

2013-01-01

# Investigating the Validity of a Survey Intended to Gauge Mathematics Problem-Solving Disposition along the Impulsive-Analytic Dimension

Miguel Mendoza

University of Texas at El Paso, [suncityguy@gmail.com](mailto:suncityguy@gmail.com)

Follow this and additional works at: [https://digitalcommons.utep.edu/open\\_etd](https://digitalcommons.utep.edu/open_etd)



Part of the [Educational Assessment, Evaluation, and Research Commons](#), [Mathematics Commons](#), and the [Science and Mathematics Education Commons](#)

---

## Recommended Citation

Mendoza, Miguel, "Investigating the Validity of a Survey Intended to Gauge Mathematics Problem-Solving Disposition along the Impulsive-Analytic Dimension" (2013). *Open Access Theses & Dissertations*. 1681.  
[https://digitalcommons.utep.edu/open\\_etd/1681](https://digitalcommons.utep.edu/open_etd/1681)

INVESTIGATING THE VALIDITY OF A SURVEY INTENDED TO GAUGE  
MATHEMATICS PROBLEM-SOLVING DISPOSITION ALONG THE  
IMPULSIVE-ANALYTIC DIMENSION

MIGUEL MENDOZA

Department of Mathematical Sciences

APPROVED:

---

Kien Lim, Ph.D., Chair

---

Osvaldo Morera, Ph.D.

---

Amy Wagler, Ph.D.

---

Benjamin C. Flores, Ph.D.  
Dean of the Graduate School

Copyright ©

by

Miguel Mendoza

2013

## **Dedication**

For my mother, Lucina Uribe Mendoza. Thank you for instilling a love of education in me.

INVESTIGATING THE VALIDITY OF A SURVEY INTENDED TO GAUGE  
MATHEMATICS PROBLEM-SOLVING DISPOSITION ALONG THE  
IMPULSIVE-ANALYTIC DIMENSION

by

MIGUEL MENDOZA, B.A.

THESIS

Presented to the Faculty of the Graduate School of  
The University of Texas at El Paso  
in Partial Fulfillment  
of the Requirements  
for the Degree of

MASTER OF ARTS IN TEACHING

Department of Mathematical Sciences  
THE UNIVERSITY OF TEXAS AT EL PASO

December 2013

## **Abstract**

This study investigates the validity of the Likelihood-to-Act (LtA) survey developed to assess respondents' problem-solving disposition along the impulsive-analytic dimension. Sixteen pre-service teachers were strategically selected from a pool of 495 LtA survey respondents to participate in a semi-structured task-based interview approximately one hour in length. The interview consisted of three main parts: a problem solving section, a consistency section, and a classification section.

- Part 1 of the interview was developed to get interviewees to solve five open-response problems meant to elicit an impulsive response. Interviewees were also asked to verbalize their thought process in order for the researchers to gain some insight into their problem-solving disposition. Four of the five problems were extracted directly from the LtA survey and one additional unrelated problem was added because it cleverly elicits an impulsive response.
- Part 2 of the interview involved having interviewees retake 10 of the LtA items and also had them explain any interview responses that varied from the original response by two or more points.
- In Part 3, each interviewee classified the problem solving act described in each of 10 LtA items (different from those previously seen in Part 2) as being representative of either an impulsive or analytic disposition.

A moderately high correlation of 0.693 ( $p = 0.004$ ) was found between the problem solving data in Part 1 and the original LtA survey scores. In addition, two-scale analysis showed that 60.2% of the corresponding original survey responses matched the problem solving disposition scores assigned by the rubric-scoring team. Reliability issues arose in the consistency section where 63 out of 150 responses to LtA items during the interview were inconsistent with classroom responses to the same LtA items. Survey-taking context issues and other hidden variables surfaced as participants were prompted to explain inconsistent and consistent responses.

## Table of Contents

Abstract.....	v
Table of Contents.....	vi
List of Tables .....	viii
List of Figures.....	ix
Chapter 1: Introduction.....	1
1.1 Background.....	1
1.2 The Likelihood-to-Act Survey Project .....	2
1.3 Research Questions.....	3
1.4 Limitations.....	3
Chapter 2: Theoretical Background.....	5
2.1 Thinking Dispositions and Dual Process Theory .....	5
2.2 Applying Dual Process Theory to Mathematical Problem-Solving .....	7
2.3 Problem-Solving Framework.....	8
2.4 General Likert Survey Creation and Testing.....	11
2.5 Analyzing Interview Data.....	13
Chapter 3: Research Methods.....	15
3.1 Participants .....	15
3.2 Data Collection Instruments .....	15
3.3 Semi-Structured Interview Specifics .....	17
3.4 Analysis Plan .....	22
Chapter 4: Results.....	29
4.1 Interview Problem Solving Results .....	29
4.2 Consistency between Classroom LtA and Interview LtA Responses .....	42
4.3 Classification of Survey Items.....	59
Chapter 5: Discussion.....	63
5.1 Assessing Problem Solving Disposition In An Interview Setting.....	63
5.2 The LtA Survey and Assessment of Impulsive-Analytic Disposition.....	64
5.3 LtA Item Weaknesses .....	66
5.4 Concluding Remarks .....	69

References.....	72
Appendix A - LtA Survey.....	75
Appendix B - Open-Response Problems .....	80
Appendix C - Satisfaction Placard.....	81
Appendix D - Definition Placard .....	81
Appendix E - Interview Questions .....	82
Appendix F - Interview 10-item LtA Survey .....	83
Appendix G – Disposition Code Definitions.....	84
Appendix H – Classification Items, Version A (#1-5) .....	85
Appendix H – Classification Items, Version A (#6-10) .....	86
Appendix I – Classification Items, Version B (#1-5) .....	87
Appendix I – Classification Items, Version B (#6-10).....	88
Vita .....	89

## List of Tables

Table 3.1: Interviewee Distribution by Group.....	18
Table 3.2: The Rubric for Scoring a Problem Solving Episode .....	24
Table 4.1: Instances of Solution Strategies.....	30
Table 4.2: Impulsive-Analytic Strategy Count for Each Interviewee .....	31
Table 4.3: Disposition Codes and Rubric Disposition Scores* .....	35
Table 4.4: The Three Measures of Impulsivity for Each Participant .....	36
Table 4.5: Transformed Rubric Scores .....	39
Table 4.6: LtA Likert Responses That Correspond To Open-Response Problems .....	40
Table 4.7: Transformed LtA Likert Scores.....	40
Table 4.8: Actual Matches and Mismatches.....	41
Table 4.9: 2 x 2 Contingency Table for Actual Disposition Matches and Mismatches .....	42
Table 4.10: 10-item LtA Response Summary .....	43
Table 4.11: Instances of Inconsistent Responses and Explanations .....	44
Table 4.12: Comparison of Classroom Responses with Inconsistent ReDo Responses.....	46
Table 4.13: Inconsistent and Consistent Counts for ReDo Item Responses.....	47
Table 4.14: Reasons for Inconsistent Responses and their Frequency .....	48
Table 4.15: Instances of Item Being Misinterpreted .....	55
Table 4.16: Instances of Misinterpretation from Part 2 of Interview .....	58
Table 4.17: Item Classification, Version A .....	59
Table 4.18: Item Classification, Version B .....	60
Table 4.21: Issues That Arose from Classification Explanations (Version A).....	61
Table 4.22: Issues That Arose from Classification Explanations (Version B).....	61

## List of Figures

Figure 2.1: System 1 and System 2 .....	6
Figure 4.1: A Drawing for Problem 3 .....	32

## Chapter 1: Introduction

Impulsive behavior in everyday life can manifest itself in many forms. For example, some people may be considered impulsive eaters, while others may be considered impulsive shoppers. In general, this type of spontaneous behavior is referred to as *impulsive* because one makes decisions without considering any consequences. Impulsive behavior can be found in many facets of life (Miley & Spinella, 2003), and that includes mathematics learning and problem solving. This study focuses on the assessment of impulsive behavior in mathematical problem solving.

### 1.1 BACKGROUND

One of the fields in which indicators of impulsive behavior have also been studied is in the psychological sciences. Kagan, Rosman, Day, Albert, and Phillips (1964) developed the Matching Familiar Figures Test that classified children as either *impulsive* or *reflective*. Those whose response time was faster than the median time and whose accuracy was below the median accuracy rate were described as impulsive, while those whose response time was slower than the median and whose accuracy rate was above the median were described as reflective. In the context of their study, the more desirable trait of reflectiveness was considered to be the opposite of impulsiveness.

In the context of mathematical problem-solving, Lim (2008b) discussed a similar dichotomy in his use of the terms “*impulsive anticipation*” and “*analytic anticipation*”. Lim defined impulsive anticipation, which is also known as impulsive disposition, as “a proclivity to spontaneously proceed with an action that comes to mind without considering its relevance and without analyzing the problem situation.” Lim, Morera, and Tchoshanov (2009) pointed to the following missing-value problem presented to a group of pre-service teachers by Cramer, Post, and Currier (1993) as an example of an impulsive indicator: *Sue and Julie were running equally fast around a track. Sue started first. When she had run 9 laps, Julie had run 3 laps. When Julie completed 15 laps, how many laps had Sue run?* Thirty-two out of 33 pre-service teachers mistakenly set up a proportion similar to  $9/3 = x/15$  in order to solve the problem. This type of

response is considered an indicator of impulsiveness since you are applying the proportion algorithm without first analyzing the problem situation to determine whether or not the proportion algorithm would be a valid approach.

On the opposite side of the spectrum, analytic disposition in problem solving refers to “one’s proclivity to make sense of a problem situation prior to responding with an action.” Generally speaking, analytic disposition is embodied in many different behaviors, for example, rereading problems to improve comprehension, trying more than one method to solve a problem or to improve efficiency, and considering the reasonableness of an answer. Most mathematics educators would undoubtedly prefer their students to be more analytic than impulsive. Generally speaking, habits of mind that lead to behavior indicative of impulsiveness are considered undesirable, whereas more analytical tendencies would be considered desirable (Lim & Selden, 2009).

## **1.2 THE LIKELIHOOD-TO-ACT SURVEY PROJECT**

This thesis is part of a larger effort. Lim, Morera, and Tchoshanov (2009a) created an instrument, named the Likelihood-to-Act (LtA) survey, for gauging a person’s impulsive-analytic disposition. Such an instrument has the potential for increasing teachers’ awareness of their own problem-solving disposition and inspiring them to help their students develop desirable habits of mind such as more consistently analyzing problems instead of “diving into the first approach that comes to mind” (Watson & Mason, 2007, p. 207) without considering its relevance. The LtA instrument has undergone multiple rounds of pilot-testing and revisions. This thesis serves to inform the continuing development of the LtA survey by providing data on how to accurately assess problem solving disposition and on why some survey items are weaker than others.

This thesis is also a follow-up to a pilot study (Lim & Mendoza, 2010) related to the LtA Survey Project. The purpose of that pilot study was to test and refine the LtA survey by analyzing 92 written responses from participants who were asked to state the initial actions they would take in solving certain

problems that were adapted from the LtA survey items. Relatively strong correlations between participants' disposition scores for their written responses and their LtA survey scores were found. A limitation of the pilot study was that students' reported initial actions did not necessarily reflect their actual problem solving behavior. For this thesis, several participants were videotaped and observed as they solved five open-response problems from start to finish, verbalized their thought process, retook some LtA survey items, and answered interviewer questions related to the survey item validity.

### **1.3 RESEARCH QUESTIONS**

For my research study, I investigated the effectiveness of the LtA items in measuring impulsive-analytic disposition of pre-service teachers in a classroom-setting and sought to identify specific survey weaknesses to help inform future studies of both LtA survey validity and math-based Likert survey validity in general. Consequently, the following questions were explored:

R1) How can problem solving disposition be assessed in an interview setting?

R2) To what extent does the LtA survey assess impulsive disposition in respondents?

R3) What are some of the weaknesses of the LtA survey items?

### **1.4 LIMITATIONS**

The author is aware that the results of this study were limited by the size of the sample being studied. Results from the fifteen interviews did not provide a basis for generalization and should only be taken as empirical evidence that may or may not be representative of other respondents. The author's objective was simply to get an idea of which survey items, or what parts of the items, were effective and/or not effective in gauging a respondent's mathematical problem-solving disposition, and to try to understand why.

Another limitation of this study was that most of the data was self-reported by the interviewees. The accuracy of the results depended on interviewees' openness and willingness to contribute their

thoughts about survey items and mathematical problem solving. Interviewers made every effort to make the interviewees feel comfortable during the interview process.

## **Chapter 2: Theoretical Background**

This literature review consists of five sections: dual-process theory and thinking dispositions, dual-process theory and its application to mathematics education, Schoenfeld's problem-solving framework, an introduction to the creation and testing of general Likert survey items, general test validity, and an overview of the process used to explore, review, and synthesize qualitative data in an interview-centered study.

System 1 and System 2 modes of thinking from Dual Process Theory are analogous to the impulsive-analytic construct being studied. Also, because this research is investigating impulsive-analytic disposition in the context of mathematics problem solving, Schoenfeld's well-established framework for problem solving was used to inform the exploration of the problem solving data. A general discussion of Likert item creation and test validation is included to provide readers a sense of the survey creation process. Finally, the general method used for exploring, memo-ing, and coding the interview data is presented as prescribed by qualitative researchers.

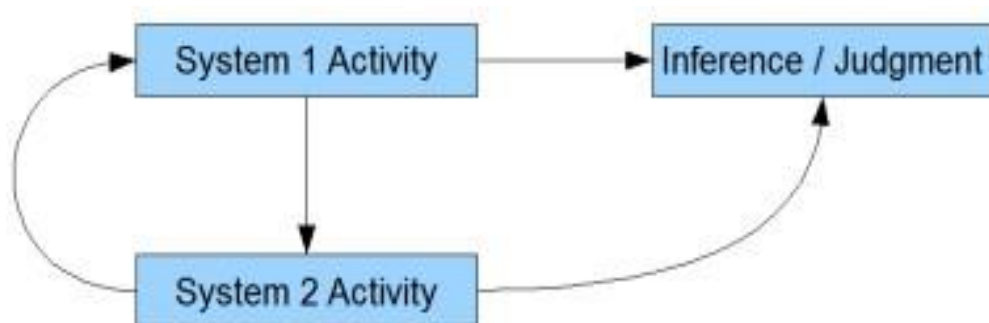
### **2.1 THINKING DISPOSITIONS AND DUAL PROCESS THEORY**

Analytic and impulsive disposition as prescribed by Lim, Morera, and Tchoshanov (2009a) are types of thinking dispositions in the context of mathematical problem solving. From a cognitive science perspective, Stanovich (1999, pp. 156-157) points out that similar concepts under different names have been described by other theorists, such as: "intellectual style (Sternberg, 1988, 1989), cognitive emotions (Scheffler, 1991), habits of mind (Keating, 1990), inferential propensities (Kitcher, 1993), epistemic motivations (Kruglanski, 1990), constructive metareasoning (Moshman, 1994), and cognitive styles (Messick, 1984, 1994)." Stanovich (1999) goes on to state that these thinking dispositions are flexible cognitive styles and should not be confused with more rigid "cognitive capacities" which are explored by instruments such as an IQ test (p. 157). Baron points out that thinking dispositions are more malleable because "you can tell someone to spend more time on problems before she gives up, and if she is so

inclined, she can do what you say” (as cited in Stanovich, 1999). In the case of analytic and impulsive disposition in mathematics problem solving, the hope is that such awareness of problem solving disposition will help survey respondents to begin shifting to an analytic disposition.

A two-system cognitive process analogous to analytic disposition and impulsive disposition is dual-process theory. According to Evans and Frankish (2009), humans have two distinct cognitive systems of reasoning: System 1 is characterized as fast, effortless, automatic, and non-conscious which in our study mirrors the indicators of impulsive disposition (e.g. acting spontaneously without considering relevance); System 2 is characterized as slow, effortful, controlled, and conscious which echoes the indicators for analytic disposition (e.g. taking time to analyze the problem situation and controlling the use of conceptually irrelevant computations). Sloman (1996) points out that the two systems often work cooperatively on some tasks. However, for the most part “humans are cognitive misers” (Stanovich, 2011), and “their basic tendency is to default to System 1 processing mechanisms of low computational expense (p. 29).” Figure 2.1 below is a graphical representation of how System 1 initially activates, and may go unchecked, lead to conceptual change, or inform further System 2 deliberation (Evans, 2006).

Figure 2.1: System 1 and System 2



Providing students with a disposition assessment would allow students to be aware of their problem solving tendencies and help them invoke their System 2 processes more regularly until their System 2 processes more routinely interfere with their automatic System 1 processes.

## **2.2 APPLYING DUAL PROCESS THEORY TO MATHEMATICAL PROBLEM-SOLVING**

The main difference between the two systems is “how easily things come to mind”, also known as their *accessibility* (Leron, 2006, p. 108). In some situations, S1 and S2 do not work together because S1 may produce a quick automatic non-normative response while S2 fails to intervene as “monitor and critic” (p. 108). Leron (2006) contends that many of the non-normative responses provided in psychological experiments, “and in mathematics education tasks, for that matter,” can be attributed to S1 and the frequent failure of S2 to intervene. (p.108) Leron (2006) points to a prime example of S1 acting unrestrained (taken from Kahneman, 2002): “A baseball bat and ball cost together one dollar and 10 cents. The bat costs one dollar more than the ball. How much does the ball cost? Almost everyone reports an initial tendency to answer ‘10 cents’ because the sum \$1.10 separates naturally into \$1 and 10 cents, and 10 cents is about the right magnitude. Frederick found that many intelligent people yield to this immediate impulse: 50% (47/93) of Princeton students and 56% (164/293) of students at the University of Michigan gave the wrong answer (p. 451).”

Leron (2006) claims that this problem creates “cognitive illusions”, similar to optical illusions talked about in cognitive psychology, and that the respondents who accepted the S1 answer uncritically “behaved irrationally”. For the respondents that answered correctly, Leron (2006) points out that S1 had also jumped to an answer, but their S2 interfered critically and made adjustments to their answer. In addition, Leron (2006) discusses the idea “that skills can migrate between the two systems” (p. 110). After someone becomes an expert at a certain task, usually after much training, that particular skill can become part of S1. Becoming an expert relates to Lim’s (2006) concept of *interiorized anticipation* which is a “way of thinking in which one spontaneously proceeds with an idea without having to analyze the problem

situation because one has interiorized the relevance of the anticipated action to the situation at hand” (p. 90). Lim (Personal Communication) believes that this “interiorized disposition” is related to impulsive disposition and analytic disposition in the same way that “expert S1” is related to S1 and S2. Generally, students whose problem solving actions are dominated by S1 are said to have an impulsive disposition, while students who are more trained to allow S2 to step in are said to have an analytic disposition.

The most important educational inference of this type of research, according to Leron (2006), “is the need to train *people to be aware of the way S1 and S2 operate, and to include this awareness in their problem-solving toolbox*” (p. 123). This belief supports the motivation for developing an instrument that allows students and educators to be aware of their own thinking disposition, and for investigating the validity of the LtA instrument.

