12-2018

Identifying Opportunities that Promote a Habit of Mind in the Quantitative Reasoning Classroom.

David Lavie Deville
University of Arkansas, Fayetteville

Identifying Opportunities that Promote a Habit of Mind
in the Quantitative Reasoning Classroom

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Mathematics

by

David Lavie Deville
University of Arkansas
Bachelor of Science in Mathematics, 2011
University of Arkansas
Master of Science in Mathematics, 2014

December 2018
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council

____
Bernard Madison, Ph.D.
Dissertation Director

____
Laura Kent, Ph.D.
Committee Member

____
Shannon Dingman, Ph.D.
Committee Member

____
Jeffrey Hovermill, Ph.D.
Committee Member

ABSTRACT

Quantitative literacy and its role in a democratic society are emerging topics of concern to mathematics educators throughout the world. Undergraduate courses in Quantitative Reasoning (QR) are a recent addition to the growing list of resources aimed at remedying quantitative illiteracy in America. This research aimed to illustrate the design and differences in the three courses in question, identify course materials that promote proficiency in QR, identify course materials that motivate students and/or encourage a habit of mind (HoM), investigate the interaction between student abilities in QR and their HoM, and investigate the relationships between context in curricular materials, student engagement, and the promotion of a HoM among students. Students were interviewed after completing three different undergraduate QR courses. A qualitative analysis of the interviews as well as course texts identified opportunities and hindrances for students in developing a HoM, how students apply a HoM, and obstacles students faced in developing a HoM.

ACKNOWLEDGEMENTS

I am especially grateful to my advisor, Bernard L. Madison. Dr. Madison supported a creative process on my part, while keeping steady and timely contact with me at all times. Without his guidance, I may not have seen this project through to the end. I would also like to thank my committee members: Shannon W. Dingman, Laura B. Kent and Jeffrey A. Hovermill. Not only were they supportive, but their requests helped me produce a more polished and impressive thesis. Further, the office staff in the UARK Department of Mathematical Sciences was an indispensable resource, and I would like to thank Mary Powers, Dorine Bower, Cait McCormack and Tami Trzeciak for all of their hard work.

I could not ask for better parents, Charles and Mary Deville, Jr. Their understanding and support developed who I am today. Additionally, I would like to thank my mother's sister, Billie Lane. She provided motivation and funding at a critical moment during this project. Last, I would like to thank my wife, Sarah Watson, for her support and trust since we first met in our freshman year at the University of Arkansas.

CONTENTS

CHAPTER 1: INTRODUCTION

Quantitative reasoning (QR), quantitative literacy (QL), numeracy, and innumeracy are emerging topics of concern to mathematics educators throughout the world. Evidence of this growing trend is found in many sources including the formation of the National Numeracy Network (NNN), Mathematical Association of America (MAA) reports, special interest groups, and works edited and authored by Madison and Steen (MAA, 1994; Madison & Steen, 2003, 2008; Steen, 1997, 2001A, 2004A;). Undergraduate courses in QR are a recent addition to the growing list of resources aimed at remedying innumeracy in America. This research focuses on three such courses: MAT 114 at Northern Arizona University (NAU), MATH 1313 at University of Arkansas' Fayetteville campus (UA), and MATH 2183 at UA. This introduction provides an overview of the conceptualizations of QL, descriptions of the courses under study, a statement of the research problem and definitions of terms used throughout the study.

Numeracy, QR, and QL are often used interchangeably in scientific literature, and there exists a diverse body of literature and opinions about their respective definitions. Conceptions of numeracy are described by Maguire and O'Donoghue (2002) in three phases that represent an "ever-increasing sophistication in the conceptualization of numeracy" (p. 154). The first, formative phase began with *numeracy* defined "as the mirror image of literacy" in the Crowther Report (1959) (O'Donoghue, 2002, p. 47). Numeracy, as defined in the Crowther Report (1959) includes an ability to reason scientifically and problem solve, much like *literacy* implies an ability to go beyond basic reading and writing skills:

> On the one hand is an understanding of the scientific approach to the study of phenomena – observation, hypothesis, experiment, verification. On the other hand is a need in the modern world to think quantitatively, to realise how far our problems are problems of degree even when they appear as problems of kind (Crowther, 1959, as cited in O'Donoghue, 2002, p. 53).

However, in “Evolution of Numeracy and the National Numeracy Network” Madison and Steen (2008) report that “this expectation of sophisticated problem solving” became lost in the years following the Crowther Report and “numeracy gradually came to mean just simple arithmetic skills normally acquired in childhood” (p. 2).

The 1982 Cockcroft Report, which “sought to revive [numeracy’s] original meaning”, marks the ushering in of numeracy’s mathematical phase (Madison & Steen, 2008, p. 3). This expanded view of numeracy places it in more of a mathematical context with an emphasis on “an ‘at-homeness’ with numbers and an ability to make use of mathematical skills which enable an individual to cope with the practical mathematical demands of his everyday life” and “[the] ability to have some appreciation and understanding of information which is presented in mathematical terms” (O’Donoghue, 2002, p. 53). Conceptions of numeracy in the mathematical phase elevate the concept above basic arithmetic, but are marked by “curriculum driven… …one size fits all” approaches with “a strong link with the curriculum of school mathematics” (Maguire & O’Donoghue, 2002, p. 155). The higher end of the mathematical phase recognizes the importance of individuality in that mathematics must be relevant to individuals and useful in their everyday lives. While this phase points to the utility of numeracy skills in everyday life, it still falls short of a clear description of how to transfer mathematical skills into ability and informed decision making in authentic and complex societal, economic, and professional contexts.

According to Maguire and O’Donoghue (2002), the largest jump in concept sophistication occurs between the mathematical and integrative phases of numeracy. The integrative phase of numeracy moves away from the previous math-centric conceptions, recognizing “the integration of mathematics, communication, cultural, social, emotional and personal aspects of each individual's numeracy in context" (Maguire & O'Donoghue, 2002, p. 156). Instead of viewing numeracy in the context of mathematics and mathematical curricula, educators in the integrative phase recognize the importance of authentic context in numeracy; students are viewed as individuals with specialized needs specific to their lives, work, and/or field of study. Practice with decontextualized mathematics is not considered adequate for increasing numeracy and what constitutes a numerate individual is specific to his or her discipline, function in society, and wants or needs in life. Whereas the upper end of the mathematical phase incorporates the need for individualized mathematics, the integrative phase recognizes that numeracy requires the marriage of mathematics with the context in which it is relevant. The integrative phase considers mathematics not only as supportive of "communicati[ve], cultural, social, emotional and personal" dimensions, but also considers mathematics as supported by the levels of context in which it is relevant (Maguire & O'Donoghue, 2002, p. 156).

The three courses relevant to this study all share a common goal, to promote QR in a population of primarily non-STEM students as a part of their university core requirements. This study aims in part to illustrate that these courses respectively represent differing conceptualizations of QR as described by Maguire and O'Donoghue (2002). The primary investigator (PI) has experience teaching each course and understands that although all of the courses share a common goal, the courses at UA differ greatly from MAT 114 at NAU. Additionally, the habits of mind (HoM) to seek out and understand quantitative information expressed in media and other avenues relevant to everyday life exhibited by the population of students who enroll in these courses as well as the courses' propensity to foster a HoM among students is of particular interest to the PI. Unfortunately, measuring a HoM is not easy and there are no standard assessments in place for doing so; to the knowledge of the PI, an instrument developed by Stuart Boersma and Dominic Klyve at Central Washington University (CWU) is the only instrument intended to measure HoM. However, no research exists on the efficacy of the instrument and the measure, HoM, is poorly understood. Fortunately, experts in the field seem to agree on a few facets of QR: core competencies exhibited by quantitatively literate individuals - described by the Association of American Colleges and Universities (AAC&U) – and the importance of real world context in the application of QR curricula.

This research aims to illustrate the design and differences in the three courses in question, identify course materials that promote proficiency in QR, identify course materials that motivate students and/or encourage a HoM, investigate the interaction between student abilities in QR and their HoM (as measured by the CWU instrument), and investigate the relationships between context in curricular materials, student engagement, and the promotion of a HoM among students.

COURSE DESCRIPTIONS

MAT 114 at NAU is named Quantitative Reasoning. MAT 114 is a modified emporium style course with elements of a flipped course design. Sections of up to 36 students meet one day of the week for 75 minutes. Students are also required to spend 75 minutes per week in the tutor occupied open laboratory (lab), where they work on homework from the week's material or material intended to prepare the students for upcoming material the following week. The course text, *Quantitative Reasoning*, is a workbook developed in house by Matthew Fahy and Gina M. Nabours. It consists of 13 modules that involve "contemporary quantitative methods, especially descriptive statistics, elementary probability, examples of statistical inference, linear and exponential models of growth and decay, and applicable discrete models" (Northern Arizona University, 2018). Around 1000 students enroll in MAT 114 each semester, and the course employs graduate students, part time faculty, full time faculty, and sometimes tenured faculty in its instruction in order to meet the demands of enrollment. Each section has a single instructor.

During class meetings, instructors primarily teach through lecture, and the students fill in blanks in their workbooks during class time. Many MAT 114 instructors choose to use PowerPoint presentations from a shared Dropbox folder that progress linearly through the text. When time allows, students are encouraged during class to attempt examples in the text either on their own or in groups. In the experiences of the PI, roughly 20%-40% of class time is allotted for focused individual or group work, with the remainder of class consisting of either lecture or class discussion. Course instructors meet weekly to discuss upcoming material, administrative items, and pedagogical concerns. Typically, instructors are encouraged to elicit answers/reasoning from students and avoid simply showing students how to complete problems in an effort to promote student engagement during lectures. There is no official mandate on the structure of class time; however, the PI believes that most classes begin with up to five minutes of administrative announcements, followed by alternations of 15 – 30 minute spans of lecture/class discussion and 5 – 15 minute spans of individual/group work with the possibility of 5 – 10 minute spans of student presentations of worked problems. This cycle (besides the administrative announcements) repeats until time is up, when there will usually be a 10 minute weekly quiz (a perforated page from the students' workbooks) and another few minutes of administrative reminders.

Lab time may be completed any time the lab is open during the week. For the Fall 2017 semester, the lab hours are from 9:00 a.m. to 9:00 p.m. on Monday and Wednesday, 9:30 a.m. to 9:00 p.m. on Tuesday and Thursday, 9:00 a.m. to 5:00 p.m. on Friday, 1:00 p.m. to 5:00 p.m. on select Saturdays, and 1:00 p.m. to 9:00 p.m. on Sunday. Students primarily use this time to work on homework/projects and to get ahead on upcoming material. Course modules correlate with each week of the regular semester, and all but 2 of the 13 course modules include a section titled "Looking Forward" that explores problems related to upcoming material, which are intended to be completed outside of regular class meetings. In the lab, students can receive help from tutors who are undergraduates employed as "Math Jacks" and instructors of Lumberjack Mathematics Center (LMC) courses. Since MAT 114 is an LMC course, the students may see their instructor(s) in the open lab; however this is not guaranteed. Further, there are online resources for the course as well as online practice problems the students are required to complete weekly. Students may use open lab time to complete their weekly online assignments; however, this is not required and the problems may be solved at home as well. Online resources are available for each module in the course text and utilize Great River Learning technologies.

Four semester tests are given outside of class with predetermined locations/times. Sections are not intermixed when they test – students test alongside their regular classmates. Tests do not vary between sections and are written by course coordinators. Students are allowed the use of a scientific calculator during examinations and select equations are provided. Tests are paper-based and typically consist of both multiple choice and short answer problem types. A two hour final exam occurs during the final exam time per university schedule.

MATH 1313 is also called Quantitative Reasoning. Sections of up to around 30 students meet two to three times a week for a total of 150 minutes per week. The course text, *Common*

Sense Mathematics, is a MAA Press textbook authored by Ethan Bolker and Maura Mast. The text focuses on understanding through common sense approaches to problems that are always situated in a real-world context; it is integrative in the sense that all problems come from a real-world context, including data sets. However, the text does not usually include full articles. Approximately 100 students enroll in the course during each semester, and the course employs graduate students, part time faculty, full time faculty, and sometimes tenured faculty in its instruction in order to meet the relatively low demands of enrollment. Each section has a single instructor.

MATH 1313 is loosely coordinated and the structure of the classes varies between sections. From conversations with course instructors, the PI believes that sections of MATH 1313 can consist of anywhere from 70% individual/group work or student presentations and 30% lecture/class discussion to 40% individual/group work or student presentations and 60% lecture/class discussion. Course instructors meet several times throughout the semester to discuss upcoming material or projects, pedagogical concerns, and administrative items. These meetings tend to be informal and discussions drift between topics naturally. There are not many instructors for the course in a given semester and discussion may be more focused on individual experiences in teaching the course. Instructors are typically encouraged to include active learning techniques in the classroom and promote involvement from students as much as possible. Also, during these meetings instructors are reminded that course instruction should not focus on the mathematical content/techniques involved in course materials, but rather emphasize understanding of contextual situations, critical analysis of media sources, and the variety of potential approaches in answering questions from the text.

The variety of implementations of MATH 1313 makes it difficult to describe a typical class. Also, some instructors utilize aspects of a flipped course design while others do not. From interviews with course instructors, the PI believes that several practices are common to most, if not all, sections of MATH 1313. Course topics are generally introduced with a short (15-30 minute) lesson on requisite mathematical skills and this type of lesson occurs weekly or every other week as needed. Some instructors report this lesson as lecture based, while others give group work problems without lecture in the hopes that students will collectively rediscover the mathematical techniques they have surely seen in prior courses. It is likely that some instructors use a combination of these tactics, and instructors utilizing a flipped course design will assign this type of lesson outside of class. Instructors report class discussion nearly always follows group, individual, or at-home work and that this discussion is essential in keeping students on track and current with course skills. Several instructors noted they implemented lessons on requisite skills only after bad classroom experiences in sections where requisite skills lessons were not included. Typically, requisite skills lessons are included in all sections where the instructor has at least one semester of experience teaching MATH 1313. Problems from the course text follow skill work and class discussion about contextual topics related to these problems is used as needed. For example, if a question involves the use of credit cards, then key aspects such as annual percentage rates and late fees may be discussed prior to specific problems from the text. Students work alone or in groups and may be asked to explain their answers to other students or groups of students. This can take up a lot of time in class and doubly so since this type of work is typically followed by a full class discussion. Further, many instructors report that student presentations on problems worked in class or in the homework fill a significant portion of class time.

Homework draws mostly from the course text and usually focuses on materials already covered in class; however, some instructors report that homework often looks ahead toward upcoming material. Instructors typically assign homework between every class and check for it at the beginning of class. Whether or not this homework is fully graded depends on both the instructor and the level of motivation at that point in the semester. Course instructors report they use quizzes on an as needed basis, usually every two to three weeks; instructors typically create their own quizzes. Two to three course projects occur throughout the semester in each section. The stated intention of the projects is for the students to design research questions like problems from the course text; however, some instructors report deviations from this structure. Some instructors ask the students to report in two to three pages on a topic of their choice that involves quantities and reasoning related to course discussions. The course includes only two semester tests given in class and a two-hour final exam scheduled by the university.

MATH 2183 is titled Mathematical Reasoning. Sections of up to around 30 students meet two to three times a week for a total of 150 minutes per week. The course text, *Case Studies for Quantitative Reasoning: A Casebook of Media Articles*, is a Pearson Learning Solutions textbook authored by Stuart Boersma, Caren L. Diefenderfer, Shannon W. Dingman and Bernard L. Madison. The text “provides a tool for educational response to the enormous QR demands that US residents face. It is the foundation for developing and delivering an ever-fresh, real-world-based course that starts or moves students forward on a path toward quantitative literacy (QL)." (Boersma et al., 2012). This text consists of case studies of media articles with warm-up questions intended to prepare students for the mathematical demands of the study questions that relate specifically to a media article presented in the textbook. Approximately 100 students enroll in the course during each semester, and the course employs graduate students, part time faculty, full time faculty, and sometimes tenured faculty in its instruction in order to meet the relatively low demands of enrollment. Each section has a single instructor.

MATH 2183, while loosely coordinated with course structures varying from section to section, appears more tightly coordinated than MATH 1313. From conversations with course instructors, the PI believes that sections of 2183 can consist of anywhere from 70% individual/group work or student presentations and 30% lecture/class discussion to 40% individual/group work or student presentations and 60% lecture/class discussion. Instructors are strongly encouraged to utilize group work, with the expectation that students, having already taken MATH 1313 or a college algebra course, exhibit a level of sophistication slightly above that of the typical MATH 1313 student. Course instructors meet several times throughout the semester and these meetings are likely very similar to the meetings with MATH 1313 instructors; meetings may be more or less frequent depending on the semester. Some 2183 instructors utilize aspects of a flipped course design while others do not.

From interviews with course instructors, the PI believes that several practices are common to most, if not all, sections of MATH 2183. In Spring 2017 sections of 2183 most instructors began class with a 10 to 20 minute news of the day activity. In this activity, students present on an article of their choice. This includes a summary of the article and a discussion of relevant mathematical content; the activity intends to spur student engagement and provide an avenue for discussion of mathematical and journalistic content. Warm-up questions from the textbook are usually assigned as homework to prepare students to work in class on the more involved and in-depth case study questions. Typically some time is allotted to discuss the warm-up questions and any difficulties students had in answering them out of class on their own. Some instructors spend time lecturing on the mathematical content of the warm-up questions while others do not. The majority of class time is spent with students working in groups on the case studies. Instructors follow up on case study group work by either discussing/lecturing about answers to study questions or allowing groups to present solutions to study questions.

