these values with a value of 0.813 for Raters A and B and a value of 0.840 for Raters A and C. The observed probabilities ($P_o$) were 0.825 and 0.850 and the chance probability ($P_c$) was 0.0625, since there is one chance in sixteen that the raters will choose the same binary operation by chance. The values for the Kappa statistic are considered to be almost perfect in terms of strength of agreement (Landis & Koch, 1977). Table 3 summarizes this information.

## 3.0 Conclusion

Excellent inter-rater reliability can be achieved using the Binary Coding Protocol for Analysis of Interviews. While there is a lack of consensus as to what constitutes sufficient inter-rater reliability, most researchers tend to agree that rates above .75 to .80 are considered excellent (Fleiss, 1981; Landis & Koch, 1977). Thus, the inter-rater reliability we achieved is more than good enough to use the Logic of Meanings as a transparent and useful evaluative tool for understanding argumentation about science concepts. It would be of interest to use the Binary Coding Protocol for the Analysis of Interviews with argumentation from content areas other than science. Since the content of the interviews used in this study were earth science misconceptions, it would be interesting to use the protocol to analyze the logical reasoning used in science misconceptions in other science content areas.

References

Cohen, J. (1960). Cofficient of agreement for nominal scales. *Education and Psychological Measurement, 20,* pp. 37-46.

Fliess, J. (1981). *Statistical methods for rates and proportions*. (Second edition). New York: John Wiley & Sons, Inc.

Franklin, B. (1992). *The development, validation, and application of a two-tier diagnostic instrument to detect misconceptions in the areas of force, heat, light, and electricity* (Doctoral dissertation, Louisiana State University and Agricultural and Mechanical College, 1992). Dissertation Abstracts International, 53/12, 4186.

Inhelder, B. & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. NY: Basic Books.

Landis, J. & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, pp. 159-174.

Mangione, K., Zellers, S., & Wavering, M. (2010). *As the world turns: Preservice teachers' alternative conceptions about earth and space science*. A paper presented at the Association for Science for Science Teacher Education, January 14-17, 2010, Sacramento, CA.

Mangione, K. (2010). *The development and validation of a two-tiered multiple choice instrument to identify alternative conceptions in earth science* (Doctoral dissertation, University of Arkansas, Fayetteville, 2010). UMI Dissertation Publishing: UMI 3438056.

Piaget, J. & Garcia, R. (1991). *Toward a logic of meanings*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Wavering, M. (2011). Piaget's logic of meanings: Still relevant today. *School Science and Mathematics 111*(5), pp. 249-252.

## ABSTRACT

A dissertation study looking at preservice teachers' alternative conceptions in earth science was completed by one of the authors. The data used for this study from the dissertation were a series of eleven interviews. **(Purpose)** The authors of this manuscript wanted to provide more in-depth analysis of these interviews, specifically to provide a detailed look at the logical structures used by the interviewees. The means to perform this analysis was Piaget's Logic of Meanings. One of the issues in this data analysis was the reliability of coding reasoning using Logic of Meanings. **(Methodology)** To overcome the reliability problem, the Binary Coding Protocol for Analysis of Interviews was developed. **(Results)** Using this protocol enabled the authors to achieve excellent inter-rater reliability. **(Conclusion)** Use of the protocol enabled the researchers to use Piaget's Logic of Meanings reliably for the analysis of text for the use of reasoning structures. **(Additional data)** (Contains 3 tables)

Table 1 (Adapted from Inhelder and Piaget, 1958, pp. 103-104)

Sixteen Binary Operations

1. Disjunction (*p v q*) = (*p and q*) v (*p and not q*) v (*not p and q*): either or both

2. Inverse, conjunctive negation (*not p and not q*): neither/nor

3. Conjunction (*p and q*): both

4. Incompatibility, inverse of conjunction (*p/ q*) = (*p and not q*) v (*not p and q*) v (*not p and not q*)

5. Implication: *p implies q* = (*p and q*) v (*not p and q*) v (*not p and not q*)

6. Inverse of implication, nonimplication: (*p and not q*)

7. Converse implication: *q implies p* = (*p and q*) v (*p and not q*) v (*not p and not q*)

8. Inverse of Converse implication: (*not p and q*)

9. Equivalence (*p = q*) = (*p and q*) v (*not p and not q*)

10. Inverse of equivalence, reciprocal exclusion: (*p vv q*) = (*p and not q*) v (*not p and q*)

11. Independence of p in relation to q: *p [q]* = (*p and q*) v (*p and not q*)

12. Inverse of independence of p in relation to q (which is also its reciprocal): *not p [q]* = (*not p and q*) v (*not p and not q*)

13. Independence of q in relation to p: *q [p]* = (*p and q*) v (*not p and q*)

14. Inverse of independence of q in relation to p (which is also its reciprocal: *not q [p]* = (*p and not q*) v (*not p and not q*)

15. Complete affirmation or tautology:

(*p * q*) = (*p and q*) v (*p and not q*) v (*not p and q*) v (*not p and not q*)

16. Its inverse, negation or contradiction (0): An assertion that there is an effect and no effect at the same time.

Table 2

Binary Coding Protocol for Analysis of Interviews

1. First word or phrase label p, next phrase label q.
2. Use –p (not p) or –q (not q) for phrases with the word not, not directly precedes word or phrase.

2a. Code negative words or prefixes as –p or –q, for example, less, non-, dis-, -mal, etc.

3. Use subscripts $p_2$, $q_2$, etc., when subject changes variables in a given question during an answer to a question.
4. For multiple codes using the same p's and q's (i.e. the same variables) for a given response add statements together to get an overall binary representing the argument. For example, binary 2 is p and –q, binary 3 is p and q. The combination is (p and q) or (p and –q), which is binary 9. This represents the overall complexity of the argument.
5. In some responses subject will restate or summarize before making an argument. This part of the response is not coded.
6. Sometimes the researcher restates the argument, analyze only what the subject says.
7. Make sure to use only p and q, not r or other symbols
8. Some phrases imply either/or. Example, "I would say no that his compass is <u>not necessarily</u> malfunctioning…" Not necessarily means could be malfunctioning (-p) or functioning (p).
9. Put answer to the question in the argument when it is stated in the response, code as p or q.
10. Keep it simple, use the simplest form of coding. If the coding becomes too complex, apply this principle.

11. When determining reliability between two scorers, use a score of one for complete agreement of the binary(ies). Use a half score for partial agreement. For example, if a rater scores an agreement (p + q) and another rater scores an argument (p + q) v (-p + -q), the agreement is scored as 0.5.

12. If a response to a question is determined uncodable by both raters, this constitutes agreement and is sored. 1.0.

Table 3

Inter-rater Reliability

| Rater | $P_o$ | $P_c$ | Kappa | Strength of Agreement |
|---|---|---|---|---|
| A and B | 0.825 | 0.0625 | 0.813 | Almost perfect |
| A and C | 0.850 | 0.0625 | 0.840 | Almost perfect |

Kappa = $(P_o – P_c) / (1 - P_c)$, where $P_o$ is observed probability and $P_c$ is the probability due to chance.