### **2.3 PROBLEM-SOLVING FRAMEWORK**

Schoenfeld’s (1992) framework for exploring mathematical cognition through problem solving was used to inform the analysis of the problem solving section of the interview. The four categories in this framework include:

- Resources
- Heuristics
- Monitoring (Self-Regulation)
- Beliefs

#### **2.3.1 Resources**

Resources refer to a problem solver’s knowledge base. Schoenfeld (1992) contends that when one is examining decision-making during problem solving, the researcher needs to know what resources are available to problem solvers. Schoenfeld also points out that misconceptions and misunderstood facts possessed by problem solvers serve as false “tools” used by the problem solver, and therefore a participant’s knowledge base may also contain things that are not true which probably means they possess

a weak mathematical foundation within the context of a specific task. A student's knowledge base affects how the student responds in a mathematical problem situation. For example, a student with a strong knowledge base is more likely to know when to rely on S1 and when to invoke S2. However, it is also important to note that a weak or strong knowledge base does not necessarily imply an impulsive or analytic disposition.

### **2.3.2 Heuristics**

On the subject of problem solving strategies, Schoenfeld (1992) discusses heuristics such as the strategies of analogy, auxiliary elements, decomposing and recombining, induction, specialization, variation, and working backwards as being more complex and fluid than previously thought. Schoenfeld points out that despite their complexity, many researchers believe it is possible to delineate learnable and useable problem solving strategies. In the context of problem solving disposition, students who spontaneously use a learned heuristic without conceptual relevance, such as drawing a picture without conceptually tying it to the problem situation would be considered to have an impulsive disposition. However, heuristics as investigative tools, such as making an educated guess and checking your answer when previous approaches have failed would be indicative of an analytic disposition.

### **2.3.3 Monitoring (Self-Regulation)**

Monitoring is a form of self-regulation that takes place while someone is problem solving. Schoenfeld (1992) describes monitoring as a problem solver's ability to keep tabs on their progress in the middle of intellectual activity. In terms of monitoring, Schoenfeld explains that the goal is to get better at planning for tasks one is given and to improve "at making corrective judgments in response to feedback from their attempts." (Schoenfeld, 1992, p.355) In his more than a hundred videotapes of college and high school students working unfamiliar problems, Schoenfeld found that about 60% of the solution attempts were of the "read, make a decision quickly, and pursue that direction come hell or high water' variety...[a]nd that first, quick, wrong decision...guarantees failure" (p. 356). Within this thesis, such a

behavior would be considered a manifestation of impulsive disposition. Schoenfeld states that in order for problem solvers to give themselves an opportunity to solve the problem by way of good self-regulation, they must unlearn “inappropriate control behaviors developed through prior instruction...[which he believes can be achieved over] a long period of time, with sustained attention to both cognitive and metacognitive processes” (p. 357). This idea seems to suggest that one indeed can progress from impulsive disposition toward analytic disposition by increasing one’s capacity to self-regulate while problem solving.

#### **2.3.4 Beliefs**

Schoenfeld (1992) interprets beliefs as “an individual’s understandings and feelings that shape the ways that the individual conceptualizes and engages in mathematical behavior” (p. 358). Schoenfeld delineates typical student beliefs about mathematics, including:

- Mathematics problems only have one right answer
- There is only one correct way to solve a mathematics problem
- If you understand the mathematics you have studied, then you should be able to solve any assigned problem in five minutes or less.
- Ordinary students cannot expect to understand mathematics; they can only memorize it and apply mechanically what they learned without understanding.
- Mathematical truth is determined when the answer is ratified by the teacher.

Schoenfeld explains that students develop their beliefs about mathematics from their classroom experiences. Students’ beliefs, in turn, influence the way they behave in certain situations. Impulsive disposition can be taken as an externalization of beliefs, such as a student thinking a problem can be solved in five minutes or less (Lim, Morera, & Tchoshanov, 2009a). In analyzing the interview data, certain explicitly stated beliefs about mathematical problem solving were useful in explaining some respondents’ disposition indicators.

## **2.4 GENERAL LIKERT SURVEY CREATION AND TESTING**

### **2.4.1 Likert-type Scale**

Likert (as cited in Clason & Dormody, 1994) created a summated scale in order to assess particular attitudes in survey respondents. Likert's sample scale is made up of a series of statements whereby the respondents rate each statement based on any one of five possible responses. Each possible response is represented by one of the following numerical values: 1, 2, 3, 4 or 5. Each number corresponds to one of the following responses: *strongly disagree*, *disagree*, *unsure*, *agree*, and *strongly agree*. However, not all Likert-type scales follow the exact same structure. Some researchers choose to offer a different response scale. For example, the LtA survey uses a six point scale.

### **2.4.2 Writing General Likert Items**

After clearly defining the variable, or construct, for a particular population, you have to then create a large sample of survey items. In the early stages of writing survey items you should brainstorm both on your own and with colleagues in order to generate appropriate items for the survey (Trochim, 2006; Griffiee, 1999). Generally, once you begin to write out the possible survey items you should follow the rules set forth by Stone (as cited in Ockert, 2005):

1. Avoid true declarations.
2. Use the present tense in your wording.
3. Avoid ambiguous questions.
4. Do not use questions that create a general consensus.
5. Write clearly and straightforward.
6. Keep questions similar in length and not too long.
7. Do not include more than one concept in each item.
8. Stay away from compound sentences.
9. Consider the respondents' verbal abilities.

10. Stay away from double negatives.
11. Do not use “and”, “or”, or any lists of examples in your items.

General Likert rules were not followed to write the LtA survey items because general Likert items do not normally involve mathematical problems.

### **2.4.3 Evaluating Survey Items**

According to Griffie (1999) allowing colleagues and other experts in the area(s) that your study focuses on to evaluate your pool of potential items, you can select the items with the best discriminating ability whether they be on one side of the response scale or the other. The researcher should also talk with item judges about why they preferred some items over others, and vice versa, in order to understand their reasoning and be able to revise any items accordingly. Students should also be included in evaluating the language of the survey in order to avoid problems with reading level appropriateness. After the evaluation stage, the researcher is armed with the knowledge that a panel of colleagues agrees that the chosen survey items sufficiently measure the theoretical construct behind the survey. According to Angoff (as cited in Griffie, 1999), judge panel approval of survey items is an example of content validity that represents a consistency in the questionnaire itself, but this in no way validates the construct of the survey which is derived from the responses collected from the survey. Once the evaluation of the survey items is complete, the researcher must pilot test the survey.

### **2.4.4 Establishing Validity**

According to McDonald (1999), “a test score is valid to the extent that it measures the attribute of the respondents that the test is employed to measure, in the population(s) for which the test is used” (p. 197). Messick (1995) makes the assertion that evaluating the degree of fit of any test information with the theoretical rationale underlying test scores, can contribute to the understanding of score meaning. Messick (1995) also describes the traditional conception of validity as including three broad types of consideration: content, criterion, and construct validities. However, McDonald (1999) points out that through the work

of such researchers as Cronbach and Messick, the general consensus among many researchers is that test validity is mainly referred to as *construct validity*. McDonald specifically describes construct validity as “the validation of an application or class of applications of a test score [that] can be taken to include every form of evidence that the score to some acceptable extent measures a specified attribute –quantifiable property or quality—of a respondent” (p. 199). Validation investigation research efforts for this thesis were based on data from task-based interviews in order to compare self-reported problem-solving behavior to actual observed problem-solving behavior.

## **2.5 ANALYZING INTERVIEW DATA**

The general process of investigating to what extent the LtA items were valid started with the analysis of the interview data using the four steps as prescribed by Hesse-Biber and Leavy (2011) for analyzing qualitative data:

1. Data Preparation Phase
2. Data Exploration Phase
3. Data Reduction
4. Interpretation

The first step involved the transcription of each interview. This transcription was done verbatim and the problem solving section of the interview also included other behavioral cues such as ‘laughs’ and ‘counts on fingers’.

The second and third steps included memo-ing and coding the data which was more of a cyclical process when analyzing qualitative data. Memo-ing was done on the interview transcripts while simultaneously watching the interview videos. Questions that were typically asked while memo-ing included:

1. What is going on here?
2. What is the person doing?

3. What is the person saying?
4. How does context affect these actions and statements? (Charmaz, 2004)

Memo-ing also included: data summaries, quotes, and written ideas about the analysis and interpretation of the data. Memos were used to create codes that helped to manage important portions of data in order to uncover any themes, patterns, ideas, and/or concepts that existed in the data (Hesse-Biber & Leavy, 2011). Code creation helped to condense the data into more analyzable pieces.

The final step of interpretation was not distinct from the memo-ing and coding process, as stated by Hesse-Biber and Leavy (2011). This step involved synthesizing the findings and reporting them in such a way that addressed the research questions of the project.

## **Chapter 3: Research Methods**

Interview participants were chosen from a group of volunteers who took the LtA survey that was administered in their mathematics course. The purpose of the interview component was to help determine what type of actions and/or thoughts by an interviewee could be indicative of impulsive and/or analytic disposition, and also to help determine whether or not the LtA items were appropriate for assessing problem solving disposition. We sought to identify specific weaknesses in survey items so as to improve the overall validity of the instrument.

### **3.1 PARTICIPANTS**

Participants who took part in the classroom data collection consisted of 495 pre-service teachers from several undergraduate mathematics courses. Of the 495 respondents who took part in the data collections, 16 were interviewed. One interview was excluded because the majority of the responses were in Spanish. This research is mainly focused on the remaining 15 interviews.

### **3.2 DATA COLLECTION INSTRUMENTS**

The following instruments were administered to participants during the data collection effort.

#### **3.2.1 Three-in-One Survey**

The three-in-one survey (Appendix A) consisted of three surveys grouped together. The first survey was an 18-item Need for Cognition Scale (Cacioppo & Petty, 1982) which measured a respondent's interest level in engaging in complex thought. Participants had to respond how characteristic of them they considered each statement to be (on a scale of 1 to 5, 1 = extremely uncharacteristic of them, 5 = extremely characteristic of them). The second survey was the Barratt Impulsive Scale (International Society for Research on Impulsivity, 2010), or BIS-11, which was a 30-item survey that measured impulsive behavioral traits. Participants had to answer how frequently they performed certain actions described in each survey item (given four choices stemming from "rarely/never" on the left, to "almost always/always" on the right). Those first two surveys consisted of typical Likert items. The third survey, which was the

main focus of this research, was the 32-item Likelihood-to-Act (LtA) survey which consisted of lengthier math-based Likert items.

### **3.2.2 Two-Part Math Test**

The math test had two parts. Each part consisted of eight multiple choice problems, and three free response word problems. The math test was given in two parts in order to determine if respondents performed better on the second part after being warned that the math problems in the first part were “tricky”. The math scores were used as one of three measures to select possible interview participants.

### **3.2.3 Semi-Structured Interview**

Interviews were conducted with selected volunteers in order to better understand why respondents selected the survey responses they did. The 60 minute interview had three main parts:

1. The first part of the interview asked the participant to solve five open-response math problems (Appendix B). Four of the problems were based on LtA items, and the fifth problem was a story problem (Problem 3) that is presumably effective in eliciting an impulsive approach was included to see if the interviewee would instead analyze the problem situation. Interviewees were asked to express out loud what they were thinking as they were trying to solve each problem. They were also prompted to select a level of satisfaction in the way they approached each problem on a scale of 1 to 5 described on Placard A (Appendix C), with 5 being *very satisfied*.
2. In the second part, the participant was asked to respond to 10 LtA items comprised of four story items, four symbolic items, and two general items which were the same for all participants and were arbitrarily chosen to represent all of the types of LtA items in the original survey. In this part, interviewees were asked to say their response out loud as they decided which Likert response to circle. For items that differed by two points or more from their original response, interviewees were asked to explain why they were selecting their

current choices and why they believed their new response differed from their original response.

3. In the third part, participants were asked to classify the act described within 10 different survey items as either *impulsive* or *analytic* after being shown the working definitions for “impulsive disposition” and “analytic disposition” as presented on Placard B (Appendix D). Interviewees were asked to note if they were confident or not confident of their classifications. This section was used as another way to have participants talk about other LtA items that they had not yet seen during the interview.

The 10 items discussed in Part 2 of the interview, along with the 10 items that were classified in Part 3 of the interview, which included two different versions that were alternated amongst interviewees (Appendix H & I), allowed us to expose interviewees to 30 of 32 LtA items during each interview. The two remaining items were general non-specific items.

### **3.3 SEMI-STRUCTURED INTERVIEW SPECIFICS**

This section describes the details of how and why interview participants were selected, how the interview was developed, and how it was conducted.

#### **3.3.1 Interview Participant Selection**

Of the 495 students who participated in the in-class survey, 93 respondents volunteered to participate in the interview activity. Initially 18 interviewees were chosen to represent 6 different groups of respondents which were determined from the results of the 93 volunteers, but due to some cancellations and subsequent substitutions only 16 volunteers were interviewed. Of those 16 interviews, one interview was mostly in Spanish and was excluded from the study. The participants for the interview portion of the study were selected based on three criteria: high/low analytic response score, high/low impulsive response score, and high/low math score. The six groups consisted of HHH, HHL, HLH, LHL, LLH, and LLL; as expected there were no data points that fell in the HLL and LHH categories. A ‘high score’ included scores

in the top quartile (Q3) of the 495 research subjects, while a ‘low score’ included scores in the bottom quartile (Q1). The first letter of each group represented the average LtA analytic response (where 3.2 and below was considered low, while 4.3 and above was considered high); the middle letter represented the average LtA impulsive response score (where 4.06 and below was considered low, and 5.00 and above was considered high); and the third letter described the total math score (where 2 or less correct responses was considered low, and 5 or more correct responses was considered high). For groups where we did not have enough interviewees, we did not restrict selection to candidates in the top or bottom quartiles in certain categories. The final interviewee distribution is shown below in Table 3.1.

Table 3.1: Interviewee Distribution by Group

<b>Group</b>	<b>Interviewee</b>
<b>HHH</b>	<b>S02, S03, S11, S13(xHH)</b>
<b>HHL</b>	<b>S06, <del>S07(Spanish)</del>*</b>
<b>HLH</b>	<b>S05, S10</b>
<b>LHL</b>	<b>S01, S04(LHx)</b>
<b>LLH</b>	<b>S08, S14, S15(LLx)</b>
<b>LLL</b>	<b>S09(LLx), S12(LLx)</b>

*Note:* An “x” represents a score in the interquartile range.

\*The interview in Spanish was excluded from this study.

Five interviewees had at least one score which landed in the interquartile range. S09, S12, and S15 all had scores represented by LLx, but S15’s math score was closer to the top quartile of math scores while S09 and S12’s math scores were closer to the bottom quartile of math scores, so S15 was placed in the LLH category as opposed to the LLL category.

### **3.3.2 Interview Development**

The interview was originally developed with the following questions in mind:

1. How can the way a student answers a math problem be categorized as being either impulsive or analytic?
2. What is the first idea that crosses the interviewee's mind?
3. Does the interviewee check the appropriateness of an idea? If so, how?
4. Does the participant consider other ideas as well?
5. In what ways are LtA items assessing and not assessing impulsive-analytic disposition?
6. Do interviewees interpret the survey items as intended by the creators?
7. What are the interviewees thinking as they respond to an LtA item?
8. Are interviewees consistent in their survey responses? If not, how do they explain their inconsistency?

### **3.3.3 Interview Process**

The interview consisted of an ice-breaking introduction, followed by the three main parts as mentioned earlier. In each interview, there was an interviewee, an interviewer, and an observer. The observer could ask questions at the end of each part of the interview. Twelve of the 16 interviews were conducted by the author's advisor in order to demonstrate proper interview etiquette and protocol, while the author observed. The remaining four interviews (including the Spanish-speaking interview) were conducted by the author, while the author's advisor observed. For each interview, an observation form was used for note taking and for guiding the interview. The observation form contained the interviewee's original responses to the 10 LtA survey items for Part 2 of the interview. Interview questions were also embedded in the observation form to guide the interviewer through the various parts of the interview.

The interview was semi-structured in that it was guided by a set of questions (Appendix E). The following is the detailed breakdown of the entire interview process:

1. *Ice-breaking introduction:* Participants were thanked for taking part in the research process and were advised that the interview would last anywhere from 50 to 60 minutes. Interviewees were informed that the interview consisted of three main parts.
2. *Part I: Problem-Solving Section*
  - a) Each interviewee solved five free response problems, four of which were derived directly from the LtA items. Problems were given one at a time with almost an entire blank page for them to show their work. No time frame for problem solving was communicated to the interviewees; they were allowed to solve the problems at their own pace. Interviewees were also asked to think out loud as they solved each problem. Prompts were given to remind interviewees to think aloud. Questions were withheld at this stage. After providing an answer to each problem, the interviewees were asked how confident they were about their answers.
  - b) For each problem, interviewees were asked to rate how satisfied they were with the way they solved it and to explain their decision. Interviewees were then asked what the first idea was that came to their mind as they tried to solve each problem. Interviewees were also asked to share any other ideas they may have thought of when solving the problem.
  - c) Interviewees were asked to characterize the way they generally approach a math problem.
3. *Part II: Item Response Consistency Section*
  - a) The interviewee was asked to answer 10 LtA items that had appeared in the 32-item LtA survey they had previously completed in the classroom. Eight of the ten items corresponded to four of the free response problems in the problem solving section of the interview. Four of those eight were items that described an impulsive

act, while the other four were items that described an analytic act. The last two items of the 10 presented in Part 2 were general items (i.e. non math-content specific). The interviewee was asked to say out loud as he or she circled a response for each item, but they were not asked to explain their choice at this time.

- b) Recorded on the observation form were the participant's original responses to the 10 LtA items. The interviewee was asked to explain any response that was two or more points away from their original response. For example, if the interviewee had originally circled a 4 or less on one of the items, and during the interview they circled a 6 for that same item, they were asked to explain why they had previously circled a different answer.
- c) Questions were posed to invite interviewees to explain how they interpreted certain words in some of the items, especially when the interviewee did not appear to have interpreted the item in the way it was intended. Sometimes questions were posed regarding consistent responses when the interviewer suspected an instance of item misinterpretation or when there was extra time due to few inconsistent responses.

#### 4. *Part III: Item Classification Section*

- a) Interviewees were asked if they had learned anything new about themselves from the interview. They were also asked if they considered themselves as someone who would spontaneously use the first idea that comes to mind when solving a problem. Finally, interviewees were asked if they believed they were more impulsive or less impulsive than their classmates when solving math problems.
- b) At this point interviewees were introduced to the working definitions of impulsive disposition and analytic disposition on a separate placard (Appendix D). They were allowed to read over the two definitions and were asked if they felt they understood

the difference between the two terms. Each interviewee was then told to use these two terms to classify a set of 10 LtA survey items that were different than the ones they saw in Part 2 of the interview. Since the LtA survey had 32 items, we created two sets of 10 items for participants to classify; the two sets were evenly distributed amongst the 16 participants.