Homework draws mostly from the course text and focuses on both materials already covered in class as well as upcoming material. For example, instructors reported frequently assigning warm-up questions for upcoming material as homework, but also noted that students were asked to complete study questions as homework given insufficient time to complete these during class. Instructors typically assign homework between every class and check for it or discuss it near the beginning of class (generally after news of the day presentations). Whether or not this homework is fully graded depends on both the instructor and the level of motivation at that point in the semester. Course instructors report they use quizzes on an as-needed basis, some more frequently than others (every 2-4 weeks); instructors typically create their own quizzes. Typically, two projects occur throughout the semester in each section. The stated intention of the projects is for the students to design case studies like in the course text (typically about a media article of the student's choice); these are more in depth than the projects in MATH 1313 since they involve development of multiple study questions relating to a media article along with a set of warm-up questions focused solely on mathematical content. The course includes only two semester tests given in class and a two-hour final exam scheduled by the university.

STATEMENT OF THE PROBLEM

The following research questions are investigated in this study.

1. What are similarities and differences between opportunities to develop habits of mind in three different QR courses?

   a. Curriculum and design

      i. Core competencies

      ii. Conceptualization of QR

2. What habits of mind do QR students demonstrate?

   a. What similarities and differences in habits of mind exist within and between students?

      i. Application

      ii. Obstacles

DEFINITION OF TERMS

The following terms are used throughout the study.

Analysis/Synthesis: "Ability to make and draw conclusions based on quantitative analysis" (Boersma et al., 2011, p. 5).

Assumptions: "Ability to make and evaluate important assumptions in estimation, modeling, and data analysis" (AAC&U, 2009).

Calculation: "Ability to perform arithmetical and mathematical calculations" (Boersma et al., 2011, p. 5).

Communication: "Ability to explain thoughts and processes in terms of what evidence is used, how it is organized, presented, and contetualized" (Boersma et al., 2011, p. 5).

Core competencies: From AAC&U's QL VALUE Rubric, they consist of interpretation, representation, calculation, application/analysis (adapted as analysis/synthesis in the QLAR), assumptions, and communication.

Habit of Mind (HoM): The habit of mind to seek out and understand quantitative information expressed in media and other avenues relevant to everyday life.

Interpretation: "Ability to glean and explain mathematical information presented in various forms (e.g. equations, graphs, diagrams, tables, words)" (Boersma et al., 2011, p. 5).

Quantitative Literacy (QL): "the power and habit of mind to search out quantitative information, critique it, reflect upon it, and apply it in their public, personal and professional lives" (Madison & Steen, 2008, p. 6).

QL Assessment Rubric (QLAR): "is an adaptation of the AAC&U VALUE QL rubric to make it more applicable to grading student work… …The QLAR is intended to measure achievement levels of the associated QL core competencies in a variety of assignments" (Boersma et al., 2011, p. 4).

Representation: "Ability to convert information from one mathematical form (e.g. equations, graphs, diagrams, tables, words) into another" (Boersma et al., 2011, p. 5).

CHAPTER 2: REVIEW OF THE LITERATURE

INTRODUCTION

QL educational literature is relatively underdeveloped, and much of it is purely theoretic in nature. This review outlines the expert opinions and beliefs that influenced the courses under study, and follows a development of QL conceptualization from mathematical to increasingly integrative. This review also includes studies and questions raised that are relevant to this research, and ends with discussion of the two primary instruments used in this research study.

LITERATURE REVIEW

A 1994 Mathematical Association of America (MAA) report aimed at giving "insight into the development of more-recent quantitative literacy/numeracy" focused on the question: "(w)hat quantitative literacy requirements should be established for all students who receive a bachelor's degree?" (MAA, 1994, p. 1). A quantitative literacy requirement subcommittee was formed in 1989, and, through discussions and investigations, the subcommittee reached four primary conclusions:

3. Colleges and universities should treat quantitative literacy as a thoroughly legitimate and even necessary goal for baccalaureate graduates.
4. Colleges and universities should expect every college graduate to be able to apply simple mathematical methods to the solution of real-world problems.
5. Colleges and universities should devise and establish quantitative literacy programs each consisting of foundation experience and a continuation experience, and mathematics departments should provide leadership in the development of such programs.
6. Colleges and universities should accept responsibility for overseeing their quantitative literacy programs through regular assessments. (MAA, 1994, pp. 1-2).

The report also "enumerat[ed] some of the principal reasons for expecting quantitative literacy of educated people" (MAA, 1994, p. 6). Notably, several important points were raised:

> One of the classic reasons for studying mathematics is that it strengthens general reasoning powers, for instance by developing problem-solving skills. While the research literature is ambiguous on this point, many thoughtful people are convinced that it is true in some sense.
>
> Increasing amounts of mathematics are needed in an increasing number of careers… ... And students, even college seniors often do not know what careers they will enter, or where their career paths will lead them. A quantitative literacy requirement helps to hold some doors open.
>
> Many adults, and especially college graduates, are very likely to assume positions in their communities and in professional organizations where quantitative literacy (e.g., the ability to deal intelligently with statistics) will come into play and may even be essential for effectiveness. A quantitative literacy requirement can thus be expected to enhance the quality of citizens.
>
> Anyone who does not have a mature appreciation of mathematics misses out on one of the finest and most important accomplishments of the human race. A quantitative literacy requirement, sensibly defined, will contribute to the spread of that appreciation.
>
> The fear of mathematics that is often called 'math phobia' or 'math anxiety' stunts the cognitive development of those who suffer from it. It is usually learned, not inborn, and a curricular component devoted to promoting quantitative literacy, if competently and compassionately taught, can be powerfully therapeutic against it.
>
> In particular, negative attitudes of parents and teachers (including guidance counselors) toward mathematics are all too easily picked up by the next generation. Statements like 'Oh, I never was good at math myself' or 'Just get this math out of the way and then forget it; you'll never need it again' or 'For punishment, you will have to do thirty extra math problems' can do enormous amounts of mischief (MAA, 1994, p. 6).

The report suggested that any remedial techniques be embedded in "compelling applied contexts" and argued that "(i)f no genuine application of a topic can be found at the appropriate level, omit it" (MAA, 1994, p. 8). The report clarified that "the standard intermediate algebra and college algebra courses are generally not of the nature proposed" and that QL efforts should push beyond simple exposure to "mathematical beauty and power" (MAA, 1994, p. 11). The report emphasized that courses aimed at QL should "place emphasis on students' doing reasoning", "capture student interest", "have genuine application", involve the use of technology, avoid "lecture and listen" and "tackle inappropriate 'beliefs' that students may carry, such as 'to do mathematics is to calculate answers'" (MAA, 1994, p. 12). The report also emphasized the importance of having students communicate mathematics in writing, as well as engage in group and project based assignments in QL courses. The genesis and underpinning philosophy of MAT 114 at NAU, in its current form, is rooted in the guidelines from this report.

MAT 114, before it was titled *Quantitative Reasoning,* was called *College Mathematics with Applications* (NAU, 1991). MAT 114 is described in the 1991-1993 course catalogue as "(c)ontemporary applications of algebra, geometry, statistics, probability and discrete mathematics", and satisfied the "liberal studies foundation requirement" (NAU, 1991, p. 290). Ideas from the MAA's 1994 report on numeracy filtered into the course and it was eventually rebranded as *Quantitative Reasoning* in 1999 (NAU, 1999, p. 628). The course is described today as it originally was in 1999, "(c)ontemporary quantitative methods, especially descriptive statistics, elementary probability, examples of statistical inference, linear and exponential models of growth and decay, and applicable discrete models", and it continues to satisfy the "liberal studies foundation requirement" (NAU, 1999, p. 628). Influences from the MAA report (1994) are seen in the course's broad mathematical curriculum that highlights areas of mathematics applicable to the real world, and the course's use of a fill in the blank style course pack – a response to the MAA's (1994) call to "place emphasis on students' doing reasoning", and to avoid "lecture and listen" (p. 12).

Educators such as Lynn Arthur Steen (1997) began to expand on the MAA's (1994) report and guided efforts in the United States to establish defining features of *quantitative literacy* (QL) and *numeracy*. Steen (1997) called *numeracy* the "new literacy of our age" (p. xv) and identified the term as synonymous with QL. Steen (1997) noted a difficulty in that "(QL) means different things to different people", and took special care to delineate QL from school mathematics, stating that QL "is really about reasoning more than 'rithmetic: assessing claims, detecting fallacies, evaluating risks, weighing evidence" (pp. xvi-xix). Steen (1997) noted that many mathematics educators at the time viewed QL "as an umbrella for important aspects of mathematics", and cited that "John Dossey [identified] these aspects as data representation, numbers and operations, variables and relations, measurement, space and visualization, and chance" (Steen, 1997, p. xxii). Steen (1997) steered toward the alternative view from George Cobb that QL is "four very different kinds of thinking: logical, algorithmic, visual, and verbal", and the idea that QL may be lumped in with the term "problem solving" (p. xxii). Steen (1997) clarified the role of mathematics in real-world applications in the following passage:

> In reality, modern high-performance work involves problems that require sophisticated reasoning and practice, but often only elementary mathematical skills. In contrast, the mathematics that students study to prepare for college requires advanced skills with abstract concepts deployed in simple (and simplistic) problem situations. (Steen, 1997, p. xxiv)

Many educators attempted to solidify a working definition of QL and further delineate it from previous math-centric conceptions. The cognitive psychologist Iddo Gal (1997) described numeracy as

> an aggregation of skills, knowledge, beliefs, dispositions, habits of mind, communication capabilities, and problem-solving skills that people need to autonomously engage in and effectively manage situations in life and at work that involve numbers, quantitative or quantifiable information, or textual information that is based on or has embedded in it some mathematical elements (p. 39).

Gal (1997) described typical mathematics problems for K-12 students as "contrived" or "presented out of context" and contrasted them with "real-life numeracy situations" which "are always embedded in a context that has some personal meaning to the people involved" (p. 39). Gal (1997) also identified the critical thinking and reading comprehension required by real-world problems as a frequent limiting factor for students. Further, he claimed the disconnect between traditional mathematical curricula and the requirements of real-world problems is compounded by the tendency of school mathematics problems to have *right* and *wrong* answers; Gal (1997) identified “numeracy situations” as situations where an optimal solution – if one even exists – is often open to interpretation (Gal, 1997, p. 40). Gal (1997) responds to this incongruity with the recommendation that mathematics educators focus more on numeracy and to help students “not only to acquire a solid foundation in the theory and processes of mathematics and statistics but also to learn to manage numeracy situations effectively and to fully realize the implications of poor management of such situations” (p. 42).

Pollak (1997) identified that “(r)eal-world problem solving involves not only mathematics but also some problematic situation outside of mathematics or some real-world difficulty crying out for systematic understanding” (Pollak, 1997, p. 91). Pollak (1997) notes that problem solving in an authentic context requires an understanding not only of the mathematics involved, but also requires the situation to be well understood enough that solutions are reasonable and/or practical. Pollak illustrates his point with a real-world problem solving case study. The length and detail involved in the case study, spanning almost seven pages of text, points to Pollak’s concept of problem solving. Pollak does not consider problem solving in the sense of the typical mathematics word problem, but as an exercise in dynamic understanding of a situational problem; a problem encountered where the nature of the problem changes with the student’s understanding of both the mathematical and contextual components. He says that a “key characteristic of real-world problem solving… … is that progress is driven by considerations of both the external world and mathematics” (Pollak, 1997, p. 101). Pollak (1997) concludes that “(a) student’s mathematics education is simply not complete if that student has not experienced the usefulness of mathematics in the larger world" and claims that "(t)his experience comes through real-world problem solving" (p. 105).

Steen (2001B) documented several early definitions of terms relating to QL; he quoted the Cockroft's (1982) definition for *numerate*, the National Adult Literacy Survey (NCES, 1993) definition for QL, the International Life Skills Survey (ILSS, 2000) definition for QL and the Programme for International Student Assessment's (PISA, 2000) definition for *mathematics literacy* respectively:

> We would wish the word numerate to imply the possession of two attributes. The first of these is an "at homeness" with numbers and an ability to make use of mathematical skills which enables an individual to cope with the practical demands of everyday life. The second is an ability to have some appreciation and understanding of information which is presented in mathematical terms.
>
> The knowledge and skills required to apply arithmetic operations, either alone or sequentially, using numbers embedded in printed material (e.g., balancing a checkbook, completing an order form).
>
> An aggregate of skills, knowledge, beliefs, dispositions, habits of mind, communication capabilities, and problem solving skills that people need in order to engage effectively in quantitative situations arising in life and work.
>
> An individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded mathematical judgments and to engage in mathematics in ways that meet the needs of that individual's current and future life as a constructive, concerned and reflective citizen. (Steen, 2001B, pp. 6-7)

Steen (2001B) noted that these definitions included varying emphasis from "basic skills" to "higher-order thinking" and some focused on individual "ability to use quantitative tools", while others focused on "the ability to understand and appreciate the role of mathematical and quantitative methods in world affairs" (p. 7). Steen (2001B) used these definitions to inform what he called "elements" of quantitative literacy: "(c)onfidence with (m)athematics", "(c)ultural (a)ppreciation", "(i)nterpreting (d)ata", "(l)ogical (t)hinking", "(m)aking (d)ecisions", "(m)athematics in (c)ontext", "(n)umber (s)ense", "(p)ractical (s)kills", "(p)rerequisite (k)nowledge", and "(s)ymbol (s)ense" (pp. 8-9). Steen (2001B) also wrote about the various ways in which QL is expressed, and claimed, "they provide a rich portrait of numeracy in the modern world" (p. 10). He listed the more sophisticated expressions of QL as "(c)itizenship", "(c)ulture", "(e)ducation", "(p)rofessions", "(p)ersonal Finance", "(p)ersonal (h)ealth", "(m)anagement", and "(w)ork" (In Steen, 2001B, pp. 10-15).

Orrill (2001) described a "liberating literacy" similar to the way the PI defines QL. Orrill (2001) described this as, "assum[ing] a much more challenging standard by which individuals command both the enabling skills needed to search out information and the power of mind necessary to critique it, reflect upon it, and apply it in making decisions" (p. xiv). Orrill (2001) noted that increases in "attention to the traditional mathematics curriculum" (p. xviii) do not necessarily correlate with an increase in QL. Orrill (2001) explained that "unlike mathematics, numeracy does not so much lead upward in an ascending pursuit of abstraction as it moves outward toward an ever richer engagement with life's diverse contexts and situations" (p. xviii). Orrill (2001) claimed that the important questions for QL educators lie not in questions regarding traditional mathematics curricula (e.g. important aspects of algebra, geometry, and calculus, and the appropriate order of introduction of these subjects to students), and stated "there is something missing in these debates when we consider them in light of the quantitative demands of contemporary life" (p. xix).

After The National Council on Education and the Disciplines (NCED) published *Mathematics and Democracy: The Case for Quantitative Literacy,* the Mathematical Sciences Education Board (MSEB) "hosted a national forum on quantitative literacy that was supported by NCED in cooperation with the Mathematical Association of America (MAA)" (Steen, 2004B, pp. xi-xii). *Achieving quantitative literacy: An urgent challenge for higher education,* produced as a part of the MAA Notes Series, "includes papers commissioned as background for the Forum, essays presented at the Forum, and selected reactions to the Forum" (Steen, 2004B, p. xii). David Bressoud (2004) asserted that it is mathematicians' responsibility to "nurture and shape a meaningful program in quantitative literacy" (p. ix) even though mathematicians are relatively ill equipped in the instruction of QL. Lynn Arthur Steen (2004B) reported "several important messages [from the Forum]":

> Quantitative literacy largely determines an individual's capacity to control his or her quality of life and to participate effectively in social decision-making.
>
> Educational policy and practice have fallen behind the rapidly changing data-oriented needs of modern society, and undergraduate education is the appropriate locus of leadership for making the necessary changes.
>
> The wall of ignorance between those who are quantitatively literate and those who are not can threaten democratic culture.
>
> Quantitative literacy is not about "basic skills" but rather, like reading and writing, is a demanding college-level learning expectation that cuts across the entire undergraduate curriculum.
>
> The current calculus-driven high school curriculum is unlikely to produce a quantitatively literate student population (Steen, 2004B, p. xii).

Steen (2004C) argued, "although most mathematical tools required for QL are taught in school, continued reflective experience with data throughout college deepens students' capacity to use these tools for productive lives and responsible citizenship" (p. 3). Steen (2004C) further noted that "college has replaced high school as the educational standard for a democratic society", but that mathematics curricula have stagnated or even narrowed, "forcing almost everyone through the bottleneck of calculus" (p. 5). Steen (2004C) wrote that "(i)n many cases, required mathematics courses actually subvert QL by convincing students that mathematics is not good for anything in the real world"; he noted that many Forum participants argued that "personal success in the new information economy requires a new set of problem-solving and behavioral skills that emphasize the flexible application of reasoning abilities", and that "(t)hese skills involve sophisticated reasoning with elementary mathematics more often than elementary reasoning with sophisticated mathematics" (pp. 9-10).