- i. Within each of the 10 items there appeared in boldface an approach for solving the problem presented in the item and respondents were asked if they felt the boldface act described was considered impulsive or analytic, and why.
- ii. Participants were also asked to note whether they were confident or not confident in their classification of each item.

All interviews were videotaped from two vantage points. The first camcorder focused on the interviewee's face and hand gestures while the second camcorder focused on their written work. The interviews were transcribed using the Transana transcribing software and then reformatted using Microsoft Word.

### **3.4 ANALYSIS PLAN**

This section contains an overview of how each interview section was analyzed. More detailed and problematic issues of methodology will be included in the discussion section with specific participant examples.

#### **3.4.1 Analysis: Part 1 of Interview**

There were a total of 75 problem solving episodes since 15 participants each solved five open-response problems (see Appendix B). The 75 episodes were first analyzed in two ways to generate two quantifiable scores. The first analysis involved data exploration and code creation to document and count the solution strategies participants used. In the second analysis, the video and transcript of each episode

was analyzed by a three member research team using a rubric and the complete problem solving episode was coded as I+, I-, U, A-, or A+. Then there was a third analysis that focused on an item-wise comparison of the rubric scores with the original LtA survey scores that corresponded with the problem solving episodes in Part 1 of the interview.

### ***Analysis 1: Coding and Counting Strategies***

In the data exploration phase, the transcript for each episode was reviewed and memos were recorded of any actions that seemed to be related to problem-solving disposition. Memos included instances dealing with procedural knowledge, conceptual understanding, problem-solving approaches, computational errors, and confidence of answer, as well as actions such as checking one's solution, drawing a picture, making computational errors, and expressing doubt. Memos along with how impulsive-analytic disposition can be inferred from the data were discussed during follow-up meetings with my thesis advisor.

In the code-creation phase, a set of codes was first created from the memos. The codes created were then collapsed into the following categories: solution strategies, mathematical knowledge, beliefs (e.g., math is tricky), and observable behaviors such as underlying keywords and finger counting. Since the problems were designed to elicit impulsive behaviors, certain solution strategies (e.g. dividing 220 by 5 for P3) were considered to be suggestive of impulsive disposition. Consequently, we focused on a set of solution-strategy codes, which was used to identify and count the number of impulsive and analytic approaches used within each episode.

In the quantification phase, it was noted whether a strategy was analytic or impulsive. For each interviewee, the number of instances of impulsive strategies was divided by the total number of instances of strategies attempted (both impulsive and analytic) to obtain the *Strategy-Count Percentage*.

## ***Analysis 2: Scoring Problem Solving Episodes with a Rubric***

In preparation for the second phase of data analysis, a rubric was created (see Table 3.2) based on the working definitions of impulsive disposition and analytic disposition in order to establish inter-rater reliability within a three member research team. Each episode was scored independently by each member of the research team by assigning one of five codes: “A+” (strong indication of analytic disposition), “A-” (weak indication of analytic disposition), “I+” (strong impulsive), “I-” (weak impulsive), or “U” (unsure). An expert-level code “E” was initially included to provide for responses where “one spontaneously proceeds with an idea without having to analyze the problem situation because one has interiorized the relevance of the anticipated action to the situation at hand” (Lim, 2006, p. 90), but there were no occurrences for “E” so that category was subsequently removed.

Table 3.2: The Rubric for Scoring a Problem Solving Episode

I+	Responds almost spontaneously to the problem with an action; no signs of studying the problem situation or considering the relevance of the action to the problem.
I-	Responds almost spontaneously to the problem with an action; weak indicators of trying to study the problem situation or considering the relevance of the action to the problem.
U	Exhibits actions that cannot be construed as impulsive or analytic or a combination of impulsive and analytic actions.
A-	Studies the problem but not thoroughly and/or consider the correctness of solution.
A+	Takes time to study the problem situation thoroughly and/or considers the relevance of planned approach before taking action.

The research team (comprised of the author, the thesis advisor, and another graduate student) met to review the rubric as well as to categorize nine problem solving episodes which were used as a training set. In scoring each response, the scorers watched and analyzed the video segment with participants’ written work and transcripts at hand. The scorers agreed to analyze an episode in its entirety instead of based solely on the solution strategy used by the interviewee. For situations where an interviewee had more than one indicator of impulsive/analytic disposition, a sub-score was assigned for each indicator,

and the entire episode was then re-evaluated to determine the final code. For example, an episode with “A+” and “I-” indicators would most likely be coded “A-”.

After scoring independently, the scorers then met to compare their codes for the remaining 66 episodes. The inter-rater reliability was 0.76 (there were 151 three-way agreements out of 198 comparisons).

As disagreements between the three raters were identified, each rater offered their explanation of why they rated the episode the way they did. Each rater kept track of the reasons behind their own scores on an excel file as part of the rubric scoring method. In cases where only one score was different, the rater with the different score was prompted to give state his reasons and to try to convince the other two raters of his accuracy. If the agreeing raters were not swayed, the agreeing raters would then attempt to convince the one different rater. These discussions included reconsideration of transcript and video evidence. Discussions continued as such until an agreement was reached. In cases where no two raters had a matching score, each rater tried to convince the other two of their position.

After calculating the inter-rater reliability as explained above, the author was made aware of more sophisticated methods for achieving a more precise inter-rater reliability than the one used for this study. Cohen’s (1960) kappa coefficient can be used between two raters. Since the rubric scoring method consisted of three raters, three separate pairwise Cohen’s kappa coefficients were found between the three raters. Cohen’s kappa between raters 1 and 2 was 0.645 (SE of kappa = 0.071); between raters 2 and 3 it was 0.623 (SE of kappa = 0.070); and between raters 1 and 3 it was 0.694 (SE of kappa = 0.067). All three strengths of agreement were considered to be *good*. The weighted kappa results were 0.841, 0.820, and 0.861, respectively, which means the strength of agreement for each of those measures was *very good*.

Then in order to quantify interviewees’ rubric scores, a numerical value was assigned to each code (1 for “A+”, 2 for “A-”, 3 for “U”, 4 for “I-”, and 5 for “I+”). The sum of the five scores for each

participant's problem solving episodes constituted the *Rubric Disposition Score*, which ranged from 5 to 25.

### ***Analysis 3: Item-wise Comparison of Rubric Disposition Scores with Original LtA Responses***

This part of the analysis consisted of an item-wise comparison between the Rubric Disposition Scores and the corresponding classroom LtA responses. This analysis was done on a two-scale basis.

#### ***Two-Scale Analysis***

In the two-scale analysis, both Rubric Disposition Scores and classroom LtA responses were transformed to a 0-1 scale ("0" being analytic, "1" being impulsive). For example, an impulsive episode with a score of I+ or I- was a "1", while an analytic episode with a score of A+ or A- was a "0". For LtA items, a response of 1-2-3 for analytic items was a "1", while a response of 4-5-6 was a "0". Conversely, an LtA response of 1-2-3 for impulsive items was a "0", while a response of 4-5-6 was a "1". Once the two sets of scores were transformed they were compared item by item in order to find the actual number of matches and mismatches between the two sets.

#### **3.4.2 LtA Response Consistency**

The item response consistency section of the interview was analyzed in five parts. The first part consisted of a quantitative summary that showed to what extent the 10 LtA items being studied in this section were consistent. Inconsistent responses in this study were interview responses that differed by at least two points on the Likert scale (1-2-3-4-5-6) from the original response provided in the classroom setting.

The second part of the analysis was a comparison of participants' classroom responses with their inconsistent responses to see in which direction of the Likert scale their answers moved. For every inconsistent response, we identified if the participant answered more impulsively or more analytically during the interview setting.

The third part of the analysis was a statistical test to see if there was a relationship between the types of LtA item involved (e.g. symbolic, story, non-math specific) and the incidence of inconsistent responses.

The fourth part of the analysis was qualitative. I examined the explanations given by participants as to why they thought their interview response was not consistent with their classroom response. The transcripts and videos were reviewed, memo-ed, and coded in order to categorize the various reasons for response discrepancies, and to find any other themes that were present in the data. Additional information offered by interviewees (i.e. beliefs, content knowledge, language interpretation) were noted especially if they were indicative of how effortful the interviewees completed the LtA survey.

The fifth part of the analysis was a qualitative study of the LtA items or parts of items that were not interpreted as intended by participants. In explaining their inconsistent answers, some participants directly and indirectly revealed some hidden issues about certain LtA items. These data were organized and presented for each problematic LtA item.

### **3.4.3 Classification of Acts Described in LtA Items**

The data in this section was analyzed in two parts. The first part was a quantitative summary of how many items were misclassified by the participants to get a sense of how well they understood the working definitions of impulsive disposition and analytic disposition. This data is intended to be used in future studies to refine the construct definitions. Nevertheless, the data is presented here for completeness sake.

The second part was qualitative, and focused on interviewees' explanations of their classification decisions. This classification section was used as an alternate way to get interviewees to talk about the other survey items not previously seen during the interview in order to identify the items that have characteristics that were problematic for them. Unfortunately, this section did not provide the amount of validity-based item discussion we had anticipated, but it did provide a rich data pool that may be used in

future studies to refine the construct definitions and/or to present disposition profile research based on the processes participants used to classify the problem solving acts described in the LtA items presented to them.

## Chapter 4: Results

The results are presented in three parts which correspond to the three main sections of the interview. Part 1 focuses on the results from the problem-solving section of the interview. Specific participant strategies for solving the five free-response problems, problem-solving behavior, and disposition cues will be presented.

Part 2 focuses on responses to the 10-item LtA survey completed during the interview. Participant explanations of inconsistent responses that differed by at least two points from their corresponding responses on the 32-item LtA survey will be presented, along with explanations for those inconsistent items.

Part 3 focuses on how interviewees classified the problem-solving actions described within a set of 10 LtA items as to whether each action is descriptive of an impulsive or analytic disposition and their explanation for each classification. This was an alternate method to getting interviewees to talk more freely about the items without reference to how they responded to those items.

### 4.1 INTERVIEW PROBLEM SOLVING RESULTS

This section presents the quantified measures that came from analyzing Part 1 of the interview which involved comparing impulsivity measures derived from solution strategies and problem solving approaches with the score for the 32-item survey taken earlier in a classroom.

#### 4.1.1 Cataloguing Solution Strategies as Impulsive or Analytic

Table 4.1 shows the list of solution strategy codes applied to the five free-response problems that participants solved in Part 1 of each interview. The table also lists the number of occurrences for each solution strategy per problem. Categorizing the solution strategies as either impulsive (shaded) or analytic (not shaded) was mostly a straightforward process because the interview free-response problems were designed to detect impulsive behaviors. For example, the two strategies of finding the common denominator for P1 and combining like terms for P2 are considered impulsive strategies

because in both cases the problem can also be solved with minimal computation. Moreover, all of the solution strategies found in Table 4.1, with the exception of DRAWPIC, were specific to the problem in which they occurred.

Table 4.1: Instances of Solution Strategies

Item	Strategy Code	Description	Instances
P1	CD10	Uses a common denominator of 10	1
	CD20	Uses a common denominator of 20	5
	CD40	Uses a common denominator of 40	8
	ADDLAST	Adds the last two fractions first	2
P2	COMBLT	Combines like terms	12
	CANCELTRM	Cancels any common terms found on both sides	6
P3	DIVIDE5	Divides the total distance by five	10
	DIVIDE4	Divides the total distance by four	6
	DRAWPIC	Draws a picture	12
	GUESS-N-CHK	Makes an educated guess and checks if it is correct	1
P4	RATIOCOMP	Creates and compares two ratios	4
	LOGIC	Uses direct reasoning to solve problem	4
	TOTALHOURS	Multiplies workers by hours to find total hours worked	10
P5	DIVIDE8	Divides the total distance by eight	1
	FORMULA	Uses speed = distance/time formula	1
	PROPORTION	Thinks of setting up a proportion	1
	DRAWPIC	Draws a picture	4
	MISSINFO	Believes problem is unsolvable; missing info	14
<b>Total number of solution strategies used</b>			<b>102</b>

*Note:* Shaded entries represent strategies indicative of an impulsive disposition.

P1 had a total of 16 strategies, P2 had a total of 18 strategies, P3 had a total of 29 strategies, P4 had a total of 18 strategies, and P5 had a total of 21 strategies. The total number of strategy instances (102) exceeds the number of problem-solving episodes (75) because some interviewees applied more than one strategy in a problem. Of those 102 observed solution strategies, 43 were considered impulsive with the majority of them occurring under the two non-contextualized problems (P1 and P2). The remaining 59 approaches that were considered analytic appeared mainly in the three contextualized problems (P3, P4, and P5).

Table 4.2 shows the number of impulsive strategies, the number of analytic strategies, and the percentage of impulsive strategies for each participant (calculated by dividing the number of impulsive solution strategies by the total number of attempted solution strategies).

Table 4.2: Impulsive-Analytic Strategy Count for Each Interviewee

Participant	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
Impulsive	5	2	2	4	2	3	5	3	3	0	3	3	2	3	3
Analytic	2	6	4	3	3	3	2	2	3	6	6	5	4	3	3
Impulsive Strategy %	62	22	33	50	40	50	71	60	50	0	33	38	33	50	43

Note: Shaded entries represent percentages greater than 50%.

Participants S1, S7, and S8 (three of the 15 interviewees) had Impulsive Strategy percentages of more than 50% which means they had more observable instances of impulsive solution strategies than analytic strategies. S1, S3, S5, S10, S11, S12, S13, and S15 (eight of the 15 participants) had Impulsive Strategy percentages of less than 50% which means they used more analytic strategy solutions than impulsive ones. S4, S6, S9, and S14 (four of the 15 participants) applied the same amount of impulsive and analytic solution strategies. The lowest Impulsive Strategy percentage score was 0% obtained by S10, while the highest score was 71% by S7.

#### 4.1.2 Assigning Rubric Disposition Scores to Problem Solving Episodes

Five episodes were chosen to illustrate how impulsive disposition and analytic disposition were inferred from the way interviewees solved the lamppost problem, P3.

##### *Impulsive, without drawing*

Participant S05 silently read the problem statement for P3 and spontaneously decided to divide 220 by 5.

S05: “So there are five lampposts and the distance from the first one to the second one is 220...(one second)...so it would be 220 divided by 5 (computes)...so it’d be 44 meters between each

lamppost.”

S05 hardly paused while working on this problem. Her facial expressions did not indicate any signs of analysis or doubts. She was one of the three participants (S5, S8, S14) who did not draw a picture to represent the problem situation. In addition, she did not seem to be constructing a mental visual of the problem situation. Her response was therefore coded as “I+”.

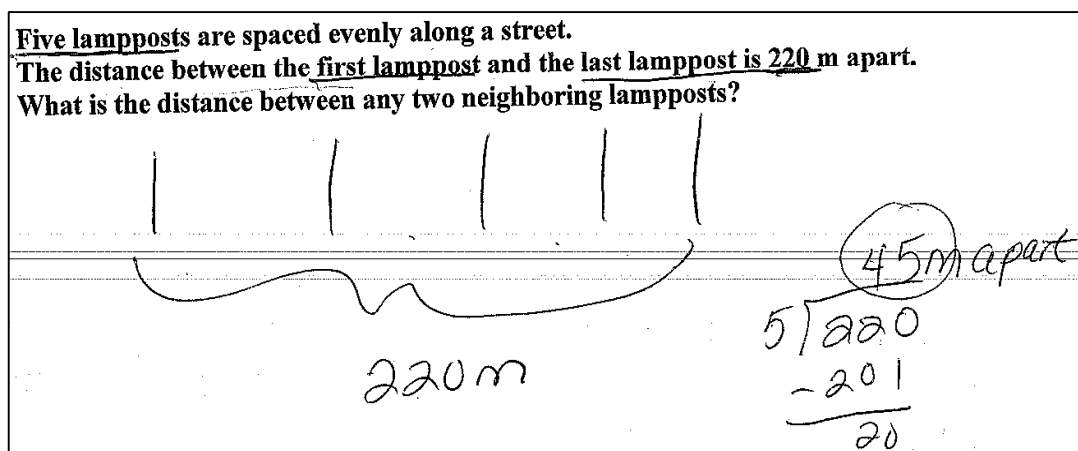
*Impulsive, with drawing*

As participant S4 read item P3, she underlined certain words. She then drew a picture to represent the problem situation (see Figure 1). However, the numbers she underlined seemed to suspend her sense making.

S04: “Five lampposts are spaced evenly along a street, so I underline things that I see with numbers...So then there’s five...(draws lampposts and labels distance)...and what is the distance between any two...(one second)... so then it is five divided into 220 ...

This response was coded “I+” because the drawing was created more as a habit rather than to gain a better understanding of the problem situation. S04 did not display any other physical cues that would suggest she was analyzing the problem situation.

Figure 4.1: A Drawing for Problem 3



### *Impulsive, ignoring the four spaces*

Like S04, participant S06 immediately began to draw a picture. Unlike S04, her picture was drawn in stages as she was rereading the problem. It was evident from the video that S06 was trying to understand the problem when she drew the picture and visually acknowledged that there were four spaces between the five lampposts (“A-“ sub-score was given for this part).

S06: “The distance between the first lamppost, and the last, is 220 meters apart, and there are 5 of them, right? (Draws lampposts) So then it is one, two, three, four, five, so between here (draws an arc over each of the four spaces) we divide 220 by 5...”

S06 ignored her figure and divided 220 by 5. Upon obtaining 44, she divided it by two to get 22 meters; she was influenced by the phrase “the distance between any *two* lampposts” and did not show any signs of considering the relevance of the dividing-by-two operation to the problem situation (“I+” for this part). The overall score was “I-” for this episode.

### *Analytic, noticing the four spaces later*

Participant S02 approached P3 in a systematic manner with a mental image of the problem situation and proceeded to draw a picture (“A+” for this part). Like S04 and S06, he was also influenced by the word “five” in the problem and decided to divide 220 by 5 (“I-” because he was not considering its relevance to the problem situation).

S02: “I do not start drawing it out because I’m a visual learner, so first off I imagine that I am actually seeing the street [and] the five lampposts, so then I begin to draw... Then I notice that there are five, since there are five, what I want to do then is 220 divided by five, which it is supposed to come out evenly...”

Soon after obtaining 44 meters, S02 added 44 to 44, and then 88 to 88 (in order to compute four times 44 meters which corresponded to the four in-between spaces). Seeing that 176 was not equal to 220, he realized that his answer was incorrect. He then resorted to a guess-and-check approach by trying 50

meters and 60 meters before finding the correct answer of 55 meters. We assigned an “A-” for this episode because S02 was thinking about the meaning of the problem and making sense of the problem (quadrupling the answer in order to check it) although he did not analyze why 220 should not be divided by 5.

*Analytic, reasoning with drawing*

Participant S10 was a fast thinker who approached P3 thoughtfully and methodically. He drew a picture (similar to Figure 1) in a purposeful manner.