Steen (2004D) argued that QL is not a remedial goal and it should be fostered in a university education. . Steen (2004E) elaborated on the differences between algorithmic mathematics and QL practices; he said that QL is "about challenging college-level settings in which quantitative analysis is intertwined with political, scientific, historical, or artistic contexts", and that "QL adds a crucial dimension of rigor and thoughtfulness to many of the issues commonly addressed in undergraduate education" (p. 22). Steen (2004E) listed several examples where QL is embedded in authentic and compelling contexts:

> *Political polling.* How can polls be so accurate? Why do they sometimes fail?
> *Clinical trials:* Why is a randomized double blind study the most reliable?
> *Tax policy:* Can lower tax rates yield greater tax revenue?
> *Vaccination strategy:* Ethics of individual vs. societal risks (e.g., smallpox).
> *Investment strategies:* The logic of diversification vs. the psychology of risk.
> *Improving education:* What data are required to infer causation from correlation?
> *Fighting terrorism:* Balancing lives vs. dollars and other incommensurate comparisons.
> *Cancer screening:* Dealing with false positives when disease incidence is low.
> *Building roads:* Why the "tragedy of the commons" often leads to slower traffic.
> *Judging bias:* Dealing with Simpson's paradox in disaggregated data.
> *Clinical trials:* Ethics of using placebos for seriously ill patients (p. 22).

Steen (2004E) explained that most of these items do "not require advanced mathematics" (pp 22-23) but rather provide a capstone to high school level mathematics that may be applied to the real world. He elaborated that "QL is anchored in specific contexts, often presented through 'thick descriptions' with rich and sometimes confusing detail" and noted that contexts are often "personal or political, involving questions of values and preferences", something wholly unfamiliar to a traditional mathematics classroom (Steen, 2004E, p. 24). Further, Steen (2004E) claimed that a "cornerstone of QL is the ability to apply quantitative ideas in *unfamiliar contexts*" and that this "insists on flexible understanding that adapts readily to new circumstances" (p. 24). Steen (2004E) also warned educators that "teaching in context is not at all easy" (p. 25).

Steen (2001B) had earlier provided a list of skills involved in QL with the warning that "skills learned free of context are skills devoid of meaning and utility"; he added that "(t)o be effective, numeracy skills must be taught and learned in settings that are both meaningful and memorable" (pp. 16-17). Steen (2001B) emphasized, "quantitative literacy is inseparable from its context" and contrasted QL with "mathematics, statistics, and most sciences" as a subject that "grows more horizontally than vertically"; while traditional mathematics identifies patterns through abstraction, Steen claimed that "numeracy clings to specifics" (pp. 17-18). Schoenfeld (2001) had also lamented the lack of real world application in common undergraduate mathematics courses, and commented that his "undergraduate courses in probability and statistics dealt largely with probability distributions; the real world was not really present" (Schoenfeld, 2001, p. 50).

Many educators who attended the MSEB forum on QL wrote about the important role of real-world context in QR. Wiggins (2003) explained what it means for a person to understand things in context. He said that when we understand,

> we (c)an explain, make connections, offer good theories… … (c)an interpret… … offer apt translations… … can apply… …adapt what we know in diverse contexts… … (h)ave perspective: See multiple points of view, with critical eyes and ears; see the big picture… …(c)an empathize… … (s)how self-knowledge: Perceive the personal style, prejudices,

> projections, and habits of mind that both shape and impede our own understanding" (Wiggins, 2003, p. 132).

Wiggins (2003) shared his belief that mathematics education is dominated by "a tacit (and false) learning theory" that assumes students must first master basic skills in "the logical order of the elements" and separate "from experiential and historical context" before they can be confronted with contextual, realistic problems. He pointed out that "(i)n no area of human performance is it true that years of drills and facts must precede all attempts to perform." Packer (2003) claimed that "mathematics education is inadequate to today's challenge, [promoting quantitative literacy]" and noted that since "(m)athematics teachers are disconnected from other faculty in many schools and colleges", "mathematics lacks context and other courses lack mathematics" (p. 34). Packer (2003) added that "it is better to build abstract thinking on a concrete base" (p. 34), and appealed to the common scenario where a student has finished a course in mathematics and is still unable to apply learned mathematics to a real-life problem. Packer (2003) believes students should begin with concrete, real-world problems and – from these – develop mathematics in the search of a solution. Packer (2003) offered a set of guidelines for determining the type of mathematics that can actually be useful to the general student population: "(o)ne helpful criterion is to restrict problems to those that American workers get paid to solve, those that American citizens should have informed opinions about, or those that American consumers actually need to solve" (p. 35). Packer (2003) explained that empirical data suggests students are generally unable to apply abstract mathematics in real-world application. He admited that topic coverage must be sacrificed in order to develop models complicated and sophisticated enough to "capture interest" (Packer, 2003, p. 38). Packer (2003) appealed to "(b)rain research" that he claimed "has shown again and again that retention of information requires context" (p. 38).

Deborah Hughes Hallett (2003) offered the following definition for QL: "(q)uantitative literacy is the ability to identify, understand, and use quantitative arguments in every-day contexts" and added, "(a)n essential component is the ability to adapt a quantitative argument from a familiar context to an unfamiliar context" (p. 91). Hallett (2003) made the distinction that "advanced training in mathematics does not necessarily ensure high levels of quantitative literacy" and claimed the "reason is that mathematics courses focus on teaching mathematical concepts and algorithms, but often without attention to context" (p. 92). "(A)lthough mathematics courses teach the mathematical tools that underpin quantitative literacy, they do not necessarily develop the skill and flexibility with context required for quantitative literacy" (Hallett, 2003, p. 92). Hallett (2003) claimed that "(m)athematics courses that concentrate on teaching algorithms, but not on varied applications in context, are unlikely to develop quantitative literacy" (p. 93). Hallett (2003) believes that students can improve their QL through analysis of complex, contextual, and novel situations. She admitted that teaching to this paradigm is "much, much harder than teaching a new algorithm" (Hallett, 2003, p. 93).

Robert Cole (2003) commented on the "narrowness of many people's disciplinary thinking"; in particular, Cole (2003) wrote about the tendency of mathematicians to view applications as a mere "delivery vehicle for the QL or mathematical techniques" (p.252).

> If QL goes down this road – a smorgasbord of techniques squeezed into a general education course – I think we run the danger of not addressing the real need outlined in the case statement in *Mathematics and Democracy: The Case for Quantitative Literacy*, namely, the need of citizens to find a use for mathematics that connects with their perception of the real world (p. 252).

Bernard Madison (2003) argued a "principal cause of the transition problems in U.S. mathematics education is the lack of an intellectually coherent vision of mathematics among professionals responsible for mathematics education" and added that "(m)athematicians similarly lack a coherent vision" (p. 153). He defined Quantitative literacy as "the ability to understand and use numbers and data in everyday life" and argued that the "need for quantitative skills" drives the increasing complexity and importance of mathematics education (Madison, 2003, p. 153). Dan Kennedy (2001) previously wrote that "(e)ven in the most 'reformed' of U.S. classrooms, students are being prepared for a capstone experience of college calculus and for embarrassingly little else" (p. 56). Kennedy added that advancements in technology "have forced many of us to confront directly the questions of what algebra and geometry we ought to be teaching with the aid of technology and what should be taught without it" (p. 58). Madison (2003) expanded on Kennedy's (2001) views, and identified the "geometry, algebra, trigonometry, and calculus (GATC)" (p. 153) sequence as the dominating mathematics curriculum for students as they move through high school and an undergraduate degree program. Madison (2003) noted that "approximately three of four" of these students will not complete the sequence and lamented that they "leave with disappointment (or worse) and fragmented mathematics skills that are not readily useful in their everyday lives" (pp. 153-154). The end result, Madison (2003) argued, is that "this system produces the world's best-educated and most creative scientists and engineers while at the same time yielding a quantitative literacy level that ranks near the bottom among industrialized nations" (Madison, 2003, p. 154; OECD, 2001).

Madison (2003) singled out "(t)he institutionalization of college algebra as a core general education course" (p. 155) as one of the most misguided hurdles in many undergraduate curricula. Madison (2003) identified the "wasteland of remedial courses" as "the most depressing of all" (p. 158) in comparison to other college mathematics courses.

> Remedial mathematics is almost always arithmetic or high school algebra. Consequently, except for returning students who have been away from school for some time, students in remedial courses are repeating material they failed to learn in earlier, possibly multiple,

> efforts. Having to repeat work, not making progress toward a degree, and studying uninspiring – and to students, illogical – subject matter makes remedial mathematics courses unusually dreary… … the proportion of students who are unsuccessful in remedial mathematics courses is often high, in the range of one-half to two-thirds (Madison, 2003, p. 158).

Madison (2003) called for "(t)wo major corrections": "(f)irst, the rigid linearity of the route to advanced mathematics must be abandoned" and "(s)econd, college mathematics courses must have independent value and not be only routes to somewhere else" (p. 162). Madison (2003) suggested that

> (b)y focusing introductory college mathematics courses on learning by using, especially learning by using technology, these courses can extend school mathematics at the same time they fill in gaps in learning. We can stop the treadmill of repeated failures in repetitious courses. We can stop telling students that they will need algebra later, perhaps in calculus and its applications. Instead, we can show students why algebra is important and what they need to master (p. 163).

The views espoused by Madison (2003) during and after the MSEB forum mark the beginning of a focused QR initiative at the University of Arkansas (UA). Pilot versions of MATH 2183 began in Fall 2004. The experimental course was originally a section of finite mathematics and focused on quantitative information found in newspaper articles (Dingman & Madison, 2010A, p. 2). Educators continued to grapple with the relationship between mathematics and QR, and the initiative at UA evolved with the views of QR educators.

Madison and Steen (2008) edited eight papers focused on the goal of producing "a list of questions that institutions might address to audit their programs of teacher education and QL education"; these questions were the focus of the June 22, 2007 Wingspread workshop (Madison, 2008, p. 4). Steen (2008) noted an unresolved problem "heard in many discussions at the Wingspread workshop" was whether or not QL is a part of mathematics and "(i)f so, why isn't it taught and learned? If not, who should teach it" (p. 13)? Steen (2008) also detailed a paradigm shift among “QL explorers” (p. 13). He said that “QL explorers have moved beyond debates about the definition of QL, not because they reached consensus but because they recognize that development of QL programs is more important” (Steen, 2008, p. 13). Steen (2008) clarified a distinction between mathematics and QL through the analogy that “mathematics is to QL as template problem solving is to authentic decision making” and used this analogy to illustrate a challenge to mathematics teachers in promoting QL among their students; mathematics teachers are often drawn to their profession “because they like to follow rules, and are most comfortable with the precision and definitiveness of a good mathematics problem” (p. 18). Steen (2008) argued that “mathematics teachers will need to encourage argument and discussion” in order “(t)o help their students become quantitatively *literate*” (p. 18). Steen (2008) summarized some of the questions raised during the workshop in the following excerpt:

> Should QL be part of a college’s mathematics requirement or organized across the curriculum with ‘Q’ courses in many departments? Might it be integrated into Comp 101 as part of every freshman’s initial exposure to college writing? Do students in non-quantitative tracks need QL, or do their current requirements suffice? What should be done for college students who do not know what fractions are or mean? ... The list of questions is endless, more than enough to fill the agenda of the next numeracy workshop (p. 23).

Richard J. Shavelson (2008) of Stanford University said that “the situated approach seems to capture current thinking about QR” and claimed that “QR is evidenced when confronted with a well contextualized, messy, open-ended, ‘real-world’ task that demands analysis, critical thinking, problem solving and the capacity to communicate a solution, decision, or course of action clearly in writing” (p. 37). He also claimed that “no single correct answer but a variety of possible answers that vary in their credibility and evidentiary base” marks many QR problems (Shavelson, 2008, p. 23). Shavelson (2008) also shifted his focus from teacher preparation to "the preparation of students generally" and asked if

> we should be talking about preparing QR in introductory college mathematics courses for the broad college audience, in general education courses, and in the mathematics major creating a pedagogy that gives the diversity of students access to both QR and the level of mathematics needed to teach in high school (p. 43).

Neil Lutsky (2008) of Carleton College proposed, "(a) fitting context for quantitative reasoning is argumentation, the construction, communication, and evaluation of arguments" (p. 59). Further, he claimed that QR could help students "construct and evaluate arguments… … because quantitative reasoning can contribute to the framing, articulation, testing, principled presentation, and public analysis of arguments" (Lutsky, 2008, p. 60). To this end, Lutsky (2008) claimed that the QR habits relevant to students "are primarily simple and non-technical" and that "(t)he teaching of quantitative reasoning across the curriculum might not only model itself on the teaching of writing across the curriculum; it might be intertwined with teaching writing" (p. 60). Lutsky (2008) called for educators to attend to why numbers are prevalent and important in everyday life and to "show others that numbers can contribute to precision in our thinking" (p. 61). Lutsky (2008) believes that QR is strongly situated in the formulation of arguments and that argumentation is inherent in most academic pursuits:

> Teaching students how to identify and find the constituent elements of an argument, how to organize arguments systematically, what kinds of statements support particular arguments effectively, how to present arguments clearly and meaningfully to an audience, how to address their own arguments reflectively, and how to evaluate others' arguments *are* fundamental to education at all levels and in almost all disciplines…. …Among other things, quantitative information may be used to help articulate or clarify an argument, frame or draw attention to an argument, make a descriptive argument, or support, qualify, or evaluate an argument" (p. 63).

Lutsky (2008) offered 10 “framing questions” he believes could help guide students in effectively applying QR in argumentation:

1. What do the numbers show?
2. How typical is that?
3. Compared to what?
4. Are findings those of a single study or source or of multiple studies or sources?
5. How were the main characteristics measured?
6. Who or what was studied?
7. Is the outcome of a study anything more than noise or chance?
8. How large is the result of a study?
9. What was the design of the study?
10. What else might be influencing the findings? (pp. 67-68)

Lutsky’s (2008) 10 questions are intended to “encourage [students] to seek relevant numbers, both when they argue and when they evaluate the arguments of others” (p. 69). Lutsky (2008) concluded that “writing assignments that invite or require quantitative reasoning” are an “essential way teachers can facilitate quantitative reasoning” (p. 69).

Milo Schield (2008) “explore[d] the possibility of delaying, minimizing, or eliminating the manipulation of common fractions as mathematical objects and of replacing it with a more applied study of fractions in the context of percentages and rates” (pp. 87-88). Though Schield’s (2008) paper pertains specifically to school age students, it also illuminates deficiencies in college students’ understandings. Schield (2008) highlighted that percentages are highly misused/misunderstood and pointed to his observation that “one college student in five could not correctly read [a] simple pie chart of percentages” (p. 92). He noted that “(c)ollege students also have considerable difficulty determining part and whole in ratios presented in tables and graphs”, in “comparing percentages and rates using ordinary English”, and in dealing with weighted averages (Schield, 2008, pp. 93-95). Schield (2008) said he hopes that students will see more value in contextual percentages and rates than in the manipulation of fractions as mathematical artifacts alone. Schield (2008) offered eight recommendations for teaching fractions in the context of percentages and rates:

1. Emphasize ordinary English.
2. Distinguish percentages from fractions.
3. Be aware of how students and adults – even very bright people – avoid common fractions.
4. Introduce arithmetic operations using percentages and rates in context.
5. Be aware of objections to increasing the focus on percentages and rates in context.
6. Identify advantages to other mathematical topics that might be introduced to help students develop their conceptual powers instead of common fractions.
7. Identify places in the curriculum to introduce or embed the study of fractions in context.
8. Identify and teach topics that college students in non-quantitative majors need to master at the school level and which are currently not being taught there (pp. 98-104)

Corrine Taylor (2008) wrote that students "need to work extensively with percents and ratios in real contexts" (Taylor, 2008, pp. 112-118). Taylor (2008) added, "case studies require students to evaluate quantitative evidence, determine reasonable analytical approaches, perform complex calculations, make decisions, and communicate not only the results but also the process" (Taylor, 2008, p. 113). Taylor (2008) said that as educators, "we need to offer more opportunities for students to make *decisions* that involve information-gathering and assessment, quantitative analyses, and communications about quantitative topics" in addition to "textbook *calculations* that use mathematics" (Taylor, 2008, pp. 118-119). Taylor (2008) concluded that in order to "best prepare students for the highly quantitative real world of business, teachers need help in creating authentic, complex problems that integrate math, research, technology, and communication skills" (p. 121). From the students' perspectives, Taylor (2008) concluded that "(s)tudents need interesting and practical examples to make it abundantly clear that mathematics skills are applicable in the real world" (p. 121).

Joel Best (2008) argued, "quantitative literacy needs to move beyond calculation to understand the social processes that shape the creation and consumption of statistics about public issues" (p. 125). To Best (2008), calculation refers to "all of the practices by which mathematical problems are framed and then solved" and he eschewed the "narrow, technical sense" (p. 125) of the term. Best (2008), a sociologist, focused on the "social construction of statistics" and its relevance to the instruction of quantitative literacy; he admited that students probably "need to improve their calculating skills", but argued, "key forms of quantitative literacy require moving beyond calculation" (p. 125). Best (2008) noted that "numbers are social constructions" since "numbers do not exist in nature", and "(e)very number is a product of human activity" (p. 126). He argued, "numbers that promise quantitative measures" are fundamental to understanding the world and that "students need to learn to think critically about these numbers"; further, Best (2008) claimed "this requires more than having a sense of how those numbers were calculated" and involves understanding "these statistics as the results of social and political, as well as mathematical, processes" (Best, 2008, p. 128).

Best (2008) provided a survey of the endless uses of numbers and statistics in society and the associated motivations, agendas, biases, and misunderstandings in their construction. From this, he argued that "(e)valuating the sorts of numbers – and the claims that such numeric evidence is used to support – that appear in news reports about public issues requires a broader set of critical thinking skills than mastering calculation" (Best, 2008, PP. 132-133). Best (2008) characterized educators promoting quantitative literacy as "mathematics educators who have become skeptical about the practical value of traditional math instruction", but claimed that "they have not gone far enough"; Best (2008) said that "quantitative literacy requires some distinctly non-mathematical – that is, more than calculation-based – skills" (p. 133). Best (2008) recognized the difficulty in incorporating his goal into QL programs because mathematics educators "who are, after all, the folks most interested in quantitative literacy… … have been trained to teach calculation, and they tend to define the problem of quantitative literacy in terms of people being insufficiently adept at calculation" (p. 134).

In Fall 2004 Bernard L. Madison initiated a quantitative reasoning course (QRCW) at the University of Arkansas in Fayetteville. "The experimental course used as its curricular guide newspaper and other public media articles and graphics collected by the instructor or submitted by students that contained quantitative information or analyses of data" (Dingman & Madison, 2010A, p. 2). The university "approved the course as MATH 2183, with college algebra as a pre-requisite" three semesters after the initial experiment (Dingman & Madison, 2010A, p. 2). Dingman and Madison (2010A) said the course "serves as the terminal mathematics or statistics course for many students" and "meets the mathematics requirement for the Bachelor of Arts degree in the College of Arts and Sciences" (Dingman & Madison, 2010A, p. 2). The course then utilized a non-standard mathematics text *Case studies for quantitative reasoning: A casebook of media articles* (2rd edition) written by Dingman, Madison, Stuart Boersma, and Karen Diefenderfer (2009). "The textbook contain[ed] 24 case studies of media articles, with each case study having warm-up exercises and study questions pertaining to the quantitative information in the article" (Dingman & Madison, 2010A, p. 3). Dingman and Madison (2010A) stated that the primary learning goal of the course "is to prepare students to answer questions analogous to the study questions about unpredictable media articles or quantitative situations they encounter in everyday life" (p. 3). Dingman and Madison (2010A) described the "News of the Day" as a "critical component of the course" where students bring in media articles with "quantitative content" (p. 3) for extra credit. Dingman and Madison (2010A) claimed

> (t)his activity [News of the Day] keeps the class content fresh and allows for student interests to surface. The discussions venture into social, economic, and political issues and create connections that some students had not previously considered. The News of

> the Day feature is aimed at extending QR beyond the course and beyond school by providing a venue for continuing practice and leveraging student interest (p. 3).

Dingman and Madison (2010A) described the student population for the QRCW as “quite diverse with about equal numbers of men and women, students from various majors, honors students, student athletes, non-traditional students, students with learning disabilities, and students who consider themselves ‘bad’ at mathematics” (p. 4). Though most students entered the course with significant algebra exposure, “algebra is minimally accessible to them”, and students in the course struggle with both proportional reasoning and their attitudes relating to mathematics (Dingman & Madison, 2010A, p. 4). Dingman and Madison (2010A) wrote that many students arrived in the QRCW with the notion that success in mathematics comes from “memorizing facts and formulas or learning how to use a particular procedure to solve problems that typically are void of context” (p. 5). Further, prior success in the mathematics classroom was uncommon among the QRCW students. These realities lead the QRCW course designers to emphasize a collaborative learning environment and focus content on contextual, real-world situations. Dingman and Madison (2010A) noted that “(m)any students struggle with the requirement that they reason in order to determine how to solve a problem” (p. 5). Students tend to ask instructors for the correct method in a problem-solving scenario “rather than spend time investigating what method would be most useful in the given situation” (Dingman & Madison, 2010A, p. 5). Students tended to focus on obtaining a correct answer rather than “understanding the process that led to that answer” (Dingman & Madison, 2010A, p. 5). Dingman and Madison (2010A) explained that students request “cues” (p. 5) for how to approach certain problem types and struggle with developing their own intuition based on context alone. Dingman and Madison (2010A) claimed that “placing the mathematical and statistical topics in real-world contexts” (p. 6) allowed students to connect the mathematics they learned in school with their everyday life and might have caused students to view mathematics as more relevant. Dingman and Madison (2010A) noted that “older, non-traditional students” “seemed more receptive to the applicability of the content under study to their everyday life, presumably due to their greater life experiences, in comparison to that of their younger classmates” (p. 6). The authors also noted that non-traditional students seemed “less receptive to the collaborative learning environment” and the authors highlighted a request by a non-traditional student for inclusion of more lecture and less group work in the course (Dingman & Madison, 2010A, p. 6).

Dingman and Madison (2010A) noted the difficulty in keeping course material fresh and relevant; especially since the media articles in the course text are inherently set in the past. They claimed that the News of the Day activity was at least a partial remedy to this. Another difficulty noted by Dingman and Madison (2010A) lies in creating case studies that are both compelling to and accessible by students.

> Although much of the mathematical and statistical content encountered in the course is generally taught in middle to early secondary grades, the embedding of the content in real-world contexts and the use of reasoning to determine solution strategies elevates the degree of sophistication for many students… … students have come to view mathematics as something completely different from what is presented in the QRCW course. As many have stated to us, they feel that they understand the concepts but struggle when they have to apply them in real-world settings where a procedure to solve the problem is not readily evident (Dingman & Madison, 2010A, pp. 7-8).

Dingman and Madison (2010A) reported that “warm-up exercises” (p. 8) were added to the course text as a concession to students’ inability or unwillingness to work on or even begin case studies where they encountered difficulty.

Dingman and Madison (2010A) said the “major challenge in assessing QL concerns the central goal of transfer of knowledge and cognitive processes to contexts that are unpredictable and of unbounded variation” (p. 8). Course assessments in the QRCW “consist[ed] of homework assignments for most class meetings, quizzes, and two examinations: a mid-term and a final" (Dingman & Madison, 2010A, p. 8). Further, "(a)lmost all problems and exercises [were] contextual, and each quiz and examination contain[ed] questions stemming from at least one media article that is new to the students" (Dingman & Madison, 2010A, pp. 8-9). Dingman and Madison (2010A) described the QRCW learning goals as aligned with the AAC&U QL VALUE rubric (2009). Dingman and Madison (2010A) use these to describe a "canonical QL situation" as "encountering; interpreting; gleaning and assuming; modeling and solving; and reflecting" (p. 9).

Dingman and Madison (2010A) reported on "two attempts at a pre- and post-course test instrument, with mostly multiple-choice items, to assess the growth in our students' QL abilities across a semester of QRCW" (p. 9). Though each testing showed "modest gains" (p. 9), Dingman and Madison (2010A) reported dissatisfaction with the instruments used. Dingman and Madison (2010A) said that due to the "wide range of contexts" (p. 10) in which QR is needed, its instruction is, at times, very demanding. "The adaptable nature of the course is quite challenging, particularly for faculty accustomed to having greater control over what is discussed and covered in the classroom" (Dingman & Madison, 2010A, p. 10). Challenges arose in the need to be well versed in areas outside of mathematics and in steering discussion back toward a quantitative focus "when, from the students' vantage point, the underlying context provides for an interesting debate and discussion" (Dingman & Madison, 2010A, p. 10). Dingman and Madison (2010A) noted that "(t)he nature of this type of classroom directly confronts instructors' ideas regarding how mathematics is taught – a manner completely different from one that many have experienced in a traditional, lecture-driven classroom format" (p. 11). Another difficulty lay in "fighting the urge as an instructor to step in and show the student how to work a problem" since “showing students how to solve a specific problem generally does not help the students solve other problems and thereby has only helped them with the roadblock in front of them” (Dingman & Madison, 2010A, p. 11).

Dingman and Madison (2010A) noted that attitudinal surveys were conducted and compared between THE QRCW, a survey of calculus course, and a “general education course taught using *For All Practical Purposes* (CoMAP, 1988) as a textbook” (pp. 11-12). The authors noted that although “the mean of the responses to each pre- and post-attitude item moved in the direction we wanted” the improvements were not dramatic and the sample sizes were relatively small (n=15 in 2007 and n=20 in 2008) (Dingman & Madison, 2010A, p. 12). The authors also sent out “email questionnaires to approximately 300 students who had finished the QRCW course in previous semesters and whose university email addresses were still active” (Dingman & Madison, 2010A, p. 12). The survey consisted of four questions:

1. How often have you practiced analyzing quantitative content of public media (newspaper, magazine, advertising flyer, online material, etc) articles since you finished the Mathematical Reasoning course? Never, Rarely, Regularly, or Often.
2. How has your confidence with quantitative reasoning changed since the course? Decreased, Stayed the same, or Increased.
3. How has your view of the importance to you of quantitative reasoning changed since the course? Decreased Stayed the same, or Increased.
4. Any other comment (Dingman & Madison, 2010A, p. 12).

Dingman and Madison (2010A)

> found that 69% (29 of 42) of respondents stated their confidence with QR had increased since the course and that 76% (32 of 42) held an increased importance to QR, although 55% (23 of 42) replied they rarely practiced analyzing the quantitative content in public media (pp. 12-13).

Dingman and Madison (2010A) also documented increased retention and success as compared to other “introductory mathematics courses” (p. 13).

Dingman and Madison (2010B) raised several questions as a result of their QRCW analyses. Dingman and Madison (2010B) noted the importance of enhancing “students’ ability to transfer knowledge and cognitive processes to solve problems in a variety of contexts” (p. 2). Dingman and Madison (2010B) observed that sometimes, “learning seems to be context-bound” and noted instances where students are unable to transfer mathematics learned in one context to another, even whenever the related mathematical processes are fundamentally identical. This raised two questions from the authors:

> What is the proper mix of abstract learning and learning in authentic situations for achieving quantitative literacy (QL)?
>
> What instructional techniques and tasks best promote transfer of knowledge and skills to a variety of contexts (Dingman & Madison, 2010B, p. 4)?

Another reported primary objective of the QRCW is to promote a productive disposition towards QL and mathematics among students. “As described in *Adding It Up,* productive disposition is the ‘habitual inclination to see mathematics as sensible, useful, and worthwhile, coupled with a belief in diligence and one’s own efficacy.’” (p. 5)” (Dingman & Madison, 2010B, p. 4). Dingman and Madison (2010B) discussed the results of the survey reported in Dingman and Madison (2010B) and noted that “QRCW students are reluctant to claim that they are good at mathematics” (p. 5). These results raised two questions from the authors:

> How can productive disposition be strengthened, particularly among students majoring in non-mathematical fields of study?
>
> Does the use of the word ‘mathematics’ rather than ‘quantitative reasoning’ or some other word such as ‘statistics’ skew responses to attitude questions (Dingman & Madison, 2010B, p. 5)?

Dingman and Madison (2010B) singled out “(o)ne of the most common and recurring mistakes [they] have observed students make in the QRCW class” as “using the wrong base for a percent" (p. 6). Dingman and Madison (2010B) defended this claim with results from a "pre- and post-test during the Spring 2008 semester" given to 95 QRCW students and 83 "students in a mathematics course using *For All Practical Purposes* (CoMAP, 1988) as a textbook" (p. 6).

> The two groups of students were similar, namely all arts, humanities or social sciences majors. The students were posed the following question: *The Fall 2007 enrollment of 18,200 was an increase of 4% over the Fall 2006 enrollment. What was the Fall 2006 enrollment?*
>
> There were five possible choices for an answer. One was the correct answer of 17,500, since raising 17,500 by 4% produces an enrollment of 18,200, while another possible solution was the result of reducing 18,200 by 4%, or 17,472. On the pre-test, 60% of the 178 students chose the incorrect 17,472, with 25% choosing the correct answer of 17,500. The other 15% of the 178 students chose one of the other three incorrect options. On the post-test, 56% of the students chose the incorrect 17,472, while 38% chose the correct answer. The QRCW students fared better than the other group with regards to this item on the post-test, but still only 44% of QRCW students chose the correct answer after similar items were considered several times in the QRCW course (Dingman and Madison, 2010B, p. 6).

Dingman and Madison (2010B) said they believe the issue is largely a combination of "reluctance to compute with an unknown" and "the tendency to multiply rather than divide, especially where percents are involved" (p. 7). Dingman and Madison (2010B) raised two questions in this regard:

> What types of instructional or curricular interventions best assist students in determining the correct base for a percent?
>
> Is multiplication more natural than division for these students? If so, when and how is this tendency developed (p. 7).

Dingman and Madison (2010B) noted that their students still struggled with algebra and its application in context even with algebra as a pre-requisite for the QRCW. They pointed to an "action conception" among students, the "first stage of a mental framework termed Action-Process-Object-Schema, or APOS" (p. 10). "These students' understandings of linear and exponential equations/functions are restricted, often to a single equation that can be evaluated at specific points and whose expression can be manipulated" (Dingman & Madison, 2010B, p. 10). In general, students wanted to evaluate or manipulate expressions and rarely search for overall patterns or trends. "(T)he students' algebra is more likely a fragmented collection of methods and operations" and "(t)his kind of fragmentation is known to reduce transfer, and hence usability of the knowledge" (Dingman & Madison, 2010B, pp. 11-12). Dingman and Madison (2010B), citing Thompson (1988), and Smith and Thompson (2007), claimed that QR "should lead to algebraic reasoning, giving evidence supporting the premise that algebraic reasoning should derive from quantitative reasoning rather than as a generalization of arithmetic" (p. 12-13). Dingman and Madison (2010B) raised several questions about the relationship between fluency in algebra and proficiency in QL, including: "(c)an QL be developed solely from arithmetic and proportional reasoning?" and "(t)o what extent does mathematical fluency contribute to QL?" (p.13).

Dingman and Madison (2010B) wrote that QRCW students have a tendency to "believe that increasing is preferred over decreasing" and they termed this as the "bigger is better" (p. 13) fallacy. The authors noted several instances where students use faulty reasoning along these lines. The questions raised by Dingman and Madison (2010B) have gone largely unanswered, and many of them are mathematical in nature. An assessment rubric for QL that identifies QL specific skills would aid responses to some of the questions raised by Dingman and Madison (2010B).

QL educators have identified skills specific to QR. Van Groenestijn (2003) identified skills specific to "managing mathematical situations" in a QR context:

> *Generative mathematical understanding and insight* to give meaning to and interpret numbers and to plan appropriate mathematical actions;

> *Literacy skills* to read and understand problems and to reason about them;
>
> *Communication skills* to be able to share problems with others, discuss information, learn from others how they would solve problems, and work cooperatively;
>
> *Problem-solving skills* to identify, analyze, and structure problems, plan steps for action, select appropriate actions, actually handle problems, and make decisions; and
>
> *Reflection skills* to be able to control the situation, check computations, evaluate decisions, and come to contextual judgments (p. 232)

The management skills identified by Van Goenestijn (2003) overlap significantly with the Association of American Colleges and Universities' (AAC&U, 2009) Quantitative Literacy Value Rubric. AAC&U (2009) utilized a team of QR experts to develop a rubric that identifies 6 "core competencies":

> *Interpretation*; ability to explain information presented in mathematical forms (e.g., equations, graphs, diagrams, tables, words).
>
> *Representation*; ability to convert relevant information into various mathematical forms (e.g., equations, graphs, diagrams, tables, words).
>
> *Calculation*.
>
> *Application/Analysis*; ability to make and evaluate important assumptions in estimation, modeling, and data analysis.
>
> *Communication*; Expressing quantitative evidence in support of the argument or purpose of the work (in terms of what evidence is used and how it is formatted, presented, and contextualized) (AAC&U, 2009).

Boersma et al., (2011) translated "the QL VALUE rubric to a QL Assessment Rubric (QLAR) in order to assess student work in QRCW courses" (p. 2). Boersma et al., (2011) "address[ed] the importance of using a specific QLAR in developing specific QL course materials" and claimed "(t)his approach would improve organization and wording of study questions in the QRCW casebook" (pp. 2-3). The authors noted, "the core competencies that comprise the areas of scoring student work provide a structure for helping to organize how students learn to reason quantitatively" and "provide an anchor for further studies using other QR materials and courses" (p. 3). "The QL Assessment Rubric (QLAR) is an adaptation of the AAC&U VALUE QL rubric to make it more applicable to grading student work" (Boersma et al., 2011, p. 4). Figure 1 details the QLAR rubric. Boersma et al., (2011) described several changes to the original VALUE rubric made in the final version of the QLAR:

1. Included a column for a score of 0. The VALUE rubric had a default score of zero if the work did not meet the level-one benchmark, and a score of zero was also assigned if the core competency was not present as a part of the answer. We include a zero column to more clearly acknowledge the presence or absence of core competencies.
2. Removed the column for a score of 4. In the VALUE rubric the score of four designated a capstone achievement. Capstone proficiency requires a cumulative experience over a complete undergraduate curriculum and is not a normal consideration for student work on a specific assignment in a stand-alone QL course.
3. Provided more objective descriptions of achievement levels. Several of the descriptions in the VALUE rubric made distinctions between levels difficult because of the use of qualitative words. For example, distinguishing between 'workmanlike' and 'competent' or between 'plausible' and 'reasonable' was found to be too subjective.
4. Changed the core competency of application/analysis to analysis/synthesis. This change was made to accommodate drawing conclusions by either analyzing (that is, breaking apart circumstances) or synthesizing (that is, pulling together components) (p. 6).