S10: “I am actually not very great with, I always see these problems as geometry, so let’s just...(one second)...picture (begins to draw), easiest for me to just draw them out myself...the distance between the first one and the last one is 220, so (counts spaces between lampposts) one, two, three, four...(computes division)...55, 55, 55, 55, I always like to check my work...two of them is 110, times two, is 220.”

We coded this response as “A+” because of his initial doubt (an indication of not being impulsive), his methodical approach, and his purposeful use of his drawing to determine the solution.

Table 4.3 shows the distribution of score codes that the scoring team assigned to each problem-solving episode as well as the resulting *Rubric Disposition Score* derived from quantifying and adding the five disposition codes assigned to each participant. For the purpose of quantitative analysis, an interval scale based on ordinal scale data was used. The lowest possible Rubric Disposition Score for any participant was 5 (obtained by S10 since “A+” had a value of 1, then five episodes coded as “A+” added up to 5), while the highest possible numerical score was 25 (but no one received five scores of “I+”).

Table 4.3: Disposition Codes and Rubric Disposition Scores\*

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
Problem 1	I+	I+	I+	I+	A+	I-	I+	I+	I+	A+	I+	I+	I-	I+	I+
Problem 2	I+	A+	I+	I+	I+	I-	I-	I+	I+	A+	A-	A-	I-	A-	A+
Problem 3	I-	A-	A+	I+	I+	I-	I-	I+	I-	A+	A-	A+	A+	I+	A+
Problem 4	I-	A+	A+	A-	A+	A+	I+	A+	A+	A+	A+	A+	A+	A+	A-
Problem 5	A-	A+	A+	A+	A+	A+	A-	A+	A+	A+	A+	U	A+	A+	A+
Rubric Disp Scr	20	10	13	18	13	14	20	17	16	5	11	12	11	14	10

*Note:* Light shading denotes most analytic interviewee while darker shading denotes tie for most impulsive.

\*Cronbach's alpha for quantified codes was 0.5313 (A+=1, A-=2, U=3, I-=4, I+=5)

S10 had the lowest Rubric Disposition Score with the lowest score possible. S1 and S7 shared the highest Rubric Disposition Score. The codes “I+” and “I-” resulted in 22 out of 30 (73%) total episodes for P1 and P2, whereas “A+” and “A-” occurred in 27 out of 30 (90%) total episodes for P4 and P5. These findings were quite consistent with those from Table 4.2. However, it is important to point out that an analytic (or impulsive) solution strategy did not necessarily imply that the interviewee possessed an analytic (or impulsive) disposition. In one case, participant S4 applied the DRAWPIC (drawing a picture) analytic strategy in Problem 3, but the episode was still coded “I+”. Conversely, interviewee S15 used the impulsive and erroneous RATIOCOMP (comparing ratios) solution strategy at first in Problem 4, but the episode was coded “A-” by the coding team because one of the coders pointed out that he was being investigative in his different “impulsive” approaches and he did not provide an answer since he was not certain if his strategy was correct.

#### 4.1.3 Correlation among Impulsivity Measures

In this section, Table 4.4 illustrates to what extent the Impulsive Strategy-Count Percentage (Table 4.2) and the Rubric Disposition Score (Table 4.3) were correlated with one another, and also the extent each was correlated with the LtA\_Difference score (computed by averaging the 16 differences between the impulsive item response and the analytic item response for each pair in the original LtA survey taken in class prior to the interviews). These three measures of impulsivity for each participant

are shown below. The top three impulsive scores in each row were highlighted in the darker shade of grey. These results appear to be consistent, except for S13, who had a high LtA\_Difference score but a relatively low Rubric Disposition Score and Impulsive Strategy Percentage. The least impulsive score for each measure was also highlighted in the light grey shade, all three of these scores belonged to S10.

Table 4.4: The Three Measures of Impulsivity for Each Participant

Participant	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
Imp. Strategy %	62	22	33	50	40	50	71	60	50	0	33	38	33	50	43
Rubric Disp. Score	20	10	13	18	13	14	20	17	16	5	11	12	11	14	10
LtA_Difference*	2.81	0.31	0.88	2.00	-0.56	1.13	0.44	0.81	0.44	-2.75	-0.31	0.44	1.50	0.25	0.81

*Note:* Light shading denotes most analytic score while darker shading denotes top three most impulsive.

\*Cronbach's alpha for original LtA responses used to calculate the LtA\_Difference score was 0.8765

Twelve of the 15 interviewees (80%) had a positive value for the LtA\_Difference score which means that on average those respondents were more likely to act in the way described in the high-response-impulsive items. Three interviewees exhibited more impulsive solution strategies than analytic solution strategies, and five interviewees had Rubric Disposition Scores greater than 15 on a scale of 5 to 25.

The Pearson's correlation between Impulsive Strategy-Count Percent and Rubric Disposition Score was 0.927 with a  $p$ -value of 0.000. This high correlation can be attributed to the fact that these two scores are based on the same data source –the solution strategies used in the problem-solving part of the interview. The correlation between Impulsive Strategy-Count Percentages and LtA\_Difference scores was 0.670 with a  $p$ -value of 0.006. The correlation between Rubric Disposition Scores and LtA\_Difference scores was 0.693 with a  $p$ -value of 0.004. The correlation between the LtA\_Difference scores and Rubric Disposition Scores excluding data for P3 (the problem not found in the LtA survey) was 0.765 with a  $p$ -value of 0.001.

The author concedes the problematic nature of using a difference score such as the LtA\_Difference score presented in the measures above. These types of scores lead to a loss of data since

a respondent who may have a high average response for items that describe an impulsive act may also have a high average response for items that describe an analytic act. De Gruijter and Der Kamp (2005) suggest that difference scores can be quite unreliable even when the separate sub-scores are highly reliable because when the true score change is zero then the reliability of the difference score is zero as well. McDonald (1999) also points out that the reliability of a difference score decreases as the sub-scores are more correlated.

#### **4.1.4 Accuracy of Disposition Measures**

Even though the results of the impulsive strategy-count method of assessment were highly correlated with the results from the rubric-scoring method, it was evident that the strategy method was not quite as accurate as the rubric-scoring method. One observed reason is that a strategy that is perceived to be impulsive may actually be analytic for that participant. For example, when S06 and S13 were adding the fractions in P1 (Appendix B) they did not see the quick addition between the two outside terms, but after some discussion the research team agreed that those students seemingly chose a common denominator of 40 not because they were applying the procedure of finding the common multiple of two different denominators, but because 40 would specifically require less calculation than choosing another denominator.

The strategy-count method did tend to be just as accurate as the rubric-scoring method for items where analytic thought coincided with a change of strategy. For example, S11 began by combining like terms for P2 (Appendix B), but as soon as S11 reached the right hand side of the equation she realized that it had many of the same terms as the left side and she began cancelling out terms instead of continuing with her initial approach of combining them. S11 began with an impulsive action and then switched to an analytic action as soon as she read the entire problem. For S11, we counted both an impulsive strategy and an analytic strategy for item P2 which helped to keep her Impulsive Strategy Percent (Table 4.2) score down.

On the other hand, a change in strategy was not always captured accurately due to the predetermined disposition that each strategy represented. For example, 12 participants started P3 (Appendix B) by drawing a picture and then seemingly used that picture to decide they needed to divide to find the answer. In the strategy-count method those 12 students were given an analytic point for drawing a picture relevant to the problem they were solving. The problem with this was that not all 12 participants who drew a picture used it analytically. It seems that some were drawing a picture out of habit. For example, participants S6, S7, and S9 all drew a relevant picture but failed to recognize that they needed to divide by four instead of five. In these situations, the strategy-count method does not account for differences between drawings out of habit and drawings for studying since the analytic point a respondent received for a drawing carried the same weight as an impulsive point within the same problem. In the rubric-scoring method those same drawings received an A- score that was then combined with an I+ score for dividing by five instead of four, which left S6, S7, and S9 with a score of I- instead of a less accurate score of 1 analytic count and 1 impulsive count in the strategy-count method which essentially cancel each other out and skew the results of items found to have more or less strategies applied to them.

#### **4.1.5 Item-wise Comparison of Problem Solving Disposition Codes with LtA Likert Scale Scores**

The preceding section examined how a participant's overall problem-solving disposition, as assessed by the LtA survey, compared to the problem-solving disposition observed during the interview. This section contains an item-wise comparison of the four free-response problems solved in Part 1 of the interview that had corresponding LtA item in the LtA classroom survey. Eight LtA items represent four of the open-response problems (P1, P2, P4, and P5) since there was an analytic and impulsive version of each LtA item. There was no comparison possible for P3 of the interview since there was no corresponding LtA item for it. Also, the P5 disposition code for participant S12 was omitted because they received a "U" which was not quantifiable given the scales being used for this part of the analysis.

The purpose of this item-wise analysis was to compare the actual probability with the expected probability of disposition matches between problem-solving rubric scores and LtA item responses.

### ***Two-Scale Analysis***

Table 4.5 below contains the Transformed Rubric Scores of Table 4.3. Each row was doubled in order to facilitate comparison between the four problem-solving episodes and the corresponding eight LtA item responses in Table 4.7. As previously mentioned, Problem 3 was not included in this analysis since it had no corresponding LtA items.

Table 4.5: Transformed Rubric Scores

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
<b>Problem 1</b>	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1
<b>Problem 1</b>	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1
<b>Problem 2</b>	1	0	1	1	1	1	1	1	1	0	0	0	1	0	0
<b>Problem 2</b>	1	0	1	1	1	1	1	1	1	0	0	0	1	0	0
<b>Problem 4</b>	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
<b>Problem 4</b>	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
<b>Problem 5</b>	0	0	0	0	0	0	0	0	0	0	0		0	0	0
<b>Problem 5</b>	0	0	0	0	0	0	0	0	0	0	0		0	0	0

*Note:* The shaded empty cells represent an excluded problem-solving episode.

From Table 4.5, the incidences of “0”s and “1”s were counted and used to calculate the probability of receiving a “0” (analytic disposition code) or a “1” (impulsive disposition code) for these eight episodes. The probability of receiving a “0” was 0.59322 (70/118) and the probability of receiving a “1” was 0.40678 (48/118).

Table 4.6 below contains the class LtA responses for the items that correspond to the problem solving episodes represented in Table 4.5. Table 4.7 below contains the transformed scores of Table 4.6.

Table 4.6: LtA Likert Responses That Correspond To Open-Response Problems

		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
Problem 1a	LtA-19	1	5	3	2	6	3	6	5	3	6	6	2	4	3	3
Problem 1i	LtA-03	6	6	6	6	1	5	5	6	5	1	6	6	6	6	6
Problem 2a	LtA-01	5	6	6	6	4	5	5	2	5	6	6	4	5	5	6
Problem 2i	LtA-17	6	6	6	6	6	6	6	5	4	1	6	6	6	6	6
Problem 4a	LtA-30	1	4	5	2	6	6	3	2	1	6	5	4	2	6	2
Problem 4i	LtA-14	6	5	3	6	1	6	2	5	2	6	5	5	6	1	3
Problem 5a	LtA-06	1	2	5	1	3	1	5	2	2	6	3		5	4	3
Problem 5i	LtA-22	5	5	3	2	3	6	3	4	1	1	6		6	1	1

Note: The shaded empty cells represent an excluded problem-solving episode.

Table 4.7: Transformed LtA Likert Scores

		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
Problem 1a	LtA-19	1	0	1	1	0	1	0	0	1	0	0	1	0	1	1
Problem 1i	LtA-03	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1
Problem 2a	LtA-01	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Problem 2i	LtA-17	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
Problem 4a	LtA-30	1	0	0	1	0	0	1	1	1	0	0	0	1	0	1
Problem 4i	LtA-14	1	1	0	1	0	1	0	1	0	1	1	1	1	0	0
Problem 5a	LtA-06	1	1	0	1	1	1	0	1	1	0	1		0	0	1
Problem 5i	LtA-22	1	1	0	0	0	1	0	1	0	0	1		1	0	0

Note: The shaded empty cells represent an excluded problem-solving episode.

From Table 4.7, the probability of receiving a “0” (analytic disposition code) was 0.432203

(51/118) and the probability of receiving a “1” (impulsive disposition code) was 0.567797 (67/118).

Table 4.8 shows the number of actual matches and mismatches after all transformed entries from Tables 4.5 and 4.7 were compared item-wise. In this table, a “0” signified a match and a “1” a mismatch.

Table 4.8: Actual Matches and Mismatches

		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
<b>Problem 1a</b>	LtA-19	0	1	0	0	0	0	1	1	0	0	1	0	1	0	0
<b>Problem 1i</b>	LtA-03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<b>Problem 2a</b>	LtA-01	1	0	1	1	1	1	1	0	1	0	0	0	1	0	0
<b>Problem 2i</b>	LtA-17	0	1	0	0	0	0	0	0	0	0	1	1	0	1	1
<b>Problem 4a</b>	LtA-30	0	0	0	1	0	0	0	1	1	0	0	0	1	0	1
<b>Problem 4i</b>	LtA-14	0	1	0	1	0	1	1	1	0	1	1	1	1	0	0
<b>Problem 5a</b>	LtA-06	1	1	0	1	1	1	0	1	1	0	1		0	0	1
<b>Problem 5i</b>	LtA-22	1	1	0	0	0	1	0	1	0	0	1		1	0	0

*Note:* The shaded empty cells represent an excluded problem-solving episode.

Five out of 30 comparisons were mismatches for the fractions arithmetic problem. Thirteen out of 30 comparisons were mismatches for the algebraic problem. Fourteen out of 30 comparisons were mismatches for the bigger project problem. Fifteen out of 30 comparisons were mismatches for the unsolvable speed problem. In terms of item/problem type, mismatches among symbolic items made up 30% of comparisons, while mismatches among story items made up 48.3% of comparisons.

A total number of 71 matches versus 47 mismatches were found. This means that 60.2% of LtA responses were consistent with their associated rubric episode disposition score, while 39.8% of them were inconsistent.

Since this thesis was mostly a qualitative study and because the LtA instrument was in the early stages of being validated, the author chose to dichotomize a lot of the quantitative data presented in this study in an attempt to simply provide readers with a sense of how the quantitative data related to the qualitatively scored problem solving episodes observed during the interview. The thesis advisors made the author aware that dichotomization of data often results in the systematic loss of measurement and the statistical loss of power, as found by Cohen (1983). In this study, the dichotomization of data was directly applied in the two-way analysis above that compared each LtA item response with its corresponding rubric score. As Cohen (1983) warns, the dichotomization approach used in this study

was to simplify the data at the author’s discretion, which unfortunately comes with the cost of “degradation of measurement”.

From the preceding tables the following contingency table for actual observed matches and mismatches between problem-solving episode codes and LtA Likert responses was created.

Table 4.9: 2 x 2 Contingency Table for Actual Disposition Matches and Mismatches

	Analytic Episode	Impulsive Episode	Episode Total
Analytic LtA Likert	<b>37</b>	<b>14</b>	51
Impulsive LtA Likert	<b>33</b>	<b>34</b>	67
LtA Likert Total	70	48	118

*Note:* Shaded cells represent disposition matches between LtA responses and Rubric scores.

Statistical analysis using Fisher’s Exact Test ( $p = 0.0139$ ) showed a statistically significant association between LtA responses and Rubric Disposition Scores.

These two-scale results only gave us a sense of disposition matches between the LtA responses and the Rubric Disposition Scores since the LtA responses were transformed by dividing the response scale in half where 1-2-3 represented one disposition and 4-5-6 represented another. Reducing six-scale data to a two-scale data means that a person who was very likely to act in an impulsive manner was automatically bundled in with a person who was only somewhat likely to act in the same way.

In the event that some middle of the scale (3-4) Likert responses may have been from respondents who were indecisive about their likelihood to act choice, an analysis of only extreme Likert responses (1-2 versus 5-6) were identified and compared with their respective Rubric Disposition Scores. The results were similar (Fisher’s Exact Test,  $p = .03$ ).

#### 4.2 CONSISTENCY BETWEEN CLASSROOM LTA AND INTERVIEW LTA RESPONSES

This section presents the results for interviewees’ responses to the 10 LtA items between the two settings. Explanations of inconsistent responses are also presented.

#### 4.2.1 Interview LtA Survey Responses

Table 4.10 is a summary of responses for the 10 LtA items interviewees answered in both the classroom and interview settings (Appendix F). These 10 LtA items are labeled with the prefix *ReDo*. The responses under each item are separated into two categories: original classroom survey response (*c*), and interview survey response (*i*). The shaded columns signify the impulsive items, while the non-shaded columns signify the analytic items. There were a total of 63 inconsistent responses (responses that differed by 2 points or more between classroom and interview setting) out of 150 total responses. For example, participant S01 had an inconsistent response for item ReDo-1 when they responded with a 1 during the interview (*i*) setting compared with their original response of a 5 in the classroom (*c*) setting. For the sake of comparison, the Likert-scale, which is supposed to be ordinal, is treated as interval data.

Table 4.10: 10-item LtA Response Summary

	A		I		A		I		A		I		A		I		A		I	
	ReDo-1 (LtA-01)		ReDo-6 (LtA-17)		ReDo-7 (LtA-19)		ReDo-2 (LtA-03)		ReDo-3 (LtA-06)		ReDo-8 (LtA-22)		ReDo-10 (LtA-30)		ReDo-5 (LtA-14)		ReDo-4 (LtA-08)		ReDo-9 (LtA-16)	
	<i>c</i>	<i>i</i>	<i>c</i>	<i>i</i>	<i>c</i>	<i>i</i>	<i>c</i>	<i>i</i>	<i>c</i>	<i>i</i>	<i>c</i>	<i>i</i>	<i>c</i>	<i>i</i>	<i>c</i>	<i>i</i>	<i>c</i>	<i>i</i>	<i>c</i>	<i>i</i>
S01	5	1	6	6	1	1	6	6	1	6	5	2	1	2	6	6	5	6	5	6
S02	6	2	6	6	5	1	6	6	2	5	5	2	4	6	5	1	5	6	5	4
S03	6	3	6	6	3	1	6	6	5	3	3	1	5	6	3	1	4	6	6	5
S04	6	6	6	6	2	1	6	6	1	6	2	1	2	2	6	3	6	6	6	6
S05	4	5	6	6	6	6	1	6	3	6	3	3	6	4	1	4	6	5	4	5
S06	5	4	6	6	3	2	5	6	1	6	6	5	6	4	6	5	6	6	4	5
S07	5	5	6	6	6	2	5	6	5	4	3	2	3	2	2	5	6	6	5	2
S08	2	4	5	6	5	4	6	6	2	6	4	1	2	6	5	2	4	4	4	2
S09	5	6	4	6	3	5	5	6	2	5	1	2	1	5	2	2	2	5	2	4
S10	6	6	1	1	6	6	1	1	6	6	1	1	6	4	6	1	6	6	5	6
S11	6	4	6	6	6	1	6	6	3	6	6	2	5	6	5	3	6	5	6	6
S12	4	6	6	6	2	1	6	6	3	3	2	4	4	2	5	1	5	5	5	3
S13	5	1	6	6	4	1	6	6	5	1	6	1	2	5	6	1	6	6	6	6
S14	5	5	6	5	3	1	6	6	4	6	1	1	6	1	1	6	4	5	3	4
S15	6	6	6	3	3	2	6	6	3	5	1	3	2	2	3	6	5	6	5	4

Note: Shaded columns denote high-response-impulsive survey items.