	Achievement Level			
Quantitative Literacy Core Competency	**3**	**2**	**1**	**0**
Interpretation *Ability to glean and explain mathematical information presented in various forms (e.g. equations, graphs, diagrams, tables, words)*	Correctly identifies all relevant information.	Correctly identifies some, but not all, relevant information.	Some relevant information is identified, but none is correct.	No relevant information identified.
Representation *Ability to convert information from one mathematical form (e.g. equations graphs, diagrams, tables, words) into another.*	All relevant conversions are present and correct.	Some correct and relevant conversions are present but some conversions are incorrect or not present.	Some information is converted, but it is irrelevant or incorrect.	No conversion is attempted.
Calculation *Ability to perform arithmetical and mathematical calculations.*	Calculations related to the problem are correct and lead to a successful completion of the problem.	Calculations related to the problem are attempted but either contain errors or are not complete enough to solve the problem.	Calculations related to the problem are attempted but contain errors and are not complete enough to solve the problem.	Calculations given are not related to the problem, or no work is present.
Analysis/Synthesis *Ability to make and draw conclusions based on quantitative analysis.*	Uses correct and complete quantitative analysis to make relevant and correct conclusions.	Quantitative analysis is given to support a relevant conclusion but it is either only partially correct or partially complete (e.g. there are logical errors or unsubstantiated claims).	An incorrect quantitative analysis is given to support a conclusion.	Either no reasonable conclusion is made or, if present, is not based on quantitative analysis.
Assumptions *Ability to make and evaluate important assumptions in estimation, modeling, and data analysis.*	All assumptions needed are present and justified when necessary.	At least one correct and relevant assumption is given (perhaps coupled with erroneous assumptions), yet some important assumptions are not present.	Attempts to describe assumptions, but none of the assumptions described are relevant.	No assumptions present.
Communication *Ability to explain thoughts and processes in terms of what evidence is used, how it is organized, presented, and contextualized.*	A correct and complete explanation is clearly presented.	A partially correct relevant explanation is present, but incomplete or poorly presented.	A relevant explanation is present, but is illogical, incorrect, illegible, or incoherent.	No relevant explanation is provided.

Figure 1. The QLAR (Boersma et al., 2011, p.5)

Boersma et al. (2011) described mapping study questions from the QRCW casebook to core competencies from the QLAR that are necessary for complete solutions.

> The 24 case studies in the casebook have a total of 234 study questions. Each study question was mapped to one or more of the core competencies in the QLAR. Two of us mapped the study questions in the first half of the casebook, and two others mapped the second half. Following that, we swapped study questions to determine if each pair would independently create the same mapping as the other. Agreement of the mapping was substantial and all differences were resolved in discussions… …One of the points of disagreement between the two groups of coders was whether a question required communication or analysis/synthesis. Obviously, most answers require some communication, but communication in this rubric is '*Explaining thoughts and processes in terms of what evidence is used, how it is organized, presented, and contextualized.'* Analysis/synthesis is *'Making and drawing conclusions based on quantitative analysis.*' Drawing a distinction is sometimes difficult, but scoring under one or the other is usually the resolution. The core competency of communication, as we applied it, concerns the explanation of a process, that is, a description of the thinking and how conclusions were obtained. The actual thinking and conclusions are part of the analysis/synthesis competency.
>
> Another point of disagreement concerned the overlap of interpretation and communication, since both could involve explanations. Our final definition of interpretation allows for interpreting explanations to be scored under the interpretation competency: Interpretation – *Ability to glean and explain mathematical information presented in various forms (e.g. equations, graphs, diagrams, tables, words)*. Therefore explaining mathematical information, whether it was gleaned from text by students or given in the question prompt, was assigned to interpretation and not communication (Boersma et al., 2011, p. 8).

The QRCW casebook consists of a total of 24 case studies and 234 study questions; Boersma et al. (2011) mapped study questions to a total 467 required core competencies with agreement on 437 (94%) of these competencies (p. 8). Boersma et al. (2011) noted, "all six competencies were required in three case studies, five were required in 13 case studies, four in seven, and one case study required only two" (p. 9). Boersma et al. (2011) explained that "the casebook was compiled without these core competencies or any other set of competencies as a guide", and "(t)he only effect of core competencies or learning goals guiding the development of the casebook was in the intuition and experience of the authors" (p. 9). Table 1 shows the "frequency and prevalence of the six core competencies in the casebook" (Boersma et al., 2011, p. 9). Boersma et al. (2011) also explained that, unfortunately, the core competency related to evaluating or making assumptions is underrepresented in the questions contained in the casebook (p. 9).

Boersma et al. (2011) conducted two reliability studies on scoring student work with the QLAR; their “reliability exercises indicate that the QLAR can be used to consistently score student work and that consistent scoring can result from multiple readers” (p. 13). Boersma et al. (2011) warned that the QLAR may not be appropriate for scoring student work for a course grade;

> (u)sing a rubric such as the QLAR with appropriate preparation and weighting of scores would produce creditable scores, but much of the value of a rubric such as the QLAR lies in other areas, [such as] guiding student thinking and production of instructional materials (p. 13).

Boersma et al. (2011) suggested that introducing students to the QLAR could help them organize their thought processes and responses. In mapping “study questions from the QRCW casebook to the six core competencies in the QLAR” Boersma et al. (2011) “discovered study questions that could be significantly improved by providing more explicit directions in order to elicit student responses that were better aligned with the competencies” (p. 14). Boersma et al. (2011) claimed “(t)his was particularly helpful if the question mapped to several competencies” (p. 14).

> For example, one of our prompts was, ‘Find out how the Standard & Poor’s 500 stock index is computed.’ If we want to score that response on communication, as we did, then the prompt would be better stated as, ‘Describe how the Standard & Poor’s 500 stock index is computed.’ Asking for an explanation will highlight the need for communication (Boersma et al., 2011, p. 14).

Table 1. Frequency and Prevalence of Core Competencies in the QRCW Casebook (Boersma et al., 2011)

Frequency and Prevalence of Core Competencies in the QRCW Casebook				
Competency	**Number of questions**	**Percent of questions**	**Number of case studies (*N*=24) which have at least one question mapped to competency**	**Number of case studies which have 50% or more questions mapped to competency**
Interpretation	152	65	24	18
Representation	68	29	19	6
Calculation	101	43	21	9
Analysis/Synthesis	79	34	20	7
Assumptions	14	6	10	0
Communication	53	23	18	4

Boersma et al. (2011) wrote that "(t)he QRCW experience has pointed to the need to scaffold student thinking and performance" (p. 14). To this end, Boersma et al. (2011) added "warm-up exercises" to the case studies to better prepare students "for the open-ended study questions", and re-arranged case studies "so that earlier ones require interpretation and representation" in order to "build on this understanding when more complex prompts appear later" (p. 14).

A *habit of mind* (HoM) is often referenced in QR literature, and many of the most recently used definitions for QR contained some reference to a HoM. Gal (1997) pointed to the notion that mathematics educators should seek to empower students to become "comfortable enough with mathematics and its applications… … so they will later be willing to invest in further mathematics-based learning… … when life so demands" (p. 42). Learning mathematics when it is demanded by a life event is an example of a HoM held by quantitatively literate individuals. Though, when educators say HoM, they do not always refer to a mathematically oriented construct. Cohen (2001) argued for emphasizing QL as "not only about arithmetic and higher mathematics but also about a general skill (or habit of mind) that is required in many subjects across the curriculum" (Cohen, 2001, p. 28). Whenever QR experts use the phrase, there is often an implied overlap with mathematical underpinnings of QR and application to contexts outside of the QR classroom. Steen (2004D) noted that "(m)ost opportunities to employ mathematical concepts are not available to students at the same time that mathematical skills are first learned: only through repeated use in increasingly complex circumstances can these mathematical skills become QL habits of mind" (p. 17). He wrote that most educators interested should not take "total ownership" since otherwise "students will continue to see QL as something that happens only in the mathematics classroom" (p. 18). Further, since productive disposition references a *habitual inclination*, it can be said that QL educators believe there is a relationship between HoM and productive disposition.

Grawe et al. (2010) reported on an instrument designed at Carleton College to assess QR in written arguments by students. Grawe et al. (2010) explained that standardized assessments (e.g. utilizing multiple-choice and calculation problems) "can tell us whether students have the capacity to apply QR knowledgeably when prompted to do so", but they cannot "show whether students have strengthened a tendency to use that capacity or have developed the skills necessary to deploy the capacity effectively in contexts other than those in the test" (p. 1). Grawe et al. (2010) summarized QR as "the habit of mind to consider the power and limitations of quantitative evidence in the evaluation, construction, and communication of arguments in personal, professional, and public life" (pp. 1-2). They noted that capability in "skills-based assessments" does not imply a students "have developed the habit of mind or flexibility to apply those competencies in the context of arguments" (Grawe et al., 2010, p. 2).

Grawe et al. (2010) presented a rubric for "measuring QR in written arguments" that was "developed over four years in the context of Carleton's QR initiative" (p. 2). This rubric was intended "to assess QR at an institutional level", and was "*not* designed to evaluate individual students" (Grawe et al., 2010, p. 2). Grawe et al. applied their rubric to papers "submitted by students for Carleton College's sophomore writing portfolio" (p. 2); these papers were not submitted with the intention of QR content evaluation, but were a general requirement of all students at Carleton College. Grawe et al. (2010) applied their rubric to these papers in the hope of gaining

> insight into how we can improve instruction at the institution and to compare QR activity between large groups (e.g., the class of 2005 vs. the class of 2010, or students who major in the social sciences vs. those who major in the humanities) (p. 3).

Grawe et al.'s (2010) rubric can be viewed in Figure 2. The first section of the rubric was intended for identification purposes, the second "to assess the potential contribution of quantitative information to the paper based on the stated and implied goals of the paper itself", the third to "evaluate the extent of quantitative evidence present in the paper", the fourth to assess the "overall quality of the use of QR in the paper" (if applicable), the fifth to identify specific problems with the QR present (if applicable) and the sixth to "determine whether [the assignment] explicitly calls for the use of QR" (Grawe et al., 2010, pp. 3-8). The categories in section three were detailed in the following list:

1. No explicit numerical evidence or quantitative reasoning. May include quasi-numeric references (i.e., 'many,' 'few,' 'most,' 'increased,' 'fell,' etc.).
2. One or two instances of explicit numerical evidence or quantitative reasoning (perhaps in the introduction to set the context), but no more.

3. Explicit numerical evidence or quantitative reasoning is used throughout the paper (Grawe et al., 2010, p. 5).

The attached criteria for assessing quality of QR in section four of the rubric can be viewed in figure 3. Grawe et al. (2010) identified the items in section five as errors “common to first-year and sophomore papers”; in section five, rubric readers scored “for the *presence* of a problem” (p. 8).

Grawe et al. (2010) had high reliability in assessing QR relevance, section two of the rubric, and reported “exact agreement in more than three-fourths of cases” (p. 11). “Exact agreement was achieved in more than 80% of cases” for section three of the rubric (Grawe et al., 2010, pp. 11-12). Section four had lower inter-rater reliability with “exact agreement in over 65% of all cases”, however, after collapsing the middle two scores (two and three) into a single score to make a three-category scale, “readers achieved exact agreement in more than 75% of all cases” (Grawe et al., 2010, pp. 12-13). Grawe et al. (2010) were unable to achieve high levels of reliability in scoring for section five. Grawe et al. (2010) concluded, “a group of readers drawn from across all divisions of the academy can be trained to apply the rubric reliably” (p. 15). Further, Grawe et al. (2010) report they found the rubric to be

> an effective formative assessment tool in at least three senses. First, the process of collectively reading papers through the lens of the rubric has nurtured a focused discussion around the definition of QR, evidence of its presence, assignments that support its development, and professional development activities that might enhance QR instruction… … Second, application of the rubric to student work has helped to identify examples of weak and strong student use of QR – examples which have strengthened presentations given to a wide audience at workshops, learning and teaching center seminars, and faculty retreats. Finally, the findings of our assessment work have shaped our programming. For example, recognizing the large fraction of papers for which QR is peripherally relevant led to professional development workshops designed to encourage assignments that teach the effective use of numbers to frame an argument (p. 16).

Quantitative Reasoning in Student Writing Rating Sheet

I. Identification. **Student I.D. #:_____________** **Reader I.D. #:____**

II. Is QR potentially relevant to this paper? [rate potential contents of paper, not the assignment]

___ No or incidentally only ___ Yes, but peripherally only ___ Yes, centrally

III. What is the extent of numerical evidence and quantitative reasoning present in the paper?

[See: "Employs QR Criteria"; **Note**: This is *not* a rating of the quality of the QR shown, only its presence.]

___ rating of 1-3, review attached criteria

IV. OVERALL ASSESSMENT of Quality of implementation, interpretation, and communication of QR:

___ rating of 1-4, review attached criteria

V. Problematic characteristics of the QR present in the paper: [check all issues that detract significantly from the reader's understanding of the information presented.]

___ Uses ambiguous words rather than numbers.

___ Fails to provide numbers that would contextualize the argument.

___ Doesn't evaluate source or methods credibility and limitations.

___ Inadequate scholarship on the origins of quantitative information cited.

___ Makes an unsupported claim about the causal meaning of findings.

___ Presents numbers without comparisons that might give them meaning.

___ Presents numbers but doesn't weave them into a coherent argument.

VI. Does the assignment explicitly call for the use of QR in the paper?

___ YES ___ NO ___ NO ASSIGNMENT PRESENT

Figure 2. Scoring rubric (Grawe et al., 2010, p. 4).

A. In Papers where QR is Centrally Relevant			
Quality Score			
1	**2**	**3**	**4**
Use of numerical evidence is so poor that either it is impossible to evaluate the argument with the information presented or the argument is clearly fallacious. Perhaps key aspects of data collection methods are missing or critical aspects of data source credibility are left unexplored. The argument may exhibit glaring misinterpretation (for instance, deep confusion of correlation and causation). Numbers may be presented, but are not woven into the argument	**The use of numerical evidence is sufficient to allow the reader to follow the argument. But there may be times when information is missing or misused.** Perhaps the use of numerical evidence itself is uneven. Or the data are presented effectively, but a lack of discussion of source credibility or methods makes a full evaluation of the argument impossible. Misinterpretations such as the confusion of correlation and causation may appear, but not in a way that fundamentally undermines the entire argument.	**The use of numerical evidence is good throughout the argument.** Only occasionally (and never in a manner that substantially undermines the credibility of the argument) does the paper fail to explore source credibility or explain methods when needed. While there may be small, nuanced errors in the interpretation, the use of numerical evidence is generally sound. **However, the paper may not explore all possible aspects of that evidence.**	**The use of numerical evidence is consistently of the highest quality.** When appropriate, source credibility is fully explored and methods are completely explained. Interpretation of the numerical evidence is complete, considering all available information. There are no errors such as confusion of correlation and causation. **This paper would be an excellent choice as an example of effective central QR to be shared with students and faculty.**
B. In Papers where QR is Peripherally Relevant			
1	**2**	**3**	**4**
Fails to use any explicit numerical evidence to provide context. The paper is weaker as a result. **This paper shows no attempt to employ peripheral QR.**	Uses numerical evidence to provide context in some places, but not in others. The missing context weakens the overall paper. Or the paper may consistently provide data to frame the argument, but fail to put that data in context by citing other numbers for comparison. **Ultimately, the attempt at peripheral use of QR does not achieve its goal.**	The paper consistently provides numerical evidence to contextualize the argument when appropriate. Moreover, numbers are presented with comparisons (when needed) to give them meaning. However, there may be times when a better number could have been chosen or more could have been done with a given figure. **In total, the peripheral use of QR effectively frames or motivates the argument.**	Throughout the paper, numerical evidence is used to frame the argument in an insightful and effective way. When needed, comparisons are provided to put numbers in context. **This paper would be an excellent choice as an example of effective peripheral QR to be shared with students and faculty.**

Figure 3. **Rubric Language for Assessing Quality of QR (Grawe et al., 2010, p. 6).**

Grawe et al. (2010) admitted that more research is necessary in order to establish construct validity. The QR concept is multi-faceted, and Grawe et al. (2010) report they would like to better understand which aspects of QR their instrument captures, especially in relation to other assessment tools (p. 16).

Boersma and Klyve (2013) decided to try and measure HoM in students directly and "offer[ed] a new 'prompt-less' instrument for measuring students' habits of mind in the field of quantitative literacy" (abstract). Boersma and Klyve (2013) based their instrument off of the QLAR and aimed it at their courses, which utilized the QRCW casebook as the course text. Boersma and Klyve (2013) operated under the notion that "a quantitatively literate person will have a predisposition to employ a number of mathematical and critical thinking skills *on their own initiative* as opposed to simply responding to a series of prompts" (p. 1). Boersma and Klyve (2013) used such a predisposition to define Habit of Mind (HoM), and noted a logistically reasonable assessment of HoM is inherently difficult, if not impossible – at least in the most general sense of the construct (p. 1).

Boersma and Klyve (2013) narrowed their study "to measure whether [their] students have the inclination to"

> glean, identify and report quantitative information in direct support of a thesis statement;
> invoke quantitative reasoning to critique a statement or opinion;
> check numerical information presented in text with any accompanying graphics; and
> critically evaluate information presented graphically (p. 2).

Boersma and Klyve (2013) eschewed "available assessment instruments" as they "have not been found adequate for measuring the 'habit of mind' component of QL" (p. 2). They cited a "tendency to assess answers as opposed to the reasoning required to arrive at those answers", an untenable surplus of prompting, and overly time-intensive assessment protocols for this inadequacy (Boersma and Klyve, 2013). Although Boersma and Klyve (2013) created their own assessment for HoM, they noted it is based on the QLAR and cite its reliability in scoring student work as an advantage. They designed the HoM instrument so that it can "be administered in a single class period (50 minutes), be scored by a single instructor in a short amount of time, and lend itself to a pre- and post-intervention assessment protocol" (Boersma and Klyve, 2013, pp. 2-3). Boersma and Klyve (2013) used the HoM instrument to assess the QR course at Central Washington University (CWU) and to search for "learner differences between two populations of students at CWU: general non-STEM majors and those non-STEM majors enrolled in [the] honors program" (p. 3).

Boersma and Klyve (2013) utilized a newspaper article in application of their HoM instrument. Their reasoning behind this was multi-faceted; "(n)ewspaper articles can be an excellent source of contextually rich and quantitatively demanding material… … are situated in authentic contexts, written to be understood by a large percentage of our population, and are reasonably short" (Boersma and Klyve, 2013, p. 4). Boersma and Klyve (2013) admitted that their HoM instrument is not completely void of prompts, however they claimed "the prompts were carefully created in order not to overtly lead students to provide the type of responses our rubric was designed to identify" (p. 4). Boersma and Klyve (2013) described the newspaper article vetting process for determining a reasonable article to be used with the HoM instrument. Boersma and Klyve (2013) outlined the criteria for selecting articles in the following bulleted list.

> Be roughly 500 words to allow students to read the articles in class and have enough time to complete the assessment.
>
> Contain content of interest to college students
>
> Contain a variety of quantitative statements thereby requiring students to isolate those statements that are more central to the main theme of the article. Statements using relative quantities (percents, percentiles) and absolute quantities (specific counts) should also be present.
>
> State an argument(s) and use quantitative comparisons in support of the argument(s).
>
> Be accompanied by a graph which exhibits some discrepancies between the numerical information presented in the article.

> Be ripe for criticism – allowing for dialog on its strengths and weaknesses (Boersma and Klyve, 2013, p. 4).

Boersma and Klyve (2013) chose “Top students show little gain from ‘No Child’ efforts” by Liz Bowie “as printed in *The Baltimore Sun* (June 18, 2008)” and “Tally high for Americans at Polls this year” by Fredreka Schouten “as printed in *USA Today* (November 6, 2008)” (p. 5). Boersma and Klyve (2013) explained their prompts “are open ended and not multiple choice”, make no explicit references “to any quantitative or mathematical calculation” (p. 4) and leave only enough room for short (several sentences) responses. The HoM assessment instrument consists of a single page with five questions along with an accompanying newspaper article. These five questions are:

1. Did you understand the article?
2. What was the main point(s) of the article?
3. What facts did the author use to support the main point(s)?
4. Were there any particular strengths or weaknesses in how these facts were reported?
5. Does the graph help interpret the numerical information found in the text? Explain your thoughts (Boersma and Klyve, 2013, p. 5).

Boersma and Klyve (2013) claim these questions

> allow [them] to measure whether students have a ‘habit of mind’ to (1) glean, identify and report quantitative information in direct support of a thesis statement; (2) invoke quantitative reasoning to critique a statement or opinion; (3) check numerical information presented in text with any accompanying graphics; and (4) critically evaluate information presented graphically (p. 6).

Boersma and Klyve (2013) gave this instrument to 23 non-STEM, general population students in Fall 2009 and 40 non-STEM, honors students in Fall 2011. The assessment was first given to students within the first three class meetings of their QR course and then again during the last week of class (with the second article – students did not answer questions about the same article twice). Boersma and Klyve (2013) scored responses based on the six core competencies outlined in the QLAR, and specifically "focused on the core competencies of Interpretation and Analysis/Synthesis" (p. 7).

Boersma and Klyve (2013) noted that the first two questions from the HoM instrument do not directly measure HoM, however the first helps identify students with language difficulties while the second "measures a student's ability to identify the main point of a lengthy article" (p. 7). The third question – "What facts did the author use to support the main point(s)" – relates most closely with the "core competency 'Interpretation' of QLAR"; high scores on this item indicate a "habit of mind to seek out quantitative information in the article", an ability to "identify relevant and specific information" and an ability to "communicate these facts in one to three sentences" (Boersma and Klyve, 2013, pp. 7-8). Boersma and Klyve (2013) used the following rubric with sample responses to score this item:

> **Score 0**: No quantitative information given or alluded to. '*A study by the Brookings Institution.'*
>
> **Score 1**: Some relevant quantitative information is identified (or alluded to), but none is correct (or specific enough to be judged correct or incorrect). '*The average increase in NAEP test scores for lower and top students, teacher and public responses, and quotes from school staff.'*
>
> **Score 2**: Some relevant and correct information is identified, but not all. *'The lowest performing gained 22 points in 7 years while the highest gained 9 points.'*
>
> **Score 3**: All relevant quantitative information is correctly identified. '*The nationwide fourth- grade reading scores for the poorest-performing students have risen 16 points since 2000 compared with only 3 points for the top students. A national teacher survey showed that 60% of teachers said that the struggling students were the top priority in their school.'* (p. 8)

Boersma and Klyve (2013) noted that supporting non-quantitative facts are not rewarded for this item since HoM is the intended measure. The fourth question – "Were there any particular strengths or weaknesses in how these facts were reported?" – delineates students who focus on presentation from students who "focus on quantitative strengths and weaknesses" (Boersma and

Klyve, 2013, p. 8). Boersma and Klyve (2013) noted this item relates most closely with the "core competency 'Analysis/Synthesis'" with high scores indicating a student can draw and base conclusions on "correct and complete quantitative analysis" (p. 8). Boersma and Klyve (2013) use the following rubric with sample responses to score this item:

> **Score 0**: No strength or weakness identified or, if identified, not supported with quantitative reasoning. *All the numbers got kind of confusing.*
>
> **Score 1**: A strength or weakness is identified but is supported with incorrect quantitative reasoning (or the reasoning is not specific enough to be able to judge correctness). *They could have compared voter turnout to the '04 election better by absolute population.*
>
> **Score 2**: A strength or weakness is identified and is supported with quantitative reasoning, but the reasoning is incomplete (e.g., it contains unsubstantiated claims). *The 62.5% could be misinterpreted as the estimate of votes for Obama.*
>
> **Score 3**: A strength or weakness is identified and supported with correct and complete quantitative reasoning. *A strong weakness was that this data was calculated before official results, so these may not be the true numbers. A strength is that they presented the data in 2 different ways: in solid numbers and in percentages.* (pp. 8-9)

The fifth question – "Does the graph help interpret the numerical information found in the text?" – is scored in two ways and measures "a student's habit of mind to check the numerical information in the text with the numerical information being presented graphically and… … students' habit of mind to critically evaluate graphical information within an authentic context" (Boersma and Klyve, 2013, p. 9). The first pass in scoring this item identifies whether or not students checked numerical information presented in the graph with the numerical information presented in the article with high scores indicating they did and, in addition, were able to spot a discrepancy between the numerical information in the graph and the numerical information in the article. Boersma and Klyve (2013) used the following rubric with sample responses to score this item in their first pass:

> **Score 0**: No indication that the numbers in the article were checked against their representation in the graph. *Somewhat. The first graph supports what is mentioned in the article and the second graph helps to prove the same point.*
>
> **Score 1**: Claims, with no justification, or incorrect justification, that the graph does or does not accurately present the numerical information in the article. *Yes because it accurately compares the two levels [of] progression over 11 years*.
>
> **Score 2**: Claims, with justification, that the graph does or does not accurately present the numerical information found in the text. *Yes because it shows how the 90% students didn't really improve over 4 years and the 10th percentile students did.*
>
> **Score 3**: Correctly points out a specific discrepancy between the graphical presentation and the quantitative information found in the text. *Not really. It seems like pretty weak support to me the more I look at it. And the 16 points & 3 points don't really have a place on this graph. The points sound like a whole new graph.* (pp. 9-10)

The second pass in scoring this question is "guided by the 'Analysis/Synthesis' competency of QLAR" (Boersma and Klyve, 2013, p. 10). Boersma and Klyve (2013) detailed that, during the second pass, a grader checks "to see if students [can] draw a conclusion (regarding the usefulness of the graph) and support their conclusion with quantitative analysis" (p. 10). Boersma and Klyve (2013) used the following rubric with sample responses to score this item for the second pass:

> **Score 0**: No strength or weakness of the graph identified or, if identified, not supported with quantitative reasoning. *Yes, the graph does help, but the quantitative information was presented clearly enough to understand without the graph. It is a nice visual aid, however.*
>
> **Score 1**: A strength or weakness of the graph is identified but is supported with incorrect quantitative reasoning (or the reasoning is not specific enough to be able to judge correctness). *The x-axis is really weird. They should have just used a bar graph or pie chart showing 1960, 1968, and 2008 voter turn out. To have a more compelling chart.*
>
> **Score 2**: A strength or weakness of the graph is identified and is supported with quantitative reasoning, but the reasoning is incomplete (e.g. it contains unsubstantiated claims). *It nicely shows that the number of those that have voted has indeed increased. But having the years skip at the bottom is somewhat annoying, I'd rather they keep it consistent.*
>
> **Score 3**: A strength or weakness of the graph is identified and supported with correct and complete quantitative reasoning. *It certainly does seem to support the claim that voting turnout in 2008 was 62.5% and that in 1968, 63.8% of people voted. The 51.7% seemed a*

> *little random till I realized it's probably the lowest point between these two years. They are also nice enough to write at the bottom that this is an 'unofficial estimate'.* (p. 10)

Boersma and Klyve (2013) noted their "scoring methodology identified students who took the initiative (unprompted) to comment on any particular strengths or weaknesses they noticed in the graph" (p. 10).

Boersma and Klyve (2013) reported difficulties in determining differences between pre- and post- test scores. In particular, they had no way to determine the relative difficulty in responding to prompts between articles, if between article differences exist, since both of their cohorts pre-tested with the first article and post-tested with the second. They suggested pre- and post- testing article randomization for future studies. Further, Boersma and Klyve (2013) reported "no evidence that students' habits of mind were significantly improved by [their] course" since "(n)one of the differences in the 'Overall' column are statistically significant" (p. 12). "In terms of average score… … no group showed significant change on any question" (Boersma and Klyve, 2013, p. 12). Boersma and Klyve (2013) noted the HoM instrument is still useful in measuring "students' habits of mind and quantitative reasoning abilities", and showed evidence that responses to the non-quantitative question two, "What was the main point(s) of the article", have a positive correlation with scores on question 4, "Were there any particular strengths or weaknesses in how these facts were reported" (p. 12). Of people who responded incorrectly to question two (N=50), the mean score on question 4 was 0.60, and of people who responded correctly to question two (N=65), the mean score on question 4 was 1.06 (Boersma and Klyve, 2013, p. 12).

Boersma and Klyve (2013) believe they are able to measure students' habit of mind since they attained relatively high reliability with the instrument whenever they compared scores with each other; “(f)or most of the problems, we found we disagreed in 20-25% of the cases” (p. 13). Boersma and Klyve (2013) note that “in scoring the second pass at question 5… …we found that our scores had initially disagreed in more than fifty percent of cases” (p. 13), and consequently made refinements to the rubric to remedy this.

CONCLUSION

Educational research in QL is an emerging field, and existing literature is relatively sparse. Further, purely theoretic or anecdotal perspectives mark many of the available resources. The works of Boersma et al. (2011), Grawe et al. (2010), and Boersma and Klyve (2013) are the most specific and scientific research studies relevant to this research thesis. Very little research has attempted to connect QR specific abilities or HoM with QR courses aimed at promoting these constructs in students.

This research aims to narrow questions raised by Dingman and Madison (2010B) by framing investigations within the HoM construct. This research expands on the work of Boersma et al. (2011) with its application of the QLAR to course texts. This research also expands on the work of Boersma and Klyve (2013) through qualitative analysis of student interviews that heavily involve student interactions with Boersma and Klyve’s (2013) prompt-less instrument.

The conclusion of this research should point to possible QR course improvements through comparisons of existing courses. This research should also point to further study in the areas of HoM evaluation, QR course evaluation and obstacles students face in developing their ability in QR and/or HoM.

CHAPTER 3: METHODOLOGY

RESEARCH DESIGN

This chapter outlines the research methods used. First, the researcher describes the population and sample under study. Then, the researcher describes the instruments used in the analyses. Next, the researcher discusses the qualitative methods used, including: structured, task-based interviews; identification of contextual themes; and application of the QLAR from Boersma et al., 2011). Analyses were conducted through a combination of pre/post testing, artifact analysis, classroom observation, and case study. Further, the qualitative discussion includes a description of the three interview portions that served overlapping purposes. This was viewed under the lens of critical constructivism; individual constructions were considered alongside classroom culture/norms.

Pre/post testing utilized Boersma and Klyve's (2013) prompt-less instrument to measure the effects of the courses on habit of mind. This also gave the researcher an idea of a baseline habit of mind for students in the research population. Further, students were identified for continued investigation through case study from the scoring of the habit of mind instrument. The Quantitative Literacy Assessment Rubric (QLAR) was applied to curricular materials/assessments as a form of artifact analysis. Classroom observations informed the researcher on common instructional practices and opportunities for engagement with curricular materials. Case study through structured, task-based interviews investigated the relationship between responses on the habit of mind instrument, the students' actual level of habit of mind and the students' proficiency in QR – as measured by course grades and responses to generic QR course questions. This provided a glimpse at the varying strategies students employ when met with quantitative information and/or curricular materials/assessments, and helped identify any themes in students' habit of mind. Student held attitudes and beliefs about specific problem types/assignments were contrasted with those of the curriculum designers (assessed through formal interviews). This was a mixed methods study although it was primarily qualitative. The following section focuses on quantitative methods for the study.

POPULATION AND SAMPLE

Primarily non-STEM undergraduates at NAU and the UofA made up the population for this study. Specifically, the NAU population consisted of students who were/are enrolled in MAT 114. The UofA population consisted of students who were/are enrolled in MATH 1313 and/or MATH 2183. These students came from a variety of backgrounds and enroll in a variety of generally non-STEM fields of study. The majority of these students were traditional, young adult students with some exceptions. It is not uncommon to encounter non-traditional students who may or may not already possess an undergraduate degree. An implicit assumption of this project was that these populations are relatively similar across the two universities since both are similar in size and standards for admission. One drawback to this assumption was that the two universities vary in offered courses of study and geographical location. Also, more students enroll in MAT 114 than the QR courses at the UofA since MAT 114 is the primary course that satisfies the liberal arts credit in mathematics whereas MATH 1313 is offered as an alternative to College Algebra, and MATH 2183 has College Algebra or MATH 1313 as a pre-requisite. Further discussion is included in the Limitations section of this report.

Samples for this population came from the students enrolled in the courses under study. At the beginning of the semester students from several course sections were asked to fill out the prompt-less instrument (PLHOM) via an online link sent out through email by their course instructor. Ninety eight subjects participated in the study by responding to the PLHOM – eighty three students from MAT 114, 13 students from MATH 2183, and one student from MATH 1313. The sample from MAT 114 is likely representative of that student population, the sample from MATH 2183 is likely not completely representative of that student population due to the low sample size, and the sample from MATH 1313 is definitely not representative of that student population due to the low sample size.

INSTRUMENTATION

The primary instrument for the quantitative portion of this project is the prompt-less instrument for assessing QR habit of mind developed by Boersma and Klyve (2013). This instrument consists of several open-ended questions (prompts) coupled with a news article. Choice of articles depend on their single page length, abundance of quantitative information relevant to the central theme of the article, and their inclusion of a quantitative visual aide (e.g. line graph, pie chart, etc…).

The students are prompted to read an article in an electronic format – viewable in Appendices A and B - and respond to questions below. The instrument contains several identifying questions to keep a record of the students' name, instructor, and number of attempts in the course. The prompts are open-ended and not multiple choice; they are mostly quantitatively neutral in that they do not explicitly draw attention to quantitative information contained in the articles and they do not refer to any specific calculations. The questions are paired with a response box that allows students to type as much or as little as they would like in their responses. The questions are given as follows:

1. Did you understand the article?
2. What was the main point(s) of the article?
3. What facts did the author use to support the main point(s)?
4. Were there any particular strengths or weaknesses in how these facts were reported?
5. Does the graph help interpret the numerical information found in the text? Explain your thoughts.

These five questions are identical to the questions contained in the instrument that Boersma and Klyve detail in "Measuring Habits of Mind: Toward a Prompt-less Instrument for Assessing Quantitative Literacy". In this paper, Boersma and Klyve (2013) explain how the Quantitative Literacy Assessment Rubric (QLAR) can assess individual student work in the context of the prompt-less instrument. In order to grade the questions, Boersma and Klyve's (2013) process is adapted to the chosen articles.

The first question, "Did you understand the article", identifies language difficulties. Responses are removed if a student identifies they have trouble understanding the written language (English). The second question, "What was the main point of the article", measures how well a student can recognize the main point of the article. The last three questions provide the bulk of the data obtained from this instrument.