#### 4.2.2 Distribution of Inconsistent Responses

The incidence of inconsistent responses per participant per item is more readily seen in Table 4.11 below. The items for the 10 LtA items being studied in Part 2 can be grouped into three categories: symbolic items, story items, and non-specific general items. Items are also listed in corresponding pairs. For example, item ReDo-1 is the analytic version of an item while ReDo-6 is the impulsive version of the same mathematical problem.

Table 4.11: Instances of Inconsistent Responses and Explanations

	Symbolic Items				Story Items				General Items		Inconsistent
	ReDo-1	ReDo-6	ReDo-7	ReDo-2	ReDo-3	ReDo-8	ReDo-10	ReDo-5	ReDo-4	ReDo-9	
S01	1	0	0	0	1	1	0	0	0	0	30%
S02	1	0	1	0	1	1	0	1	0	0	50%
S03	1	0	1	0	1	1	0	1	1	0	60%
S04	0	0	0	0	1	0	0	1	0	0	40%
S05	0	0	0	1	1	0	1	1	0	0	40%
S06	0	0	0	0	1	0	1	0	0	0	40%
S07	0	0	1	0	0	0	0	1	0	1	30%
S08	1	0	0	0	1	1	1	1	0	1	60%
S09	0	1	1	0	1	0	1	0	1	1	50%
S10	0	0	0	0	0	0	1	1	0	0	20%
S11	1	0	1	0	1	1	0	1	0	0	50%
S12	1	0	0	0	0	1	1	1	0	1	40%
S13	1	0	1	0	1	1	1	1	0	0	50%
S14	0	0	1	0	1	0	1	1	0	0	40%
S15	0	1	0	0	1	1	0	1	0	0	30%
Inconsistent	46.7%	13.3%	46.7%	6.7%	80%	53.3%	53.3%	80%	13.3%	26.7%	

*Note:* Shaded cells represent responses that were explained by interviewees.

A “1” represents the presence of an inconsistent response, while a “0” indicates that the participant’s interview response was consistent with their initial class response. For this study, it was arbitrarily decided that a consistent response was any interview setting response that differed by less than two points from the classroom setting response. For example, since the difference between S05’s

classroom response of 4 for item ReDo-1 is only one point away from her interview response of 5, her response was considered consistent. A “1” was chosen to represent an inconsistent response since we were more interested in analyzing inconsistent items than consistent items.

A total of 64 response explanations were collected during the interview. Fifty-nine out of 64 explanations were provided by participants for inconsistent responses. Shaded cells in Table 4.11 indicated cases where participants were asked to explain an inconsistent response between the classroom setting and the interview setting. In instances where participants had too few inconsistent responses to explain in the time allotted, they were asked to explain some consistent responses (e.g. ReDo-1 & ReDo-7 for S04, and ReDo-4 & ReDo-5 for S06). There were also four instances where an inconsistent response was not discussed due to a time constraint (e.g. ReDo-5 for S13, ReDo-5 for S15, ReDo-6 for S09, and ReDo-10 for S12).

#### **4.2.3 Comparing Classroom LtA Responses with Corresponding ReDo Responses**

Table 4.12 provides a look at the direction that participants took when answering the ReDo items compared with their original responses to the same items in the classroom setting. “+I” denotes when an interviewee chose a survey response during the interview that was more indicative of an impulsive disposition, while “+A” signifies when they chose a more analytical response.

Table 4.12: Comparison of Classroom Responses with Inconsistent ReDo Responses

	Symbolic Items				Story Items				General Items	
	ReDo-1	ReDo-6	ReDo-7	ReDo-2	ReDo-3	ReDo-8	ReDo-10	ReDo-5	ReDo-4	ReDo-9
S01	+I				+A	+A				
S02	+I		+I		+A	+A		+A		
S03	+I		+I		+I	+A		+A	+A	
S04					+A			+A		
S05				+I	+A		+I	+I		
S06					+A		+I			
S07			+I					+I		+A
S08	+A				+A	+A	+A	+A		+A
S09		+I	+A		+A		+A		+A	+I
S10							+I	+A		
S11	+I		+I		+A	+A		+A		
S12	+A					+I	+I	+A		+A
S13	+I		+I		+I	+A	+A	+A		
S14			+I		+A		+I	+I		
S15		+A			+A	+I		+I		

The two symbolic items that describe an impulsive act, ReDo-2 and ReDo-6, had the smallest percentages of inconsistent responses. The average interview LtA responses on the Likert scale for the two symbolic impulsive act items was 5.7 for ReDo-2 and 5.4 for ReDo-6. Considering the heavily overall impulsive rubric scores for P1 and P2, the high rate of consistency for both ReDo-2 and ReDo-6, and the high average Likert response of applying the impulsive strategies described in these items suggest that most interviewees were indeed more likely to begin by finding a common denominator for ReDo-2, and to begin by combining like terms for ReDo-6.

For symbolic items in general, interviewees selected a more impulsive response in 13 out of 17 inconsistent items (76.5%). For story items, interviewees selected a more analytic response in 27 out of 40 inconsistent responses (67.5%). For general items, interviewees selected a more analytic response in 5 out of 6 inconsistent responses (83.3%). These results reinforce the findings for Part 1 of the interview where the interviewees were found to be more analytic while working on story items but impulsive while working on symbolic items

#### 4.2.4 Statistical Analysis of Inconsistent Response Counts and Item Types

Table 4.13 shows the total inconsistent and consistent counts for the different types of items.

Table 4.13: Inconsistent and Consistent Counts for ReDo Item Responses

Item Type	Analytic Items		Impulsive Items		Total Count
	Inconsistent Count	Consistent Count	Inconsistent Count	Consistent Count	
Symbolic Items	14	16	3	27	60
Story Items	20	10	20	10	60
General Items	2	13	4	11	30
Total Items	36	39	27	48	150

There were more inconsistent responses among story items (40/60) as compared to symbolic items (17/60) and general items (6/30). Statistical analysis showed a strong association between item type (symbolic, story, general) and the incidence of inconsistent responses when analyzed using both Pearson's Test of Independence ( $\chi^2 = 25.55$ ,  $df = 2$ ,  $p = 0.000003$ ) and Fisher's Exact Test ( $p = 0.0000026$ ).

Considering that most of the Rubric Disposition Scores for story items were analytic, it seems that many participants were not putting forth an adequate amount of effort on story items when completing the LtA survey in the classroom. It is possible that some participants responded carelessly in the classroom and did not analyze the problem situations as well as they did during the interview. In general, interviewees were more analytic on story items because they were in the interview setting and also from having solved the related open-response problems at the beginning of the interview. The results from this section suggest a possible difference between students' problem-solving behavior which is generally more analytic in solving an unfamiliar story problem, and impulsive-analytic disposition when responding to an LtA story item which seems to involve more ineffectual reading and artificial understanding.

There were slightly more inconsistent responses among high-response analytic survey items (36/75) than high-response impulsive survey items (27/75). However, the association between the number of inconsistent responses and impulsive-analytic item type was not significant as found using

Pearson's Chi-squared Test with Yates' Continuity Correction ( $\chi^2 = 1.75$ ,  $df = 1$ ,  $p = 0.186$ ) and Fisher's Exact Test ( $p = 0.185$ ).

#### 4.2.5 Participant Reasons for Inconsistent Responses

Table 4.14 provides a breakdown of explanations provided by participants for why they believed their interview setting response varied by two points or more from their classroom setting response. Since each inconsistent response explanation could be linked to either a classroom setting response (c) or an interview setting response (i), each explanation was assigned a setting first, then inconsistent response reason categories were generated as new explanations arose. As shown below, 13 inconsistent response reason categories were identified (see Appendix G for Code Definitions). Six inconsistent response reasons were found in the classroom setting only, and another two inconsistent response reasons were found in the interview setting only. The other five inconsistent response reasons were found in both survey settings.

Table 4.14: Reasons for Inconsistent Responses and their Frequency

	Symbolic Items								Story Items								General Items				Total	
	ReDo-1		ReDo-6		ReDo-7		ReDo-2		ReDo-3		ReDo-8		ReDo-10		ReDo-5		ReDo-4		ReDo-9			
	c	i	c	i	c	i	c	i	c	i	c	i	c	i	c	i	c	i	c	i	c	i
NOTSERIOUS	1		1						3						1				1		7	0
MISREAD					2				3												5	0
NOREADWELL	1								1				1						1		4	0
RUSHED	1				1				1												3	0
ASSUMEDSIMPLICITY	1				2																3	0
NORLX	1														1						2	0
SOLVECHANGE		2								1		6		4		1					0	14
INCONSISTENT		1				2								1							0	4
INFLUENCED			1					1	2		5				2	1					10	2
NORECALL	1				3				1		2		1			1					8	1
LANGUAGE				1		1								1	1	1		2	1	1	2	7
EXPERIENCE														1	3	1			1	1	4	3
WKCONTENT	1								2			1									3	1

*Note:* Lightest shade denotes interview specific explanations, medium shade denotes explanations used in both settings, and darkest shade denotes classroom specific explanations.

For the 64 responses that were explained, a total of 83 reasons were provided by interviewees because some participants cited more than one possible reason to explain an inconsistent response. Of these 83 reasons, 32 out of 83 (38.6%) were used to explain interview setting responses, while 51 out of 83 (61.4%) were used to explain classroom setting responses.

The most commonly used inconsistent response explanation interviewees gave for a classroom response was INFLUENCED with 12 occurrences which refers to cases where a respondent felt that they were more likely to act in the manner described within the survey item merely because it was stated in the item. For example, for ReDo-8 (LtA-22) several participants indicated that in the classroom setting they were more likely to use  $d = rt$  to find the speed as stated in the item (an incorrect approach), but when they actually analyzed the problem situation during the interview setting, they solved the problem analytically and did not consider the impulsive act described in the item.

The second most frequent explanation given for classroom responses was NORECALL with 9 occurrences, which was a code assigned to cases where a participant could not remember why they chose a particular response. This may have been due to the two month interval between when the LtA survey was originally completed, and when the interviews were held.

Specifically, an example of story item inconsistency bias can be seen in ReDo-8, which looks to have been superficially read by at least five interviewees that claimed that they had been influenced by the impulsive act described in the item when responding to it in the classroom. Consequently, S01, S02, S08, S11, and S13 went from an average Likert response of 5.2 to 1.6 likelihood that they would use  $d = rt$  to find the speed. These interviewees seemingly responded more analytically in the interview setting after having analyzed ReDo-8 in the interview. It is difficult to know whether these interviewees would have answered analytically if they had not had to solve the problem at the start of the interview, but this result showed that interviewees tended to be more analytic when they actually took time to analyze and understand the problem situation that was presented to them.

#### **4.2.6 Explanations of Inconsistent Responses Exclusive to the Classroom Setting**

One of the reasons that interviewees appeared to be more analytic during the interview is due to the setting where they were asked to think aloud as they solved the problems. Reasons related to classroom setting are presented below.

##### ***Classroom Survey Does Not Count Against Us***

Some participants expressed their lack of seriousness in taking the survey in the classroom setting. S01 claimed that she did not read the item well in two of her three inconsistent items (any items where there was a two or more point difference between her class response and her interview response). S01 explained the difference of mood was because “[the researcher explained that] the first [survey] does not count against us” and so she just wanted to “hurry up and do it”. For the interview setting she said, “I am relaxed, I’m looking at it, thinking about it, saying it out loud and I think this is more accurate.” S01 explained that she has two ways of doing math problems, “if it counts I do it one way, if it does not then another way.” Since she felt that the classroom survey did not count, she “just answer[ed] the question [and] did not verify.” This suggests that when completing this non-graded participatory activity, participants might be skimming over survey items without giving them much thought. This lack of seriousness affects the reliability of the LtA instrument.

##### ***Difference of Approach in Classroom versus Interview Setting***

S09 had six inconsistent items, she claimed in three of them (ReDo-3, ReDo-9, ReDo-10) she might not have read the item well: “When I did [the survey in class], I did not really read it the way I am looking at it [during the interview].” Moreover, S09 also believes, “since I already worked with [the problems in Part 1 of the interview] and had to use my head, my answers changed.” S09 then said, “me and math do not get along” so she normally has to take her time with math problems. S09 also points out that seeing math problems without working them out is different and she would do “whatever comes to my head”, but working them out and then talking about them caused her to be “thinking and thinking”

about the problems. In a similar manner, S13 attributed one of her six inconsistent items to changing her mind after having solved the problem in Part 1 of her interview. S13 explained that the reason she was *extremely unlikely* to use the  $d = rt$  formula for the speed problem was because in “[the classroom survey items] I was not asked to solve [them], I [was] just asked to [say] how likely I [was] to use [a] strategy” and because of this “I did not see that [the speed problem] was not solvable.” S13 said that just reading the item without having to work it out “it makes sense to [use the speed formula], but then...when you have to actually work it out you see that it is not workable.” S13 mentioned that she felt she had to hurry because the survey was timed. She explained that since “[the researcher] asked [the class]...to write down the time when they were doing it, [she thought]...I have to hurry...whatever okay...[that answer] sounds good.” This explanation shows how some students may have applied shallow thinking to some of the survey problems in the classroom setting because they felt rushed. These explanations suggest that in a classroom setting some respondents tended to answer the LtA survey with the first thing that comes to mind, but in a problem-solving setting where they were allotted more time they were more likely to analyze the problem situations contained in the survey items.

### ***Classroom Setting Allowed For More Irrelevant Manipulation of Quantities***

Some participants explained that they had been influenced by the numbers in the suggested solutions for the impulsive items of the survey. S02 claimed that the first time he took the survey in a classroom: “I was actually excited to...get my hands on the numbers and play with them” because it had been a while since he worked with numbers in his math classes. In the classroom setting, S02 wanted to “[look] at the numbers and then [try] to solve [them] as fast as possible.” S02 was influenced by the numbers in the items in three (ReDo-3, ReDo-5, ReDo-8) out of five inconsistent items. His interview responses to those three items would have made him seem more analytic. In regard to his interview setting frame of mind, S02 said, “I am actually reading the problem...I am more calmed down...more serious about everything in general.” In explaining a classroom response, S02 says about dividing the

numbers in the problem to find the speed in ReDo-3: “I felt that this was right, but then when I did it [during Part 1 of the interview], I felt like analyzing the problem was better.” During the interview, S02 pointed out that even if he had not solved the speed problem in Part 1 of the interview, he would have analyzed the item in Part 2 of the interview unlike in the classroom, and that in general he is “more like today”. He explained that the reason he was impulsive in the classroom was because “in any math problem...we look for answers no matter what it’s saying, [you] look at the 8 [o’clock] and 2.4 [miles].” S02’s desire to manipulate numbers in the classroom setting was probably due to a lack of seriousness which geared him to act more impulsively.

### ***Influenced by the Act Described Within The Item***

S04 had two inconsistent items, and she explained that in both cases she was influenced by the suggested solutions found within the survey items. S04 explained that the survey item where you need to find the speed “gives you the idea of dividing...because of that...[it] got me thinking oh yeah, well if you divide...this by 8 you would find [the speed]...[but] right now when you gave me the problem without that I was like...you need the information.” S04 believes that having the suggested solution relieved her from having to think about the problem in the classroom, whereas in Part 1 of the interview she thought “well 8 o’clock, distance, but what time did she leave?” S04 agreed that if a solution is suggested she would just agree with it without really thinking about the problem. In Part 1 of her interview, S04 analyzed the problem situation and figured out that there was not enough information to find the speed. This behavior differed from the way she responded to the corresponding LtA item in the classroom setting.

Similarly, in the bigger project comparison item (ReDo-5), S05 seemed to be influenced by the suggested solution during the interview because she answered with somewhat likely to compare rates in the interview setting compared with extremely likely to not compare rates in the classroom setting. Upon reflection during the interview she revises her response in Part 2 from a 4 to a 1 because it is more

representative of her. S05 explained, “I get it why [comparing rates] is here, but I [did not] think of it.” S05 was unable to explain why comparing rates was a viable approach, so it is unclear why she thought she was somewhat likely to compare rates since in Part 1 of the interview she only multiplied and compared the total hours worked for each project. Out of five inconsistent items, S11 denoted that one of those responses was influenced by the suggested solution presented in the survey item. During the class survey, S11 responded as *extremely likely* to use the  $d = rt$  formula because she “focus[ed] on the formula to answer the [survey item], but not so much the problem.” Having determined in the problem solving part of the interview that the problem was unsolvable, S11 pointed out that “the first thing I thought of...[was to] use [the] formula...but then when I [tried] to solve it is when I realized I cannot...I am missing information.” It seems that in this case S11 was not analyzing the problem situation in the classroom because the suggested solution was compelling to her. It seems that items with suggested solutions can be problematic in assessing a respondent’s disposition because we may not be able to tell if a participant is exhibiting their true disposition for that particular item, or if they are simply being influenced by the suggested solution stated within the item.

### ***Problem Seemed Familiar and Simple Until Now***

Participants such as S03 claimed that some of their approaches in class were more based on assumed simplicity and familiarity instead of analysis. After seeing “*solve for x*” (ReDo-1), S03 explains, “I thought maybe it was likely that I would notice the solution at [that] time” and she agreed that she made that assumption without really checking to see if she could solve it or not. She explains: “After having to solve [the problem], I realized that I wasn’t trying to notice the solution...I noticed that...now it would take me...a while to compute.” S03 assumed she could notice the solution to the problem based on her familiarity of the problem procedure. Similarly, in the fraction addition item she thought, “I might study the fractions, but then...noticed...they are not common denominators so...I wouldn’t even try to predict the answer unless they were common denominators.” Without actually

having to work out the problems, S03 assumed she would answer analytically on the basis of familiarity with the proposed problem.

These interviewee explanations suggest that settings (classroom versus interview) may impact the way certain respondents will complete the LtA survey items.

#### **4.2.7 Explanations of Inconsistent Responses Exclusive to the Interview Setting**

The main explanation provided by interviewees about an inconsistent response that was specifically related to the interview setting was SOLVECHANGE with 14 occurrences which was a code assigned to instances where an interviewee realized that they would solve a problem differently after actually having solved it in Part 1 of the interview. At least six different interviewees reported this reason when explaining an inconsistent response for ReDo-8, which was the unsolvable speed item that described an impulsive act.

#### **4.2.8 Items Not Interpreted As Intended**

One of the themes that arose from analyzing the transcripts was interviewees' misinterpretation of survey items. Five of the 10 LtA items in Part 2 of the interview had multiple instances of participants interpreting a term or phrase within some survey items in a way that was not intended by the survey creators.

Table 4.15 below shows the summary of items that were misinterpreted along with the participants whom misinterpreted it and a short explanation. Interviews did not reveal any instances of unintended interpretation for items ReDo-2, ReDo-3, ReDo-6, ReDo-8, and ReDo-9.