The third question, "What facts did the author use to support the main point(s)" assesses habit of mind in regard to interpretation, the "ability to glean and explain mathematical information presented in various forms" (Boersma et al., 2011, p. 9). High scoring students "must (1) have the habit of mind to seek out quantitative information in the article, (2) identify relevant and specific information, and (3) communicate these facts" (Boersma and Klyve 2013, pp. 7-8). The following rubric from Boersma and Klyve (2013) is used to score this question (sample student solution in italics – not from Boersma and Klyve):

0. No quantitative information given or alluded to. *Studies and research done on the past about incarceration*.
1. Some relevant quantitative information is identified (or alluded to), but none is correct (or specific enough to be judged correct or incorrect). *Incarceration rates with crime rates and studies that were researched.*
2. Some relevant and correct information is identified, but not all. *The author states issues with this theory by saying, for example, that incarceration rates were increasing for years before crime started going down. He also uses the fact that criminologists now tend to believe that incarceration accounts for a fraction of the drop in crime, like 25%*.
3. All relevant quantitative information is correctly identified. *Studies claiming 58% of the crime drop during the 1990s was due to incarceration were based on old, incomplete data. Also, as incarceration goes up, its general effectiveness decreases, and incarceration rates were up years before the crime rates began to decrease. Criminologists now only attribute close to 25% of the drop in crime to increased incarceration and this effect is even less so when considering violent crime.*

This question only prompts for supporting facts and does not ask specifically for quantitative facts. The instrument ignores non-quantitative comments since it aims to measure a habit of mind towards quantitative reasoning (Boersma and Klyve 2013, pp. 7-8).

The fourth question, "Were there any particular strengths or weaknesses in how these facts were reported?" measures students' propensity to consider quantitative information when determining the effectiveness of an argument. This question falls more in the Analysis/Synthesis competency from the QLAR. The following rubric from Boersma and Klyve (2013) is used to score this question (sample student solution in italics – not from Boersma and Klyve):

0. No strength or weakness identified or, if identified, not supported with quantitative reasoning. *Yes, they were not in depth just told, so it was weak.*
1. A strength or weakness is identified but is supported with incorrect quantitative reasoning (or the reasoning is not specific enough to be able to judge correctness). *There were not a lot of statistics within the writing itself, but there were graphs provided to visualize the data*.
2. A strength or weakness is identified and is supported with quantitative reasoning, but the reasoning is incomplete (e.g., it contains unsubstantiated claims). *One strength that supported this article was all the percentages and time periods that were recorded throughout the past 25 years based on crime rates. But a weakness in this article was the statistics because one author claims that 58% of violent crimes were do to incarceration, but who are the people he is basing this percentage on? Is this number based on the entire world, U.S., all criminals ever recorded? The fact is very vague.*
3. A strength or weakness is identified and supported with correct and complete quantitative reasoning. *A strength is that they reported the original statistic that led people to believe incarceration reduces crime, 58% of the drop in crime attributable to increased incarceration, and followed up with more recent and complete studies that found only around 25% of this reduction attributable to increased incarceration. A weakness is the lack of information/statistics involving actual rates of incarceration. There is no way to*

understand the scope of the changes in incarceration rates over the time period discussed in the article.

Again, the fourth question does not prompt readers to respond with quantitative information, and any responses that do use quantitative information should indicate a clear habit of mind (Boersma and Klyve 2013, pp. 8-9).

The fifth question, “Does the graph help interpret the numerical information found in the text? Explain your thoughts”, measures the “students habit of mind to check the numerical information in the text with the numerical information being presented graphically” and the students’ propensity/ability to critically reflect on the figures’ usefulness in supporting the author’s points (Boersma and Klyve 2013, p. 9). The following rubric from Boersma and Klyve (2013) is used to score this question (sample student solution in italics – not from Boersma and Klyve):

0. No indication that the numbers in the article were checked against their representation in the graph. *The graphs are clear and easy to interpret*.
1. Claims, with no justification, or incorrect justification, that the graph does or does not accurately present the numerical information in the article. *Yes, it clearly shows that the increased incarceration rate decreased the crime rate.*
2. Claims, with justification, that the graph does or does not accurately present the numerical information found in the text. *I do not feel the images helped the information as crime did not seem to go down as drastically as it seemed in the article; some crime actually appeared to rise.*

3. Correctly points out a specific discrepancy between the graphical presentation and the quantitative information found in the text. *Not really, the graphs are not on the same scale as the percentages posted within the writing.*

This rubric identifies the students who check to see if information in the article aligns with the information presented in the graphs. Question 5 is scored a second time utilizing the following rubric (sample solution in italics – not from Boersma and Klyve):

0. No strength or weakness of the graph identified or, if identified, not supported with quantitative reasoning. *Yes, it helps the reader be able to quickly compare and contrast different factors and data.*
1. A strength or weakness of the graph is identified but is supported with incorrect quantitative reasoning (or the reasoning is not specific enough to be able to judge correctness). *No it does not, the text is wrong because and numerical is right. It said that it had decreased by 25 percent which it had not in the graph.*
2. A strength or weakness of the graph is identified and is supported with quantitative reasoning, but the reasoning is incomplete (e.g. it contains unsubstantiated claims). *Kind of, the second one shows homicide actually increasing, while others decreased and they are talking about violent crimes.*
3. A strength or weakness of the graph is identified and supported with correct and complete quantitative reasoning. *The figures weakly support the author's point. The first figure appears to support the point that increases in incarceration results in diminishing returns since the effectiveness of incarceration goes down. However, the measure of 'effectiveness' is not discussed/defined in the article, and the figure itself is supposedly*

from crime data in Texas, which is not necessarily representative of the nation as a whole.

This scoring method rewards students who take the initiative to report on strengths and weaknesses of the use of the graphs in the article. The total habit of mind score is the sum of the four individual 0-3 point scores, resulting in a final 0-12 point score for an individual.

An analysis of the quantitative data from the habit of mind instrument involves searching for differences in pre/post testing. There are several considerations when comparing pre/post scores. It is not clear that the 0-12 point habit of mind score is actually on a ratio scale. Educational research designs often ignore this limitation and, as such, this project does not aim to argue this point. There are two articles used as a basis for the habit of mind instrument and their application is partially random. At NAU, 4 instructors asked their sections of MAT 114 to complete the habit of mind instrument; 2 instructors pretested with article A and post tested with article B, while 2 instructors pretested with article B and post tested with article A. At the UofA, 6 instructors asked their sections of MATH 1313 or 2183 to complete the habit of mind instrument; 3 instructors pretested with article A and post tested with article B, while 3 instructors pretested with article B and post tested with article A. This partial randomization attempted to wash out the effects from using two different articles for pre/post testing on the individual level.

No difference in the variability (and possibly the means) in pretest scores indicated that the two articles coupled with the habit of mind questions are equally valid measures of habit of mind. Comparisons were drawn with two-tailed unpaired *t*-tests and a two-tailed paired *t*-test. Statistical analysis was done with the free online software GraphPad Software QuickCalcs' *t*-test calculator. The software can be found at the address:

https://www.graphpad.com/quickcalcs/ttest1.cfm.

QUALITATIVE METHODS

The majority of this project is a qualitative study from two primary sources – student interviews/classroom observation and article analysis of course materials. First, structured, task based interviews were conducted as follow-ups to the initial pretest for habit of mind. The intention of the follow-up interviews was three-fold; a re-introduction of the article and questions from the habit of mind instrument offered a chance to see firsthand how students interacted with the instrument, while follow-up questions allowed the researcher to probe at the level of prompting required to elicit a QR oriented response and gather information about student held beliefs about their QR course. Second, an article analysis of course materials provided a bigger picture of how well QR core competencies were represented in current curricular materials. Further, an additional article analysis identified real-world contextual themes in exercises from the texts.

STRUCTURED, TASK-BASED INTERVIEWS

Interviews were a follow-up to the pre application of the habit of mind instrument. For each course – MAT 114, MATH 1313, and MATH 2183 – all willing subjects received a follow up interview; three interview subjects were solicited through email and never completed an online version of the PLHOM. Nine subjects from MAT 114, five subjects from MATH 1313, and one subject from MATH 2183 were interviewed. Variability among subjects allowed for a closer look at how students from different levels of habit of mind might interact with QR related media and curricular materials. The interviews were structured for three primary areas of data collection – HoM evaluation, student interaction with course-like materials and the student impressions of the course. The interviews were audio recorded and the interviewer took notes for the duration.

First, the habit of mind instrument from the subject's pretest was reintroduced and the subject was asked to quickly read the article and complete the last 3 questions from the habit of mind instrument: "What facts did the author use to support the main point(s)", "Were there any particular strengths or weaknesses in how these facts were reported", and "Does the graph help interpret the numerical information found in the text? Explain your thoughts" (Boersma et al., 2013). From this, the researcher observed how the subject interacted with the PLHOM and answered any questions that the subject has. In fact, questions from subjects about the instrument shed light on potential problems with the instrument. Follow up prompting by the researcher explored how much prompting was required to point responses toward higher scores on the habit of mind instrument. For example, subjects who did not identify a QR related fact the author uses to support his main point were asked "were there any other facts the author uses to support the main point" and, absent of a QR related fact, "were there any quantitative facts the author uses to support the main point". Appendix E details the interview protocol for the initial stage of the interview. During interviews, the PI scored subjects' responses to the PLHOM using the rubric provided by Boersma et al. (2013). Initial scores dictated the type of prompting students received during the interview (Appendix E). Subjects who improved their score on the prompt-less instrument after initial prompting were further prompted according to higher level responses as time allowed. Appendix F details issues in scoring student responses to the prompt-less instrument.

The second portion of the interview investigated how students interact with course-like materials. Here, course-like materials were chosen to represent a QR oriented question/prompt from the course and any purely mathematically oriented questions supported a follow up QR oriented question. The questions in this portion of the interview may be viewed in Appendix E.

The third portion of the interview, viewable in the interview protocol (Appendix E), assessed student opinions of the course. Specifically, questions were designed to elicit responses that expose what students believe they will take away from the course, the students' ideas about the nature of QR and how they see – if at all – the differences between a course in QR and a traditional mathematics course. Upon completion of the interview, students received $20 for the hour of their time. This expense was paid for out of pocket by the PI. In addition to student interviews, several QR classes were observed, but not audio recorded, and field notes were taken with student anonymity in mind.

Qualitative analyses followed *topic-oriented* ethnographical analysis outlined by James Spradley (1980). This "narrows the focus to one or more aspects of life known to exist in the community", where – in the context of the study – the community refers to the learning community in the NAU and UofA QR classrooms, and the focus is on HoM and learning opportunities for QR/HoM (Spradley, 1980, p. 31). In line with Spradley's (1980) "verbatim principle", interviews are audio recorded and later transcribed by hand.

A first pass of the field notes from classroom observations, field notes from student interviews and the transcribed audio sought to identify patterns of behavior. These include patterns of behavior in the classroom, patterns of behavior students showed when faced with the articles in the first part of the interview and patterns of behavior students showed when faced with course-like materials in the second part of the interview. Here, it was beneficial to identify semantic relationships that students use to give meaning to QR materials and relate these to the habits they show when faced with QR materials (Spradley, 1980, p. 89). Specifically, the researcher sought to identify semantic relationships in the *analytic domains* related to QR and HoM. This process represented an "*in-depth investigation*" of classroom/learner culture and individual's HoM (Spradley, 1980, p. 101).

Next, the researcher began forming taxonomic relationships between recurring terms and actions from classroom and interview data. Finally, all of the data was revisited and the researcher sought to identify cultural themes – "any principle(s) recurrent in a number of domains, tacit or explicit, and serving as a relationship among subsystems of cultural meaning" (Spradley, 1980, p. 141). This was to attempt to form generalities in HoM and student interaction with QR courses. Spradley (1980) stated this involves a search for "those cognitive principles that appear again and again" (p. 144). From this, the researcher drew conclusions about commonalities and differences in the experiences, impressions, opportunities to develop HoM and opportunities to engage in QR between the courses at UofA and NAU.

ARTICLE ANALYSIS OF COURSE MATERIALS

Article analysis of course texts and sample assignments/projects identified opportunities provided by the texts for students to engage in QR core competencies outlined by the QLAR and opportunities for students to engage with truly contextual – real world – applications of QR. Questions from *Case Studies for Quantitative Reasoning* by Madison et al. are already coded by the QLAR rubric by Boersma et al. (2011). The researcher similarly coded MAT 114 exercises from the course text, *Quantitative Reasoning* by Matthew Fahy and Gina Nabours in order to draw comparisons. Further, contextual/content themes from the two texts were identified in order to give substance to course descriptions.

CHAPTER 4: ANALYSIS

INTRODUCTION

Article analyses of course texts for MAT 114, MATH 1313 and MATH 2183 seek to inform about HoM opportunities students encounter based on curricular materials. The first article analysis describes topic organization and the real-world, contextual themes available for development through exercises in course texts, and connects this with the QR conceptualization under which each course was developed. The content-out approach in the MAT 114 course text aligns with the mathematical conceptualization of QR, while the context-out approaches in MATH 1313/2183 align with the integrative conceptualization of QR. An additional article analysis identifies core competencies required by a sample of problems from the MAT 114 and MATH 1313 course texts, and compares these with each other and with a similar analysis of the MATH 2183 course text conducted by Boersma et al. (2011).

Interview analyses seek to inform about HoM opportunities students encounter in each course based on student impressions of course-like materials and student impressions of QR courses. Further, the interview analyses seek to inform about HoM exhibited by these students and identify obstacles to student application of HoM. This includes identification of areas in students' lives where they apply and practice a HoM. An additional quantitative analysis of the pre/post testing with the prompt-less instrument (PLHOM) from Boersma and Klyve (2013) seeks to inform about HoM exhibited by a larger cohort of students and identify any possible gains in students' HoM as a result of their completion of a QR course. Other than this the analyses contained herein are qualitative in nature.

ARTICLE ANALYSIS – TOPIC ORGANIZATION AND CONTEXTUAL THEMES

Topic organization in MAT 114 versus MATH 1313 and MATH 2183 illustrates an important distinction between QR philosophies behind the courses at NAU and UA. MAT 114 exhibits a traditional course design with course modules arranged by building mathematical content (content-out). Mathematical content drives the curricular framework and contextual or real-world forays occur only as the developed mathematical content allows. In contrast, UA course developers take a context –out approach to curriculum development. MATH 1313 course text authors Ethan Bolker and Maura Mast describe this approach in the text's preface; "Each chapter starts with a real story that can be best understood with careful reading and a little mathematics" (Bolker and Mast, 2016, p. xiii). Further, in the introduction to the MATH 2183 course text, authors Madison et al. (2012) state "every QR problem is a contextual problem" and that "every mathematical or statistical topic investigated is one that is contained in or is useful in critiquing a public media article" (p. v). This context-out approach makes curricular progression in the UA courses difficult to elucidate and this progression is further muddied by subtle differences in curricular focus between MATH 1313 and MATH 2183. Example course progressions with problems typical to student experiences are provided to better compare and contrast the curricular structure underlying the three QR courses.

MAT 114 consists of 13 course modules that may be further subdivided into what the PI sees as three main course progressions:

- Basic statistics and the central limit theorem (BSCLT)
- Linear/exponential models and finance (LEMF)
- Graph theory and scheduling (GTS)

Table 2 illustrates the relationship between course module, progression, and content topic. The problems shown in Figures 4, 5, 6, and 7 occur in class exercises or homework assignments encountered by every MAT 114 student. They represent typical problems students encounter throughout the course and highlight the first course progression (BSCLT) in MAT 114.

Table 2. Course module, topic, and progression.

Module	Topic	Course Progression
1	Basic statistics (population, sample, parameter, statistic, and sample size)	BSCLT
2	Stem and leaf plots, histograms, outliers and skewedness	BSCLT
3	Center and spread of a data set (standard deviation and the 5 number summary)	BSCLT
4	Normal distributions and z-scores	BSCLT
5	Introduction to probability (random, disjoint, and independent events probability)	BSCLT
6	Probability – further topics (multi-step probability calculations and expected value)	BSCLT
7	Margin of error and calculating confidence intervals (central limit theorem)	BSCLT
8	Fundamentals of functions and models (linear models and the correlation coefficient)	LEMF
9	Exponential Models	LEMF
10	Compound interest (savings and loans formulas)	LEMF
11	Multi-step compound interest problems and basic amortization tables	LEMF
12	Introduction to graph theory (Basic terminology, Euler circuits, and Hamiltonian circuits)	GTS
13	Graph representations and scheduling	GTS

The first module introduces the five vocabulary words *population, sample, parameter, statistic,* and *sample size.* The first module also develops some information on statistical bias and explores the idea that samples may or may not be representative of a population. There are several short answer type exercises included in the module, however the problem type in Figure 4 appears 4 times in written materials for the module. Module 2 introduces students to several data visualizations including stem and leaf plots, bar graphs, pie charts, and histograms. By the time students have completed the *Looking Forward* section of Module 2 on page 25 of the MAT 114 workbook, they should be able to calculate a 5 number summary and standard deviation by hand from a set of raw data. Module 3 adds to this with the rule of thumb calculation that "approximately 95% of the data falls within two standard deviations of the mean" (Fahy and Nabours, p. 32). Figure 5 is a problem from the Module 3 exercises (completed in class); while other problem types might better represent Module 3, this problem better illustrates the BSCLT course progression and segues into the development of the standardized z-score for a normally distributed set of data (Module 4).

HW 1.2. An NAU student organization recently conducted a poll in an effort to gauge student support for increased campus transit. Surveys were conducted at three locations on campus – outside the [Health and Learning Center], in the student union, and outside Cline library. Of the 324 students who were interviewed, 81 stated they would be in favor of an increase in student fees to pay for additional campus buses.