Table 4.15: Instances of Item Being Misinterpreted

Item	Participant	Interpretation
<b>ReDo-1</b>	S04, S05, S06, S07, S09	“Study and notice the solution” means to combine like terms
<b>ReDo-4</b>	S01, S05, S06	“Analyze the terms in the equation” means to study a word problem
	S03, S04, S07, S13	“Analyze the terms in the equation” means to combine like terms
	S09	“Terms” means mathematical definitions in textbooks, instead of equation terms
<b>ReDo-5</b>	S06, S08, S14	“Compare rates (e.g., comparing 30/8 to 20/3)” means comparing the workers and hours of each project instead of comparing ratios
<b>ReDo-7</b>	S04, S09, S15	“Predict” means to find the answer through mental computation
	S11, S14	“Predict” means to only estimate the answer
<b>ReDo-10</b>	S14	“Predict” means strictly estimating without finding an exact answer

For ReDo-1, participants S04, S05, S06, S07, and S09 interpreted *how likely are you to study and notice the solution* as “[seeing] that there are terms like coefficients with the letters so you know to have add those...terms together, and the numbers...add those together...to get x by itself”(S04). S09 explained that she was “comfortable with this kind of problem...I think because you just got to move...the variables to one side.” All five participants had consistent responses ranging from 4 to 6 on the Likert scale, but they simply combined like terms when they solved this problem. Hence, these five participants did not interpret ReDo-1 in the way it was meant to be interpreted.

For ReDo-4, participants S01, S05, and S06 interpreted *how likely are you to analyze the terms in the equation* as likelihood to analyze a story problem. For example, S01 interpreted the act as “[looking] for certain keyword or phrases,” and S05 commented “maybe it is a word problem, the instructions...what they are trying [to get] you to look for.” S06 said “ [the] first thing I try to do is picture it and if I have to draw it I will draw it and...analyze it first before I get into it.” It is clear that these participants were erroneously interpreting the terms of an algebraic equation as words in a story problem. S01, S05, and S06 all had consistent responses of either 5 or 6 on the Likert scale.

S03, S04, S07 and S13 had a different interpretation of ReDo-4. These participants interpreted analyzing the terms of the equation as “analyzing the like terms and thinking about combining them”(S03). S07 said that “it means numbers and variables...separating them.” Three of the four

participants responded with a six in both the classroom and the interview setting; only S03 had an inconsistent response by going from a 4 to a 6 on the Likert scale.

S09 interpreted the “terms” in ReDo-4 differently. She seemed to interpret the terms as vocabulary terms in mathematics: “the terms, as in like the math terms, like maybe...terms that nobody is used to unless you practice it...on a daily basis.” Later in the interview, S09 agreed that the math terms she mentioned earlier were mathematical words found in math textbooks. S09’s response went from a 2 to a 5.

For ReDo-5, participants S06, S08, and S14 interpreted the numerator and denominator in each fraction in *how likely are you to compare rates (e.g., comparing 30/8 to 20/3)* as standalone quantities instead of a composite unit of rate measured in terms of workers per hour. S06 explained “I did compare the two projects in matter of [hours] worked, considering the 30 and the 8, and the 20 and the 3, but I did not put it over like [the fractions stated in the survey].” S06 saw the 30/8 and 20/3 as a summary of the four quantities being compared instead of two ratios. Of these three participants, S06 had the only consistent response going from a 6 to a 5 on the Likert scale. S08 had an inconsistent response by going from a 5 to a 2 on the Likert scale, but admitted that her class response was due to her misunderstanding of the presented ratios, she explained “I think I misunderstood what the 30 over 8 and 20 over 3...I meant comparing as far as comparing...the amount of workers and the amount of hours.” S14, on the other hand, went from a 1 in the classroom to a 6 in the interview. She was comparing rates as comparing total hours worked: “to me [comparing rates means] this is bigger than this one...there is more...[because] 30 times 8 is obviously going to be bigger than 20 times 3.” S06, S08, and S14 all solved the bigger project comparison problem in Part 1 of the interview by multiplying and finding the total hours worked on each project. The fact that none of these three interviewees used any ratios in their calculations shows the propensity of some respondents not to interpret the term “rate” as a measure that is computed by dividing one quantity by another quantity.

For ReDo-7, participants S04, S09, and S15 all interpreted the word *predict* in the statement *how likely are you to begin by studying the fractions to see if you can predict the answer* as a mental computation. S04 pointed out “I need to write it down and add it, I cannot do it at the top of my head unless it is two plus two,” while S15 said “I was not likely to try to do [it] in my head and start finding...common denominators...[because] it is too hard for me.” Although S09 did not mention that mental computation was difficult for her, she interpreted studying and predicting as “[looking at the problem and] know[ing] the steps I am supposed to do.” Of the three participants, only S09 had an inconsistent response going from a 3 to a 5 on the Likert scale so she possibly was more comfortable with the idea of mentally computing the answer in the interview setting. S04 went from a 2 to a 1 while S15 went from a 3 to a 2, both remaining consistent. Since all three of these participants applied the common denominator approach in Part 1 of the interview, all three might have assumed that they had to mentally solve the problem using the common denominator method when asked to predict the answer. For some interviewees, this item seems to be measuring how comfortable they are with mental computation instead of how likely they are to recognize the more simple and efficient analytic approach of mentally adding the last two fractions first.

S11 and S14 had a slightly different interpretation of the word *predict* in ReDo-7. Both participants contended that predicting was a form of estimation to see “more or less [what the answer] will be,” but that it is “not the actual answer”...just “close enough...to see if they got the right answer”(S14). S11 said that she “would not predict anything”...because she “would kind of go towards what is being asked...[and] would try to solve [the problem] instead of predict.” Both of these participants had inconsistent responses, S11 went from a 6 to a 1 while S14 went from a 3 to a 1. S11 selected a 1, which signifies *extremely unlikely*, because predicting was seen as an undesirable approach since it does not give an exact answer. S14, on the other hand, selected a 1 because she regarded the common denominator approach she used for solving the problem as different from predicting.

For ReDo-10, participant S14 had a sense of the relative magnitude of two products but she did not consider that as prediction. S14 changed from a 6 to a 1 on the Likert scale for this item. She chose *extremely unlikely* in the interview because seeing which project was bigger: “obviously 26 times 10 is...going to be more than 18 times 7” was not considered predicting for her. Interestingly, she admitted that she “would still have to calculate [the total hours worked]”(S14) even though it is obvious to her which multiplication is bigger.

A shaded cell in Table 4.16 (a version of Table 4.11 with different shading) indicates an instance of misinterpretation for that item. Among the 22 instances of misinterpretation, 14 of them occurred when the participants were consistent in their responses across the two settings (classroom and interview). This is interesting because an item may elicit consistent responses (i.e., reliable) but may be misinterpreted by participants (i.e., not valid).

Table 4.16: Instances of Misinterpretation from Part 2 of Interview

	Symbolic Items				Story Items				General Items	
	ReDo-1	ReDo-6	ReDo-7	ReDo-2	ReDo-3	ReDo-8	ReDo-10	ReDo-5	ReDo-4	ReDo-9
S01	1	0	0	0	1	1	0	0	0	0
S02	1	0	1	0	1	1	0	1	0	0
S03	1	0	1	0	1	1	0	1	1	0
S04	0	0	0	0	1	0	0	1	0	0
S05	0	0	0	1	1	0	1	1	0	0
S06	0	0	0	0	1	0	1	0	0	0
S07	0	0	1	0	0	0	0	1	0	1
S08	1	0	0	0	1	1	1	1	0	1
S09	0	1	1	0	1	0	1	0	1	1
S10	0	0	0	0	0	0	1	1	0	0
S11	1	0	1	0	1	1	0	1	0	0
S12	1	0	0	0	0	1	1	1	0	1
S13	1	0	1	0	1	1	1	1	0	0
S14	0	0	1	0	1	0	1	1	0	0
S15	0	1	0	0	1	1	0	1	0	0

Note: Shaded cells denote an incidence of misinterpretation.

### 4.3 CLASSIFICATION OF SURVEY ITEMS

In this section interviewees were presented with the definitions of impulsive and analytic disposition and then were asked to classify the act described in each of 10 LtA items as either impulsive or analytic.

#### 4.3.1 Classifying LtA Items as Analytic or Impulsive

Two distinct versions of the 10-item classification exercise were alternated among the interviewees: Version A (Appendix H) and Version B (Appendix I). Table cells containing incorrect classifications were shaded.

Table 4.17: Item Classification, Version A

Version A	A01	A02	A03	A04	A05	A06	A07	A08	A09	A10	Misclassified
	LtA-09	LtA-02	LtA-11	LtA-23	LtA-28	LtA-13	LtA-10	LtA-15	LtA-05	LtA-32	
S01	A	I	I	A	A	A	A	A	I	I	30%
S04	I	A		I	I	A	I	I	A	A	78%
S07	I	A	A	A	A	A		A	A	A	67%
S08	A	A	I	I	I	A	I	A	A	I	60%
S09						A	A	A			33%
S10	A	A	I	A	I	I	A	A	I	A	10%
S13	A	I	I	A	I	I	I	A	A	A	20%
S15	A	A	I	A	A	A	I	A	A	A	50%

*Note:* Shaded cell denotes an item that was classified wrongly.

Version A had 31 out of 72 misclassifications. There were 8 instances where a classification was not provided. Items A01, A03, A04, A08, and A10 were misclassified two or less times. A06 and A09 were both misclassified by 6 out of 8 participants.

Table 4.18: Item Classification, Version B

Version B	B01	B02	B03	B04	B05	B06	B07	B08	B09	B10	Misclassified
	LtA-25	LtA-18	LtA-27	LtA-07	LtA-12	LtA-29	LtA-26	LtA-31	LtA-21	LtA-16	
S02	I	A	A	I	A	A	I	I	A	A	10%
S03	I	A	A	I	A	A	I	I	A	I	0%
S05	A	I	A	A	I	A	I	I	A	A	50%
S06	I	I	I	A	A	A	A	A	I	A	70%
S11	A	A	A	I	A	A	A	I	I	A	40%
S12	A	A	I	I	A	I	I	A	I	A	60%
S14	A	A	A	A	A			I		A	43%

Note: Shaded cell denotes an item that was classified wrongly.

Version B had 26 out of 67 misclassifications. There were 3 instances where the participant was not able to provide an answer. Items B02, B03, B05, B06, B07, and B08 were misclassified two or less times. B10 was misclassified by 6 out of 7 participants.

Table 4.19 is an extension of Table 4.4 by including the inconsistency percentages in Table 4.11 and misclassification percentages in Tables 4.17 and 4.18 to the three measures of impulsivity. Table 4.20 shows the correlation among these five measures. As reported earlier, the correlations between measures of impulsivity are high. The relatively high correlation values between the misclassification percentage and the impulsive strategy-count percentage and between misclassification percentage and rubric disposition scores suggest that students who are impulsive in solving problems (Part 1 of the interview) are more likely to misclassify the LtA items.

Table 4.19: Comparing Measures of Impulsivity, Inconsistency, and Misclassification

Participant	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
Imp. Strategy %	62	22	33	50	40	50	71	60	50	0	33	38	33	50	43
Rubric Disp. Score	20	10	13	18	13	14	20	17	16	5	11	12	11	14	10
LtA_Difference	2.81	0.31	0.88	2.00	-0.56	1.13	0.44	0.81	0.44	-2.75	-0.31	0.44	1.50	0.25	0.81
Inconsistency %	30	50	60	40	40	40	30	60	50	20	50	40	50	40	30
Misclassification %	30	10	0	78	50	70	67	60	33	10	40	60	20	43	50

Note: Light shading denotes highest analytic/most consistent/most accurate classification scores while darker shading are top three most impulsive/most inconsistent/most inaccurate classification scores.

Table 4.20: Correlation Coefficients for the Five Measures

	Imp. Strategy %	Rubric Disp. Score	LtA_Difference	Inconsistency %	Misclassification %
Imp. Strategy %	1				
Rubric. Disp. Sc.	0.92	1			
LtA_Difference	0.67	0.69	1		
Inconsistency	0.08	0.11	0.26	1	
Misclassification	0.64	0.52	0.28	-0.18	1

#### 4.3.2 Further Misinterpretation Issues Revealed from Classification Explanations

After participants classified each LtA item, they were asked to explain their decision. Some reoccurring item weaknesses from Part 2 of the interview were substantiated further, and some new survey item issues were also discovered. These issues were summarized in the tables below.

Table 4.21: Issues That Arose from Classification Explanations (Version A)

Item	Participant	Interpretation
A01	S01, S04, S07, S08	“Predict” is an impulsive action or it is telling you to guess
A02	S04, S10, S15	“Either-Or” gives a choice between the two presented approaches only
	S08	Was confused by both approaches stated within item
A04	S08	“Without doing elaborate computations” means knowing answer immediately
A05	S07	An “algorithm” has to work
A07	S01, S04	“Without any computations” means instantly knowing or having an answer
A08	S01, S04	“Obtain answer almost instantly” means no writing and you might make a mistake
A10	S01	“Analyze the quantities” was interpreted as analyzing terms of an algebraic equation
	S08	“Analyze the quantities” is asking if irrelevant quantities should be analyzed

Table 4.22: Issues That Arose from Classification Explanations (Version B)

Item	Participant	Interpretation
B03	S06, S12	“Predict” is an estimate or rounded answer and not performing an action is impulsive
B05	S14	Does not know what Arithmetic expression means, thinks algebraic equation
B09	S14	Not aware that the slope is a numerical value

Version B of the classification exercise had 16 instances of misinterpretation compared with only four instances for Version A. This does not necessarily mean that one set of items is better than the other; it could also just be that interviewees working on Version B were possibly more willing to share their thoughts than those working on Version A.

A new issue concerning grammatical structure arose with A02. Several students expressed a sense of confusion seemingly brought on the Either-Or conjunction contained within the item. Some respondents thought they were required to choose one of the two impulsive acts described within the item due to this.

Other similar issues to the ones in Part 2 of the interview were seen in Part 3 as well, such as misinterpretations of “predict” and not knowing the meaning of certain mathematical terms (e.g., “algorithm” and “arithmetic expression”). For example, “quantities” is taken as a generic term by S01 to refer to terms in an algebraic equation. Interestingly, S01 also interpreted “terms of an algebraic equation” in ReDo-4 as the *terms in a word problem*. This shows how some respondents with weak math vocabularies (and quite possibly weak content knowledge) may interpret different mathematical terms.

## **Chapter 5: Discussion**

This section discusses the results in relation to the three research questions: (R1) How can problem solving disposition be assessed in an interview setting? (R2) To what extent does the LtA survey assess impulsive disposition in respondents? (R3) What are some of the weaknesses of the LtA survey items?

### **5.1 ASSESSING PROBLEM SOLVING DISPOSITION IN AN INTERVIEW SETTING**

As the results from the pilot study (Lim & Mendoza, 2010) suggested that the examination of the solution strategies on participants' written work is a viable way of assessing problem solving disposition. In this study, the results of the strategy-driven method of assessment were found to be highly correlated with the results from the research scoring team. Nevertheless, it is important to note that the research team found that through the use of a rubric they were able to detect more subtle analytic acts within impulsive strategies which may have gone unnoticed in the strategy-method, especially for symbolic items. This means survey creators will have to decide whether or not to include multi-step procedural items in subsequent revisions of the LtA survey if they deem these subtle analytic acts to be important for overall disposition assessment.

Survey creators will have a similar decision to make with problems that lend themselves to the application of multiple strategies. Refinement of these items will need to include more than one strategy per item if they hope to detect a change from an impulsive approach to an analytic one, or vice versa.

Due to the disagreement for certain Rubric Disposition Scores, the research team was able to discern different dispositions within seemingly impulsive behavior and find that even when an "impulsive strategy" is applied, there may an analytic act involved which arguably improved the overall accuracy of the Rubric Disposition Scores. The strategy-count method appears to have been a more biased method of scoring problem-solving disposition since multi-strategy problems weighed more heavily than problems where less strategies were used.

## **5.2 THE LTA SURVEY AND ASSESSMENT OF IMPULSIVE-ANALYTIC DISPOSITION**

### **5.2.1 Correlation among Part 1 Measures**

All of the pair wise correlations in Table 4.4 between the LtA\_Difference scores, Impulsive Strategy-Count Percentages, and the Rubric Disposition Scores indicated at least a moderately high association between the different pairs of measures. This stage of analysis suggests that impulsive disposition could be gauged through a survey (assuming the survey is valid) and shows the viability of quantifying such behavior through the rubric-scoring method in order to corroborate problem solving actions assessed by the LtA scores. Also, the correlation between the LtA responses and the Rubric Disposition Scores for only the problems adapted directly from the LtA survey (P1, P2, P4, and P5) was strong which may be interpreted as a sign that the instrument is headed in a feasible direction.

### **5.2.2 Item-Wise Comparison of Transformed Rubric Scores with LtA Survey**

This part of the analysis was performed more to determine to what extent the Rubric Disposition Scores may have been gauged by the classroom LtA responses. Based on the Fisher's Exact tests, there seems to be an association between students' problem-solving responses as quantified by the Rubric Disposition Score and students' likelihood to act impulsively or analytically to the mathematical situations. Nevertheless, it is important to note the distinction between students' actual problem-solving behaviors and their responses to LtA survey items. The problem solving mode of assessment of impulsive-analytic disposition, although more accurate, is extremely expensive, time consuming, and not practical for large-scale implementation.

### **5.2.3 Consistent Items between Classroom and Interview Settings**

Two of the most consistent items were the impulsive versions of the symbolic items (ReDo-2 and ReDo-6). The question remains whether these two items are good items because not only were they reliable from one setting to another, they also received highly impulsive responses in the ReDo survey which matched what respondents received in the Rubric Disposition Scores as well. Or are these potential

items that need to be validated further in case they are measuring the familiarity of applying a very common procedure internalized by students? And if it turns out that they are indeed testing the familiarity of applying a common procedure to a procedural problem, what does that say about using symbolic items in general to assess impulsive disposition? These types of items offer the opportunity for future research studies that can help to determine if they are reliably valid, reliably invalid, or a combination of both.

#### **5.2.4 Classroom Setting Versus Interview Setting**

Given the potential and portability of a valid LtA instrument for mass scale data collection, survey creators will need to decide if some reliable items are assessing someone as impulsive because they may be acting impulsively in a setting that lends itself to intuitive responses or if they are being impulsive problem solvers because they are not reading survey items with the attention to detail one may provide in an interview/testing environment. If survey-taking setting becomes a variable in how respondents respond to LtA items, is it fairer to say that the survey is instead measuring how well participants pay attention in a particular setting? This is an interesting distinction to make for future LtA item studies. How can the interview setting be controlled to where a respondent can be assessed to be impulsive and not have to wonder if it is the usual way they solve a problem?

#### **5.2.5 Survey Item Type and Inconsistency Bias**

The inconsistency bias found between types of survey items and the number of inconsistent response is an interesting find for further survey validity study because it lets us know that survey item type so far does have an effect on survey results. Survey creators could use this information to decide whether current survey types work well together, or if some should be removed from the LtA instrument, or if they should attempt to add a new type of item such as a graphic item. The idea of adding items with graphical aids may help researchers determine which subjects are analytic and impulsive without worrying whether the seemingly analytic picture the subject creates is due to a habit that results in a picture that does nothing for their conceptual understanding of the problem situation. Such an item could be similar

to the lamppost problem but given after the lamppost problem is solved. Then researcher can observe if students who ignored their picture earlier will still ignore a picture that is given to them as part of the problem situation. This could lead to new way of isolating habits from impulsive-analytic disposition assessment in observed problem solving.

### **5.3 LTA ITEM WEAKNESSES**

The following are some LtA item weaknesses uncovered as part of my qualitative findings that can be used to revise and refine the LtA instrument with the goal of improving validity among the items.

#### **5.3.1 Study and Notice the Solution**

Survey creators may need to revise the phrase “study and notice the solution” as some respondents interpreted the phrase to mean *recognize the procedure needed and begin to combine like terms*. This phrase could potentially cause respondents to answer analytically when they should be responding more on the impulsive side of the scale.

#### **5.3.2 Analyze the Terms in the Equation / Analyze the Quantities**

The phrase “analyze the terms in the equation” was misinterpreted to mean various things. This phrase might be better placed with an algebraic equation so that respondents immediately relate the word “terms” within the phrase with the actual terms of an algebraic equation instead of interpreting “terms” to mean technical words found in mathematical textbooks.

Similarly, “analyze the quantities” was interpreted in different ways. Interestingly, this phrase was paired with the phrase “when asked to solve a word problem” and a respondent still interpreted “quantities” as terms of an algebraic equation, which may mean the interviewee had a weak command of mathematical terms since she had also interpreted “terms of the equation” as studying a word problem. Fortunately, the effect on her responses to both items seems non-existent since the way she seemingly interpreted the items would assess her disposition likelihood in the same manner.

### 5.3.3 Compare Rates (e.g., comparing 30/8 to 20/3)

This example of misinterpretation shows that providing extraneous information such as an example within an item sometimes does not help respondents interpret the item correctly. Finding an alternative to the concept of “comparing rates” may help improve this item. Some respondents thought that comparing rates simply meant to compare the quantities from each project and not to compare the ratios obtained from dividing the quantities of workers and hours for each project (an impulsive strategy). The fractions that appear at the end of the item were also misunderstood to represent the analytic strategy of comparing the quantities from each project and being able to notice the answer or know how what to multiply to find the answer.

### 5.3.4 Predict

The word “predict” was probably the most misinterpreted part of an item since it appeared a few times throughout the survey. The general consensus among some respondents was that that it referred to guessing, estimating or rounding answers, giving inexact answers, or being restricted to mental computation. Replacing this word with a less vague word will improve the validity of those items by avoiding inaccurate responses from those respondents who expressed an aversion to “predict”.

### 5.3.5 Either–Or Statement

Grammatical structure within an item was also observed as a possible detriment to an interviewee’s understanding of an item. Some respondents reported that the *Either-Or* statement contained within A02 was troublesome for them. Keyes (1963) talks about types of mental language and how it is interpreted in the brain, and he states that verbal maps such as Either-Or statements are mentally restraining because nothing in life is all or nothing so an Either-Or statement can be problematic for the brain to process. In this case the Either-Or statement may restrict some students from approaching the item with an open mind by limiting them to an understanding that they must accept one of the two impulsive acts described in the

item. Some respondents did report that they felt the item was asking them to choose one of the strategies (which were both impulsive strategies).

### **5.3.6 Without (Elaborate) Computations / Obtain the Answer Almost Instantly**

Some respondents thought that “without computations” and “obtain an answer almost immediately” meant that you should immediately know an answer similar to expert level, others saw it as mental computation that excludes writing of any sort which may lead to a computation mistake. These phrases are good to point out since there does seem to be respondents who feel anxiety from the thought of strict mental computation. Respondents who feel this way and interpret these phrases as mental computation would probably answer impulsively whether or not they acted so. Survey creators should consider revising these phrases as a way to avoid causing anxiety in people who need to write their thoughts down, whether it be merely a comforting habit, or a way to analyze problem data.

### **5.3.7 Algorithm / Arithmetic Expression**

*Algorithm* and *arithmetic expression* are two mathematical terms that some respondents did not understand. These types of findings will inform LtA revisions for new items and/or item refinements. Keeping in mind that some math terms may be problematic will help keep LtA items more consistent and comprehensible. A respondent with a lack of mathematical vocabulary may consequently possess weaker content knowledge compared to other respondents, but that does not necessarily mean that they are more impulsive. Avoiding words that affect their command of an item will help improve the validity for those respondents.

### **5.3.8 Revision and Refinement of Items**

Some respondents seem to have personal reactions to certain words within the survey. Sometimes their perception of a word or phrase is enough to affect their Likert response which may result in an imprecise response. These findings will inform the creators that one way to improve validity is to keep

the math concepts within the items as simple as possible so that survey item focus remains on a respondent's desired action, instead of on their flawed understanding of an item.

As general Likert item creation literature suggests, the author agrees that every effort should be made to keep the items as short and clearly stated as possible while making sure to avoid extraneous explication (e.g. without doing any computation) and conjunctions (e.g. Either-Or). These efforts should include brainstorming of possible new items and revising language structure as well as pilot testing these ideas informally with different respondents before adding them to the pool of survey items.

#### **5.4 CONCLUDING REMARKS**

The main point of this study was to offer a qualitative approach for the purpose of investigating the validity of the LtA survey instrument. Qualitative research helps us understand the phenomena, and uncover hidden issues that cannot be revealed through strictly quantitative methods. Through the analysis of interviews, qualitative research allowed us to choose a representative sample from the LtA respondent pool and find out what each one thinks about problem solving, the survey items, and the constructs of the study. The qualitative approach is quite important in this study since we are dealing with newly formed constructs (although the phenomena is not new) and a yet to be validated instrument that aspires to measure impulsive-analytic disposition using math-based Likert items which is quite an innovative undertaking in the field of mathematics education.

For this study, mixed methods were used among the different parts of the interview to quantify the findings and better relate them to other parts of the study. In the first part of the results, we were able to assess impulsive disposition in two ways: counting the impulsive strategies used in solving math problems, and creating a rubric to guide a research team that scored all observable disposition cues of interviewees (using the same problems). The two measures were highly correlated with each other, and both were moderately correlated with the original LtA responses which showed the viability of such a method for validation research.

The second part of the results included an item by item comparison to complement the measures of interviewee disposition comparisons that took place in the first part. A 2 x 2 contingency table was created and tested with Fisher's Exact test to uncover whether or not any relationship between LtA responses and Rubric Disposition Scores. Less than two-thirds of the LtA responses actually matched up with the rubric scores assigned to each item in the comparison. The p-value of 0.0139 revealed that there was a strong association between the two measures.

The third set of results informed us of what items appear to have a reliability problem and why that may be the case according to interviewee explanations of inconsistent responses. Evidence of a survey-taking setting variable was uncovered where some interviewees seemed to lack focus in the classroom and tended to be more analytic in the interview setting, possibly because they had increased focus in the interview setting and because they solved the problems within the LtA items in the first part of the interview. Incidence of inconsistent responses was found to have a strong association with the mathematical type of survey item. The majority of story item responses were inconsistent and the responses tended to be more analytic in the interview setting. Interestingly, high-response analytic symbolic items seemed to have reliability issues as their interview LtA responses tended to lean more impulsive on the Likert scale.

The fourth and fifth sets of results were combined into overall occurrences of item misinterpretation. Qualitative research revealed that some respondents had problems self-reporting a precise measure of their disposition because of misinterpreted words and phrases that made some of them give a reverse response of what they meant to reply. Words like "predict", "algorithm", "terms", "rate" were interpreted wrongly by some, while the meaning of some phrases like "study and notice the solution" were altered by some interviewees to refer to an impulsive approach to a problem instead of an analytic one.

These findings are valuable for the revision of problematic LtA survey items and the refinement of those items that appear to have observable aptitude for the assessment of impulsive disposition.

## References

- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42, 116-131.
- Clason, D. L., & Dormody, T. J. (1994). Analyzing data measured by individual Likert-type items. *Journal of Agricultural Education*, 35(4), 31-35.
- Clement, J. (2000). Analysis of clinical interviews: Foundations and model viability. In A. Kelly & R. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 547-589). Mahwah, NJ: Lawrence Erlbaum Associates.
- Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, 7(3), 249-253.
- Cramer, K., Post, T., & Currier, S. (1993). *Learning and teaching ratio and proportion: Research implications*. In D. T. Owens (Ed.), *Research ideas for the classroom: Middle grades mathematics* (pp. 159-178). NY: Macmillan.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Cuoco, A., Goldenberg, E. P., & Marks, J. (1996). Habits of mind: an ongoing principle for mathematics curricula. *Journal of Mathematical Behavior*, 15, 375-402.
- De Gruijter, D. N. M. & Van der Kamp, L. J. T. (2005). *Statistical test theory for education and psychology*. Graduate School of Education, University of Leiden, Netherlands.
- Evans, J. St. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Science*, 7, 454-459.
- Evans, J. St. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review*, 13(3), 378-395.
- Evans, J. St. B. T., & Frankish, K. (2009). *The duality of mind: An historical perspective*. In J. St. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 1-35). Oxford; New York: Oxford University Press.
- Goldin, G. A. (1998). Observing mathematical problem solving through task-based interviews. In A. Teppo (Ed.), *Qualitative research methods in mathematics education* (pp. 40-62). Reston, VA: National Council of Teachers of Mathematics.
- Griffiee, D. T. (1999, January). Questionnaire construction and classroom research. *The language teacher*. Retrieved from Jalt Publications Website:  
<http://www.jalt-publications.org/tlt/articles/1999/01/griffiee>
- Harel, G. (2001, Fall). PUPA's two complementary products: Taxonomy of students' existing proof schemes and DNR-based instruction. *International Newsletter on the Teaching and Learning of Mathematical Proof*, 1-11. Retrieved from  
<http://www-leibniz.imag.fr/DIDACTIQUE/preuve/Resumes/Harel/Harel01.pdf>
- Hendricks, C. (2009). *Improving schools through action research: A comprehensive guide for educators* (2<sup>nd</sup> ed.). Upper Saddle River, NJ: Pearson.
- Hesse-Biber, S. N., & Leavy, P. (2011). *The practice of qualitative research* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Sage.

- International Society for Research on Impulsivity. (2000-2010). Barratt impulsive scale. Retrieved January 6, 2010, from Impulsivity Website:  
<http://impulsivity.org/BIS-11/history-and-development-of-the-bis>
- Kagan, J., Rosman, B. L., Day, D., Albert, J., & Phillips, W. (1964). Information processing in child: Significance of analytic and reflective attitudes. *Psychological Monographs*, 78(1, Whole No. 578).
- Kahneman, D. (Nobel Prize Lecture, 2002). Maps of bounded rationality: A perspective on intuitive judgment and choice. In T. Frangsmyr (Ed.), *Les Prix Nobel* (pp. 416-499). Retrieved from Nobel Website:  
<http://www.nobel.se/economics/laureates/2002/kahnemann-lecture.pdf>.
- Keyes Jr., K. S. (1963). How to develop your thinking ability: A guide to sound decisions. New York, NY: McGraw-Hill.
- Leron, U., & Hazzan, O. (2006). The rationality debate: Application of cognitive psychology to mathematics education. *Educational Studies in Mathematics*, 62, 105-126.
- Lim, K. H. (2006). Students' mental acts of anticipating in solving problems involving algebraic inequalities and equations. *The Selected Works of Kien Lim*.
- Lim, K. H. (2008a). Advancing U.S. pre-service teachers' mathematical sophistication via mathematical tasks. *EduMath*, 27, 2-24.
- Lim, K. H. (2008b). *Students' mental acts of anticipating: Foreseeing and predicting while solving problems involving algebraic inequalities and equations*. Germany: VDM Verlag.
- Lim, K. H., Morera, O., & Tchoshanov, M. (2009a). Assessing problem-solving dispositions: Likelihood-to-act survey. *Proceedings of the 31st Annual Meeting of the Psychology of Mathematics Education-North American Chapter* (pp. 700-708). Atlanta.
- Lim, K. H., & Selden, A. (2009b). Mathematical habits of mind. In S. L. Swars, Stinson, D. W., & Lemons-Smith, S. (Eds.), *Proceedings of the Thirty-first Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 1576-1583). Atlanta: Georgia State University.
- Lim, K. H., & Mendoza, M. (2010). Inferring impulsive-analytic disposition from written responses. In P. Brosnan, D. B. Erchick, & L. Flevares (Eds.), *Proceedings of the 32<sup>nd</sup> Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 759-767). Columbus, OH.
- McCoach, D. B., & Siegle, D. (2003). The school attitude assessment survey-revised: A new instrument to identify academically able students who underachieve. *Educational and Psychological Measurement*, 63(3), 414-429.
- McDonald, R. P. (1999). Test theory: A unified treatment. Mahwah, N.J.: Erlbaum Associates.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749.
- Miley, W., & Spinella, M. (2003). Impulsivity and academic achievement in college students. *College Student Journal*. Retrieved November 18, 2009, from AccessMyLibraryWebsite:

<http://www.accessmylibrary.com/article-1G1-112720418/impulsivity-and-academic-achievement.html>

- Ockert, D. (2005). Substantive scale verification: A Likert scale analysis and critique of university student pedagogical activity preferences. *JALT Hokkaido Journal*, 9, 48-64.
- Papazova, E. B., & Pencheva, E. S. (2006, January). Psychological type and the regulation of the individual behavior. *East & West: A Multicultural Research Conference*, Honolulu, HI. Retrieved November 12, 2009, from Type and Culture Website: [http://typeandculture.org/Pages/C\\_papers06/PenchevaType&Behavior.pdf](http://typeandculture.org/Pages/C_papers06/PenchevaType&Behavior.pdf)
- Russell, J., Tastle, W. J., & Wierman, M. J. (2008). A new measure to analyze student performance using the Likert scale. *Information Systems Education Journal*, 6(35), 1-9.
- Schoenfeld, A. H. (1985). *Mathematical problem solving*. Orlando, FL: Harcourt Brace Jovanovich.
- Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334-370). New York: Macmillan.
- Sloman, S. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin* 119, 3-22.
- Stanovich, K. E. (1999). *Who is rational?: Studies of individual differences in reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Stanovich, K. E. (2009). *Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory?* In J. St. B. T. Evans & K. Frankish (Eds.), *In two minds: dual processes and beyond* (pp. 55-88). Oxford; New York: Oxford University Press.
- Stanovich, K. E. (2011). *Rationality and the reflective mind*. New York, NY: Oxford University Press.
- Stavy, R. and Tirosh, D. (2000). *How Students (Mis-)Understand Science and Mathematics: Intuitive Rules*. Teachers College Press.
- Trochim, W. M. K. (2006, October 20). Likert scaling. *Research methods knowledge base*. Retrieved November 14, 2009, from Social Research Methods Website: <http://www.socialresearchmethods.net/kb/scallik.php>
- Watson, A. & Mason, J. (2007). Taken-as-shared: A review of common assumptions about mathematical tasks in teacher education. *Journal of Mathematics Teacher Education*, 10, 205-215. [Campbell, W. G. 1990. *Form and Style in Thesis Writing, a Manual of Style*. Chicago: The University of Chicago Press.]

## Appendix A - LtA Survey

### Likelihood-to-Act Survey

7	6				
---	---	--	--	--	--

For each item, indicate as honestly as you can how likely you are to act in the manner specified in the statement by circling the number using this scale:

- 1 = Extremely Unlikely
- 2 = Unlikely
- 3 = Somewhat Unlikely
- 4 = Somewhat Likely
- 5 = Likely
- 6 = Extremely Likely

Start Time:

:

- |  |   |   |   |   |   |   |
|--|---|---|---|---|---|---|
| 1. $78 + 987x + 654 + 321x = x + 987x + 654 + 321x$<br>When asked to solve for $x$ , how likely are you to begin by studying the equation and noticing the solution?   | 1 | 2 | 3 | 4 | 5 | 6 |
| 2. Given that 4 candy bars cost \$2.15.<br>When asked to find the cost of 40 candy bars, how likely are you to begin by EITHER finding the cost per candy bar OR setting up a proportion?  | 1 | 2 | 3 | 4 | 5 | 6 |
| 3. $\frac{3}{4} + \frac{1}{10} + \frac{9}{10}$<br>When asked to find the answer for the above arithmetic expression without using a calculator, how likely are you to begin by finding the common denominator?   | 1 | 2 | 3 | 4 | 5 | 6 |
| 4. When solving a problem in mathematics, how likely are you to read and understand the problem thoroughly before deciding what to do?   | 1 | 2 | 3 | 4 | 5 | 6 |
| 5. Given that a line that passes through two points: (17, 17) and (43, 43).<br>When asked to find the slope of the line, how likely are you to use the slope formula $m = \frac{y_2 - y_1}{x_2 - x_1}$ ?   | 1 | 2 | 3 | 4 | 5 | 6 |
| 6. Paula is cycling from home to school.<br>At 8 o'clock she has already cycled 2.4 miles.<br>When asked to find her cycling speed, how likely are you to analyze the problem situation instead of dividing 2.4 by 8?  | 1 | 2 | 3 | 4 | 5 | 6 |
| 7. $\frac{4}{5} \times \frac{7}{9} + \frac{1}{5} \times \frac{7}{9}$<br>When asked to find the answer for the above arithmetic expression without using a calculator, how likely are you to use the order of operations (i.e., multiply first and then add)? | 1 | 2 | 3 | 4 | 5 | 6 |
| 8. When asked to solve an algebraic equation, how likely are you to analyze the terms in the equation before you start solving it?   | 1 | 2 | 3 | 4 | 5 | 6 |
| 9. $(x - 7)(x - 4) = 0$<br>When asked to solve for $x$ , how likely are you to study the equation and predict the solution?  | 1 | 2 | 3 | 4 | 5 | 6 |

## Appendix A - LtA continued

- |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|
| 10. A 40 gram box costs \$7.00. A 21 gram box costs \$3.30.<br>When asked to decide which box is a better buy, how likely are you to choose your answer without having to perform any computations?   | 1 | 2 | 3 | 4 | 5 | 6 |
| 11. $8044 \times 0.25$<br>When asked to find the answer without using a calculator, how likely are you to multiply 8044 by 25 and then move the decimal point by two places to the left?  | 1 | 2 | 3 | 4 | 5 | 6 |
| 12. When asked to find the value of an arithmetic expression, how likely are you to study the numbers in the expression to see if you can find a quicker way to get the answer and avoid elaborate computations?  | 1 | 2 | 3 | 4 | 5 | 6 |
| 13. $\frac{3x+3}{x+1} = \frac{x}{4}$<br>When asked to find a value of $x$ that satisfies the above equation, how likely are you to cross-multiply and then solve $4(3x+3) = x(x+1)$ for $x$ ?   | 1 | 2 | 3 | 4 | 5 | 6 |
| 14. Project P took 30 workers, each working 8 hours, to complete.<br>Project Q took 20 workers, each working 3 hours, to complete.<br>When asked to determine which project was bigger in size, how likely are you to compare rates (e.g., comparing 30/8 to 20/3)? | 1 | 2 | 3 | 4 | 5 | 6 |
| 15. $1545.9 + 694.8 - 545.9$<br>When asked to find the answer without using a calculator, how likely are you to study the decimals and obtain the answer almost instantly?  | 1 | 2 | 3 | 4 | 5 | 6 |
| 16. When asked to solve a word problem, how likely are you to begin by looking for key words to help you decide which formula or which operation to use?  | 1 | 2 | 3 | 4 | 5 | 6 |
| 17. $90 + 1234n + 567 + 89n = n + 1234n + 567 + 89n$<br>When asked to solve for $n$ , how likely are you to begin by combining like terms?  | 1 | 2 | 3 | 4 | 5 | 6 |
| 18. Given that 6 bottles of mineral water cost \$2.10.<br>When asked to find the cost of 30 bottles, how likely are you to notice the relationship between 30 and 6 and use it to obtain the answer?  | 1 | 2 | 3 | 4 | 5 | 6 |
| 19. $\frac{3}{5} + \frac{11}{12} + \frac{1}{12}$<br>When asked to find the answer for the above arithmetic expression without using a calculator, how likely are you to begin by studying the fractions to see if you can predict the answer?                       | 1 | 2 | 3 | 4 | 5 | 6 |
| 20. When solving a problem in mathematics, how likely are you to use the first idea that comes to mind?   | 1 | 2 | 3 | 4 | 5 | 6 |
| 21. Given that a line that passes through two points: (35, 35) and (82, 82).<br>When asked to find the slope of the line, how likely are you to predict the value of the slope without performing any computation?  | 1 | 2 | 3 | 4 | 5 | 6 |

## Appendix A - LtA continued

- |   |                       |
|---|-----------------------|
| 22. Jimmy is walking from home to school and at 7 o'clock he has already walked 1.4 km. When asked to find his rate of walking, how likely are you to use the $d = rt$ or $r = \frac{d}{t}$ relationship and obtain 0.2 km/hour?  | 1   2   3   4   5   6 |
| 23. $\frac{3}{8} \times \frac{1}{7} + \frac{3}{8} \times \frac{6}{7}$<br>When asked to find the answer for the above arithmetic expression without using a calculator, how likely are you to study the fractions and obtain the answer without doing any elaborate computations?            | 1   2   3   4   5   6 |
| 24. When asked to solve an algebraic equation, how likely are you to begin by using a formula or a rule?  | 1   2   3   4   5   6 |
| 25. $(x - 5)(x - 8) = 0$<br>When asked to solve for $x$ , how likely are you to multiply out the terms (i.e., FOIL) and then solve $x^2 - 13x + 40 = 0$ using the quadratic formula?  | 1   2   3   4   5   6 |
| 26. A 1.5 pound cheese costs \$4.00. A 2.9 pound cheese costs \$8.50. When asked to decide which chunk of cheese is a better buy, how likely are you to compute and compare their unit cost (i.e., $4/1.5$ versus $8.5/2.9$ )?  | 1   2   3   4   5   6 |
| 27. $0.25 \times 4804$<br>When asked to find the answer without using a calculator, how likely are you to predict the answer without actually multiplying the numbers?  | 1   2   3   4   5   6 |
| 28. When asked to find the value of an arithmetic expression, how likely are you to immediately use an algorithm that you think will work?  | 1   2   3   4   5   6 |
| 29. $\frac{6x+2}{3x+1} = \frac{x}{5}$<br>When asked to find a value of $x$ that satisfies the above equation, how likely are you to study the equation and predict that the answer is 10?   | 1   2   3   4   5   6 |
| 30. Project ABC took 26 workers, each working 10 hours, to complete. Project XYZ took 18 workers, each working 7 hours, to complete. To determine which project was bigger in size, how likely are you to visualize the two scenarios and predict the answer without doing any computation? | 1   2   3   4   5   6 |
| 31. $3875.4 + 367.9 - 875.4$<br>When asked to find the answer without using a calculator, how likely are you to begin by adding 3875.4 and 367.9 and then subtract 875.4?   | 1   2   3   4   5   6 |
| 32. When asked to solve a word problem, how likely are you to analyze the quantities and understand how those quantities are related?   | 1   2   3   4   5   6 |

Have you taken this survey before? Yes / No

If yes, when did you take it? \_\_\_\_\_

End Time:

:

## Appendix A - Need for Cognition

7	6				
---	---	--	--	--	--

Please indicate as honestly as you can to what extent each statement is characteristic of you.  
Please use the following scale:

- 1 = extremely uncharacteristic of you (not at all like you)  
2 = somewhat uncharacteristic  
3 = uncertain  
4 = somewhat characteristic  
5 = extremely characteristic of you (very much like you)

- |   |   |   |   |   |   |
|---|---|---|---|---|---|
| 1. I would prefer complex to simple problems.   | 1 | 2 | 3 | 4 | 5 |
| 2. I like to have the responsibility of handling a situation that requires a lot of thinking.   | 1 | 2 | 3 | 4 | 5 |
| 3. Thinking is not my idea of fun.  | 1 | 2 | 3 | 4 | 5 |
| 4. I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.                   | 1 | 2 | 3 | 4 | 5 |
| 5. I try to anticipate and avoid situations where there is likely a chance I will have to think in depth about something.                     | 1 | 2 | 3 | 4 | 5 |
| 6. I find satisfaction in deliberating hard and for long hours.   | 1 | 2 | 3 | 4 | 5 |
| 7. I only think as hard as I have to.   | 1 | 2 | 3 | 4 | 5 |
| 8. I prefer to think about small, daily projects to long-term ones.   | 1 | 2 | 3 | 4 | 5 |
| 9. I like tasks that require little thought once I've learned them.   | 1 | 2 | 3 | 4 | 5 |
| 10. The idea of relying on thought to make my way to the top appeals to me.   | 1 | 2 | 3 | 4 | 5 |
| 11. I really enjoy a task that involves coming up with new solutions to problems.   | 1 | 2 | 3 | 4 | 5 |
| 12. Learning new ways to think doesn't excite me very much.   | 1 | 2 | 3 | 4 | 5 |
| 13. I prefer my life to be filled with puzzles that I must solve.   | 1 | 2 | 3 | 4 | 5 |
| 14. The notion of thinking abstractly is appealing to me.   | 1 | 2 | 3 | 4 | 5 |
| 15. I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought. | 1 | 2 | 3 | 4 | 5 |
| 16. I feel relief rather than satisfaction after completing a task that required a lot of mental effort.                                      | 1 | 2 | 3 | 4 | 5 |
| 17. It's enough for me that something gets the job done; I don't care how or why it works.  | 1 | 2 | 3 | 4 | 5 |
| 18. I usually end up deliberating about issues even when they do not affect me personally.  | 1 | 2 | 3 | 4 | 5 |

## Appendix A - BIS-11

7	6				
---	---	--	--	--	--

DIRECTIONS: People differ in the ways they act and think in different situations. This is a test to measure some of the ways in which you act and think. Read each statement and put an X on the appropriate circle on the right side of this page. Do not spend too much time on any statement. Answer quickly and honestly.

		①	②	③	④
		Rarely/Never	Occasionally	Often	Almost Always/Always
1	I plan tasks carefully.	①	②	③	④
2	I do things without thinking.	①	②	③	④
3	I make-up my mind quickly.	①	②	③	④
4	I am happy-go-lucky.	①	②	③	④
5	I don't "pay attention."	①	②	③	④
6	I have "racing" thoughts.	①	②	③	④
7	I plan trips well ahead of time.	①	②	③	④
8	I am self controlled.	①	②	③	④
9	I concentrate easily.	①	②	③	④
10	I save regularly.	①	②	③	④
11	I "squirm" at plays or lectures.	①	②	③	④
12	I am a careful thinker.	①	②	③	④
13	I plan for job security.	①	②	③	④
14	I say things without thinking.	①	②	③	④
15	I like to think about complex problems.	①	②	③	④
16	I change jobs.	①	②	③	④
17	I act "on impulse."	①	②	③	④
18	I get easily bored when solving thought problems.	①	②	③	④
19	I act on the spur of the moment.	①	②	③	④
20	I am a steady thinker.	①	②	③	④
21	I change residences.	①	②	③	④
22	I buy things on impulse.	①	②	③	④
23	I can only think about one thing at a time.	①	②	③	④
24	I change hobbies.	①	②	③	④
25	I spend or charge more than I earn.	①	②	③	④
26	I often have extraneous thoughts when thinking.	①	②	③	④
27	I am more interested in the present than the future.	①	②	③	④
28	I am restless at the theater or lectures.	①	②	③	④
29	I like puzzles.	①	②	③	④
30	I am future oriented.	①	②	③	④

## Appendix B - Open-Response Problems

*Five Open-Response Problems from Part I of Interview:*

- P1. Find the answer for  $\frac{3}{4} + \frac{1}{10} + \frac{9}{10}$
- P2. Solve for  $n$ :  $90 + 1234n + 567 + 89n = n + 1234n + 567 + 89n$
- P3. Five lampposts are spaced evenly along a street. The distance between the first lamppost and the last lamppost is 220 m apart. What is the distance between any two neighboring lampposts?
- P4. Project P took 30 workers, each working 8 hours, to complete. Project Q took 20 workers, each working 3 hours, to complete. Which project was bigger in size?
- P5. Paula is bicycling from home to school. At 8 o'clock she has already cycled 2.4 miles. What is her speed?

## **Appendix C - Satisfaction Placard**

*Placard A.* A placard explaining a satisfaction scale from one to five:

5 = Very Satisfied

4 = Satisfied

3 = Neutral

2 = Unsatisfied

1 = Very Unsatisfied

## **Appendix D - Definition Placard**

*Placard B.* A placard showing the working definitions for “impulsive disposition” and “analytic disposition”:

Impulsive disposition refers to a tendency to spontaneously apply an idea that comes to mind without checking its relevance or appropriateness.

Analytic disposition refers to a tendency to understand and analyze a problem situation before performing actions to solve the problem.

## Appendix E - Interview Questions

### *Interview Questions:*

#### **Part 1**

1. Solve this problem. Say out what you are thinking as you solve it.  
Use the blank space below for your written work.
2. How confident are you that your answer is correct?
3. Now I like you to share with us the way you solved those problems.
  - a. On a scale of 1 to 5, how satisfied are you with the way you solve this problem?  
Why?
  - b. What was the first idea that came to your mind as you tried to solve this problem?
  - c. What other ideas, if any, did you think of?
4. In general, how would you characterize the way you approach a math problem?

#### **Part 2**

1. I'm going to get you to take this 10-item survey, which is a short version of the one you did in class.  
This time we need you to say out loud as you read the item and decide which number to circle. Is that OK?
2. (For each discrepant item)  
You circled \_\_\_\_ for this question. Can you explain why you circled \_\_\_\_?  
Interestingly, for the same item you circled \_\_\_\_ in class. Why?

#### **Part 3**

1. (General questions)
  - a. Did you learn something new about yourself from this interview?
  - b. Do you consider yourself as someone who spontaneously uses the first idea that comes to mind to solve a problem?
  - c. When it comes to solving math problems, do you think you are more impulsive or less impulsive than your classmates?
2. Let me introduce you to these two terms: impulsive disposition and analytic disposition.  
Your job is to use these two terms to classify the survey items.
  - a. Do you think the boldface act described in this question is impulsive or analytic?  
Are you confident of your answer?
  - b. Why are you not confident? Or why do you think the act is impulsive or analytic?

## Appendix F - Interview 10-item LtA Survey

1.  $78 + 987x + 654 + 321x = x + 987x + 654 + 321x$   
When asked to solve for  $x$ , how likely are you to begin by studying the equation and noticing the solution?
2.  $\frac{3}{4} + \frac{1}{10} + \frac{9}{10}$   
When asked to find the answer for the above arithmetic expression without using a calculator, how likely are you to begin by finding the common denominator?
3. Paula is cycling from home to school.  
At 8 o'clock she has already cycled 2.4 miles.  
When asked to find her cycling speed, how likely are you to analyze the problem situation instead of dividing 2.4 by 8?
4. When asked to solve an algebraic equation, how likely are you to analyze the terms in the equation before you start solving it?
5. Project P took 30 workers, each working 8 hours, to complete.  
Project Q took 20 workers, each working 3 hours, to complete.  
When asked to determine which project was bigger in size, how likely are you to compare rates (e.g., comparing 30/8 to 20/3)?
6.  $90 + 1234n + 567 + 89n = n + 1234n + 567 + 89n$   
When asked to solve for  $n$ , how likely are you to begin by combining like terms?
7.  $\frac{3}{5} + \frac{11}{12} + \frac{1}{12}$   
When asked to find the answer for the above arithmetic expression without using a calculator, how likely are you to begin by studying the fractions to see if you can predict the answer?
8. Jimmy is walking from home to school and at 7 o'clock he has already walked 1.4 km. When asked to find his rate of walking, how likely are you to use the  $d = rt$  or  $r = \frac{d}{t}$  relationship and obtain 0.2 km/hour?
9. When asked to solve an algebraic equation, how likely are you to begin by using a formula or a rule?
10. Project ABC took 26 workers, each working 10 hours, to complete.  
Project XYZ took 18 workers, each working 7 hours, to complete.  
To determine which project was bigger in size, how likely are you to visualize the two scenarios and predict the answer without doing any computation?

## **Appendix G – Disposition Code Definitions**

**NOREADWELL** – Did not read item well

**NORELAX** – Claims to be more relaxed during interview than in class

**RUSHED** – Participant felt rushed when completing survey in class

**NORECALL** – Participants cannot recall why he/she circled a certain response

**SOLVECHANGE** – Participant’s response changed from having solved the problem in Part 1

**WKCONTENT** – Participant has a problem understanding an item due to weak content knowledge

**INCONSISTENT** – Participant seems to be inconsistent in approach/strategy for no specific reason

**ASSUMEDSIMPLICITY** – Participant assumes they will be able to solve a problem by the act described in the survey item, but after a closer look they decide they would need to take another route

**NOTSERIOUS** – Participant states he/she was not serious or is observed to not take an item response seriously

**INFLUENCED** – Participant’s response is influenced by the quantities or wording within the survey item, this can influence from the act described in the item, or by irrelevant numbers within the item

**MISREAD** – Participant reads the survey item in a way that changes what the item was meant to ask, usually a careless mistake from reading too fast or not carefully enough

**LANGUAGE** – Participant interprets a word or words of a survey item not as intended

**EXPERIENCE** – A previous experience related to math problem solving affects a participant’s response, such as a recent class lesson or homework experience

## Appendix H – Classification Items, Version A (#1-5)

- Q5. When asked to find the value of an arithmetic expression, how likely are you to **immediately use an algorithm that you think will work?**

Q4.  $\frac{3}{8} \times \frac{1}{7} + \frac{3}{8} \times \frac{6}{7}$

When asked to find the answer for the above arithmetic expression without using a calculator, how likely are you to **study the fractions and obtain the answer without doing any elaborate computations?**

Q3.  $8044 \times 0.25$

When asked to find the answer without using a calculator, how likely are you to **multiply 8044 by 25 and then move the decimal point by two places to the left?**

Q2. 4 candy bars cost \$2.15.

When asked to find the cost of 40 candy bars, how likely are you to **begin by EITHER finding the cost per candy bar OR setting up a proportion?**

Q1.  $(x - 7)(x - 4) = 0$

When asked to solve for  $x$ , how likely are you to **study the equation and predict the solution?**

## Appendix H – Classification Items, Version A (#6-10)

- Q10. When asked to solve a word problem, how likely are you to **analyze the quantities and understand how those quantities are related?**

- Q9. Given that a line that passes through two points: (17, 17) and (43, 43). When asked to find the slope of the line, how likely are you to **use the slope formula  $m = \frac{y_2 - y_1}{x_2 - x_1}$ ?**

- Q8.  $1545.9 + 694.8 - 545.9$   
When asked to find the answer without using a calculator, how likely are you to **study the decimals and obtain the answer almost instantly?**

- Q7. A 40 gram box costs \$7.00. A 21 gram box costs \$3.30.  
When asked to decide which box is a better buy, how likely are you to **choose your answer without having to perform any computations?**

- Q6.  $\frac{3x+3}{x+1} = \frac{x}{4}$   
When asked to find a value of  $x$  that satisfies the above equation, how likely are you to **cross-multiply and then solve  $4(3x + 3) = x(x + 1)$  for  $x$ ?**

## Appendix I – Classification Items, Version B (#1-5)

- P5. When asked to find the value of an arithmetic expression, how likely are you to **study the numbers in the expression to see if you can find a quicker way to get the answer and avoid elaborate computations?**

P4.  $\frac{4}{5} \times \frac{7}{9} + \frac{1}{5} \times \frac{7}{9}$

When asked to find the answer for the above arithmetic expression without using a calculator, how likely are you to **use the order of operations (i.e., multiply first and then add)?**

P3.  $0.25 \times 4804$

When asked to find the answer without using a calculator, how likely are you to **predict the answer without actually multiplying the numbers?**

P2. 6 bottles of mineral water cost \$2.10.

When asked to find the cost of 30 bottles, how likely are you to **notice the relationship between 30 and 6 and use it to obtain the answer?**

P1.  $(x - 5)(x - 8) = 0$

When asked to solve for  $x$ , how likely are you to **multiply out the terms (i.e., FOIL) and then solve  $x^2 - 13x + 40 = 0$  using the quadratic formula?**

## Appendix I – Classification Items, Version B (#6-10)

P10. When asked to solve a word problem, how likely are you to **begin by looking for key words to help you decide which formula or which operation to use?**

P9. Given that a line that passes through two points: (35, 35) and (82, 82). When asked to find the slope of the line, how likely are you to **predict the value of the slope without performing any computation?**

P8.  $3875.4 + 367.9 - 875.4$   
When asked to find the answer without using a calculator, how likely are you to **begin by adding 3875.4 and 367.9 and then subtract 875.4?**

P7. A 1.5 pound cheese costs \$4.00. A 2.9 pound cheese costs \$8.50. When asked to decide which chunk of cheese is a better buy, how likely are you to **compute and compare their unit cost (i.e., 4/1.5 versus 8.5/2.9)?**

P6.  $\frac{6x+2}{3x+1} = \frac{x}{5}$   
When asked to find a value of  $x$  that satisfies the above equation, how likely are you to **study the equation and predict that the answer is 10?**

## **Vita**

Miguel Mendoza has a Bachelor of Arts in Mathematics with a minor in Communication Studies from the University of Texas at El Paso. He has co-presented and presented two studies related to the assessment of problem solving disposition at the PME-NA Conferences of 2010 and 2012. He was born and raised in El Paso, TX and currently resides there. He may be contacted by e-mail at [suncityguy@gmail.com](mailto:suncityguy@gmail.com).