Match each vocabulary word below with its *value* or *description* on the right.

____ Population	a) 324
____ Sample	b) 81
____ Parameter	c) 25%
____ Statistic	d) Campus buses
____ Sample size	e) All NAU students
	f) The students who were interviewed
	g) The percentage of *all* NAU students in favor of increased student fees to pay for additional buses

Figure 4. Module 1 HW problem 2. (Fahy and Nabours, 2016, p. 7)

After exploration of normal distributions and z-scores in Module 4, the course modules make a horizontal move to develop some basic probability theory. Figure 6 represents much of the material covered in Module 5 with three questions designed around basic theoretical probabilities, a multiplicative rule for problems involving "and", and basic conditional probabilities respectively.

EX 3.5. Two brands of computer printers were tested by a consumer group. For each brand, a sample of 50 printers was tested. Researchers measured the number of days until the printer experienced a significant malfunction. The mean and standard deviation for each sample are stated below.

Brand A	**Brand B**
$\bar{x} = 71$	$\bar{x} = 68$
sd $= 21.7$	sd $= 6.2$

For each brand, calculate an approximate range that captures 95% of the data set for that brand.

Based on these samples, which brand would you prefer to buy? Why?

Figure 5. **Module 3 Exercise 3.5. (Fahy and Nabours, p. 33)**

EX 5.6. Suppose that all that's left of a bag of Skittles are three of each color (yellow, red, orange, green, and purple).

If you take two out of the bag to eat, what is the probability that the second Skittle is red given that the first was orange?

If you take three out of the bag, what is the probability they are all purple?

If you take one Skittle out of the bag, what is the probability it is yellow given that it is not green?

Figure 6. **Module 5 Exercise 5.6. (Fahy and Nabours, p. 64)**

Module 6 builds on the ideas from Module 5, develops the relationship between sample size and empirical/theoretical probabilities, and connects this with expected value. Once the *Looking Forward* section of Module 6 is completed, students have defined margin of error and used it to produce 95% confidence intervals. Module 7 formalizes this with a discussion of the Central Limit Theorem. Module 7 is the capstone of the BSCLT course progression; Figure 7 represents a capstone type problem in Module 7. Problems like this are likely encountered on exams, and instructors encourage students to interpret a 95% confidence interval (CI) as “we are 95% confident that the [insert specific parameter description] lies between [95% CI lower bound] and [95% CI upper bound].

> **HW 7.2.** A recent report released by the Coconino County Sustainability Commission found that, of the 225 households studied, 117 had outdated, inefficient major appliances. Construct and interpret a 95% confidence interval for the true percentage of households in Coconino County [that] have outdated, inefficient major appliances.
>
> Coconino County can apply for federal funding to implement sustainability programs if they can show that the *majority* (that is, more than 50%) of all households in the county have outdated, inefficient major appliances. Does the study described above support this claim?

Figure 7. Module 7 HW problem 2. (Fahy and Nabours, p. 92)

These problems illustrate the content driven development of the basic statistics and probability theory involved in MAT 114’s first course progression towards the Central Limit Theorem. Note that though context is inserted, it is clear that the mathematical content drives the progression of topics (i.e. basic statistics and probability problems segue into problems utilizing results of the central limit theorem).

Course progressions for MATH 1313 and MATH 2183 are not as clear cut and more nuanced. To illustrate this and further distinguish content-out from context-out course design in QR, example MATH 1313 problems are shown below in Figures 6, 7, 8 and 9. These are chosen to represent typical problems encountered in the course and to show some of the many themes that could be developed by a MATH 1313 instructor. All of the following problems were either explored in class or assigned to students enrolled in Dr. Bernard Madison's section of MATH 1313 during the 2017 spring semester. They were also recommended to other course instructors by Dr. Madison near the beginning of the semester. It is important to note that the course progression represented by these problems is only a possibility, and it is plausible that some sections used different questions entirely and developed alternative themes. Also, course progressions in MATH 1313 are flexible, non-linear and often overlapping; for example, a single exercise about a state's taxes could provide practice with course themes like extrapolating data, fact checking, finding information, dealing with large numbers, estimation, and understanding percentages. Themes do not progress linearly in that they only appear when useful for a contextual problem/situation. For example, a theme about conservation could include problems dealing with hybrid vehicles; however, connected problems, as seen below, can crop up months apart during course.

Exercise 2.9.14. [W][S][Section 2.2][Goal 2.1] Hybrids vs. nonhybrids: The 5-year equation.

On February 23, 2011 Matthew Wald blogged at *The New York Times* about a study in *Consumer Reports* saying that

> A car buyer who lays out an extra $6,200 extra [sic] to buy the hybrid version of the Lexus RX will get the money back in gas savings within five years, according to Consumer Reports magazine, but only if gasoline averages $8.77 a gallon. Otherwise, the nonhybrid RX 350 is a better buy than the Hybrid 450h. [R57]

Wald notes that the study assumes

- the car will be driven 12,000 miles a year.
- gas will cost $2.80 a gallon.
- the hybrid gets 26 miles per gallon, the nonhybrid, 21.

a) Show that the computation is wrong – that at $8.77 per gallon of gas you can't save $6,200 in 60,000 miles of driving.
b) Show that you can save that much with that much driving if gas costs $8.77 per gallon more than $2.80 per gallon.
c) The *Times* blogger was reporting on a study from *Consumer Reports* magazine. Do you think the error was the blogger's, or the magazine's? What would you have to do to find out which?
d) Write a response to post as a comment on the blog.

Figure 8. Exercise 2.9.14 (Bolker and Mast, 2016, pps. 41-42)

Figure 8 depicts an example problem from chapter 2 of *Common Sense Mathematics*. The [W] indicates that the problem is "worthy" and good for working in class; the [S] indicates the problem is included in the solution manual. Goal 2.1 is "explicitly manipulate units in expressions" (Bolker and Mast, 2016, pps. xv, 23). Further, the problem may be placed within several themes a course like MATH 1313 can develop: personal finance, extrapolating data, fact checking, conservation (subtopic: hybrid vehicles), rates in the media, and argument in writing. The problem in Figure 9, generally assigned around a week or two after the problem in Figure 8, fits within the title of chapter 3, "Percentages, Sales Tax and Discounts" (Bolker and Mast, 2016). Additionally, it builds on themes of extrapolating data, fact checking, rates in the media, understanding percentages, dealing with large numbers, and facts and figures in public policy (subtopic: taxes). Note that [C] refers the complexity of the problem and Goal 3.3 is to "master strategies for deciding how to arrange percentage calculations" (Boker and Mast, 2016, pps. xv, 57).

Exercise 3.10.7. [S][C][W][Goal 3.3] New taxes?

On January 14, 2013 *The Boston Globe* reported that Massachusetts could raise $1 billion a year by increasing the income tax rate from 5.25 to 5.66 percent. [R89]

a) Find the 2013 taxable income in Massachusetts.
b) Find the total revenue from this income tax at the 5.25 percent rate.
c) Compare your answer to the state budget. Are the numbers consistent?

Figure 9. Exercise 3.10.7 (Bolker and Mast, 2016, pps. 66-67)

The next example from chapter 5 of *Common Sense Mathematics* signals a significant jump in time for students. Figure 10 illustrates an example students would likely encounter at least 3 weeks after the problem in Figure 9. In this time students have continued to explore other themes in the course and have more practice with evaluating arguments, dealing with percentages (including taxes), measuring and dealing with average values, indices, and calculations involving inflation. Note that Goal 5.1 is to "compute means using weighted averages" and Goal 5.2 is to "investigate what it takes to change a weighted average" (Bolker and Mast, 2016, p. 99).

The exercise in Figure 10 continues the themes dealing with large numbers, extrapolating data, understanding percentages, rates in the media, and facts and figures in public policy (subtopic: unemployment); the only new layer involved is seeing the unemployment rate as a national average, however, weighted averages do form a critical component of good responses to this exercise.

Exercise 5.7.7. [S][Section 5.3][Goal 5.1][Goal 5.2] Five million unemployed.

In *The Hightower Lowdown* (Volume 12, Number 5, May 2010) you could read

- **5 MILLION PEOPLE** (about 10% of the workforce are out of work).
- **UNEMPLOYMENT IS HEAVILY SKEWED BY CLASS**. Among the **wealthiest 10%** of American families (incomes above $150,000), only **3% are unemployed-** a jobless rate that rises as you go down the income scale. Among the **bottom 10%, more than 30% are out of work**. [R155]

What average unemployment rate for the middle 80% of families fits with the given values for the top and bottom 10% to work out to the overall (weighted) average unemployment rate of 10%?

Figure 10. Exercise 5.7.7. (Bolker and Mast, 2016, p. 107)

Figure 11 too signals a significant jump in time since students would likely encounter the problem about one month after the problem in Figure 10. In this time, the students continue to gain experience with averages, explore statistics and distributions in a variety of contexts (many dealing with facts and figures in public policy), and develop ideas about modeling with linear functions (based around applications involving electricity bills and income taxes). Note that Goal 7.3 is to “stress the meaning and units of the slope and intercept” and parts/goals related to work with spreadsheets are omitted (Bolker and Mast, 2016, p.151). Exercise 7.8.11 (Figure 11) continues course themes like personal finance, conservation (subtopic: hybrid vehicles), argument in writing, extrapolating data, facts and figures in public policy (subtopic: tax rebates), dealing with large numbers, and rates in the media. Exercise 7.8.11 (Figure 11) could represent a capstone question since it is about a specific topic previously explored in Exercise 2.9.14 (Figure 8), and the new mathematical concept (linear modeling) is motivated through added depth in questioning and context. Further, the problem builds on skills developed previously in the course (e.g. units and conversions, discounts, and appropriate precision).

It is important to stress that the course themes laid out for MATH 1313 are possibilities and there are no data to support that the connections drawn here were also made in practice. The multitude of possible themes reflects the flexibility of the course and the versatility of the exercises in *Common Sense Mathematics*. Ultimately, course instructors are free to follow their own themes and explore contexts of their choice; however, MATH 1313 instructors typically move linearly through the textbook and because these exercises are recommended by the course coordinator, it is likely that most MATH 1313 students encounter these problems.

Exercise 7.8.11. [S][Section 7.5][Goal 7.3] Hybrid payback.
The "Best & Worst Cars 2011" issue of *Consumer Reports* provides the following data for the new Toyota Camrys:

	conventional	hybrid
cost	$19,720	$26,575
fuel economy	26 MPG	34 MPG

Assume gasoline costs $3.50/gallon.

1. Questions about the conventional Camry.
   - i. Once you own the car, how much does it cost to run, in dollars per mile? Does your answer make sense?
   - ii. Calculate the total cost (purchase plus gasoline) to drive the conventional Camry 10,000 miles.
   - iii. Write the linear equation that computes the total cost *C* of driving the conventional Camry *M* miles.
   - iv. Identify the slope and the intercept of this equation, with their units.

   [Parts b-c OMITTED]

a) If you drive 120,000 miles will you recover in gas savings the extra initial cost of the hybrid? Write a complete sentence or two and use appropriate precision for the numbers you use to make your argument.
b) When will you recover the extra initial cost in gas savings if the government (re)instates a $3,000 tax rebate for hybrid purchases?
c) With the original initial costs, how much would the price of gasoline have to be in order for the breakeven point to occur at 30,000 miles?

   [Part g OMITTED]

Figure 11. Exercise 7.8.11 (Bolker and Mast, 2016, p. 170)

MATH 2183 is similarly structured around a context-out approach; the key differences in student experience between MATH 1313 and MATH 2183 lie in the richness of contexts encountered by students. Context in *Common Sense Mathematics* is noticeably truncated. Exercises often involve article blurbs or quick statements and questions therein are often specific and leading. *Case Studies for Quantitative Reasoning: A Casebook of Media Articles* (3rd ed.), authored by Boersma et al. (2012) is the course text for MATH 2183. Though already described as a context-out course, a better description would be media-out since all of the material for the course is designed around case studies of media articles. *Case Studies* (3rd ed.) is organized by QR topics, but study questions, the text's primary focus, are contextually grounded by and specific to the media (article, ad, graphic, etc…) in the 30 case studies. Described below are several study questions and associated case study materials. Most MATH 2183 students encounter these questions either in class or in homework, and the questions represent several themes an instructor could develop in a semester. Figure 12 shows three study questions from Case Study 1.1 in *Case Studies* (3rd ed.). Case Study 1.1 focuses on David Leonhardt's article "What $1.2 Trillion Can Buy". He writes, "The human mind isn't very well equipped to make sense of a figure like $1.2 trillion", the author's estimate for the cost of the war in Iraq. In the article, he justifies his estimate as a conservative compromise between the "two best-known analyses of the war's costs", and also explains the hidden opportunity costs of the war. Leonhardt explores several ways to cope with and understand the magnitude of a number like $1.2 trillion, citing that it would "pay for [both] an unprecedented public health campaign" and an enactment of 9/11 commission national security recommendations. He later contrasts the war's yearly cost of $200 billion with the cost of treating heart disease and diabetes, universal preschool, and what it might take to "make a real difference" in Afghanistan (Leonhardt, 2007). The questions in Figure 10 explore these claims and figures.

Often, one of the first real exercises students encounter in MATH 2183 is Case Study 1.1 (Figure 12). This marks the possible beginning of several course themes: understanding large numbers, units and conversions, extrapolating data, understanding rates/percentages, expressing rates/percentages in writing, critical summary, and HoM (subtopic: representing data). In contrast to questions in *Common Sense Mathematics*, the added complexity of a full media article sets apart the case studies in MATH 2183. Students see full articles, and must seek out relevant information on their own or with classmates. A full article also provides more substance for stemming inquiry, and course instructors can aim to draw students along more complex lines of reasoning. MATH 2183 is more writing focused than MATH 1313 and the writing based themes, critical summary, argument in writing, and quantification in writing, pervade the exercises throughout *Case Studies* (3rd ed.). Quantification in writing is so central to almost every course exercise that it is mostly omitted in these analyses to avoid redundancy.

“What $1.2 Trillion Can Buy” by David Leonhardt

New York Times

January 17, 2007

This article discusses how to make sense of a large quantity, specifically $1.2 trillion. Understanding such a quantity depends heavily on what one understands beforehand, namely the cost of other items.

[1, 3, OMITTED]

1. **How many $150,000 homes would the $1.2 trillion buy? How can we make sense of this number of $150,000 homes? Determine an appropriate measure (e.g., population of a particular state or city) that would help someone make sense of the number of $150,000 homes that $1.2 trillion could buy.**

4. **Compare the writer’s estimated cost of the war to Lawrence Lindsey’s estimate using a ratio and a percent. Use each comparison in a sentence.**

5. **Using Scott Wallsten’s estimate, how many days of funding the war in Iraq would produce a dollar amount equal to the annual budget of the National Cancer Institute?**

6. **Choose some measure of the size of $1.2 trillion (other than homes) that helps you understand its magnitude and express the $1.2trillion in your measure. The answer to #2 is an example of this. Explain why this measure is meaningful to you.**

Figure 12. Case Study 1.1 Study Questions (Madison et al., 2012, p. 8)

Students would likely encounter the next example (Figure 13) early in MATH 2183 as well. It can be linked to many course themes such as understanding rates/percentages, critical summary, fact checking, HoM (subtopic: understanding errors), facts and figures in public policy (subtopic: taxes), and argument in writing. This case study is a collection of letters to the editor and is best summarized by the opening statement for the case study included in Figure 13. The study questions in Case Study 2.1 (Figure 13) exemplify the increased cognitive load borne from contextual richness. Answering these questions (Figure 13) requires 7 separate article analyses and a reflection/synthesis of information from the seven articles. Further, this richer context allows for a larger variation in student analyses/opinions in response to questions such as number 6 in Figure 13. In turn, classroom discussions possibly become more interesting, insightful, unpredictable, or – hopefully not – confused; however, the possibility of increased classroom discourse from increased problem context is undeniable.

Case Study 5.3 (Figure 14) represents a building and synthesis of themes within the problems from Figures 12 and 13 as well as another platform for discussion of variance in student analyses/opinion. Two histograms represent respectively Republican and Democratic visualizations of the same tax cut. The study questions (Figure 14) play on several themes likely familiar to students who likely encounter Case Study 5.3 mid to late semester: critical summary, argument in writing, understanding rates/percentages, expressing rates/percentages in writing, extrapolating data, understanding large numbers, fact checking, facts and figures in public policy (subtopic: taxes), and visualizing data. Further, the partisan context encourages variation in student analyses/opinions.

Case Study 2.1: Letters to the Editor on Tax Rates

Resource Material: Seven Letters to the Editor, *Arkansas Democrat-Gazette*, June 30, 2003, to July 15, 2003.

Learning Goals: The learning goals of this case study include critical reading of source material and performing basic calculations with percents and ratios.

The initial letter of June 30 in this sequence of seven letters was responding to a previous letter that is not included. The initial letter is from Bob Massery. Six people wrote letters concerning Mr. Massery's statements regarding the tax rates of a person who pays $5,000 in tax on a $30,000 income and a person who pays $53,000 in tax on a $200,000 income. The six letters about Massery's letter, Pierce (July 9), Stilley (July 9), Basinger (July 10), McGuire (July 14), Herrington (July 15), and VanHook (July 15), believe Massery's calculation of tax rates are not correct. Some of the letter writers make errors when they attempt to correct Massery.

[Warm Up Exercises and Articles OMITTED]

Study Questions for Case Study 2.1

Seven Letters to the Editor

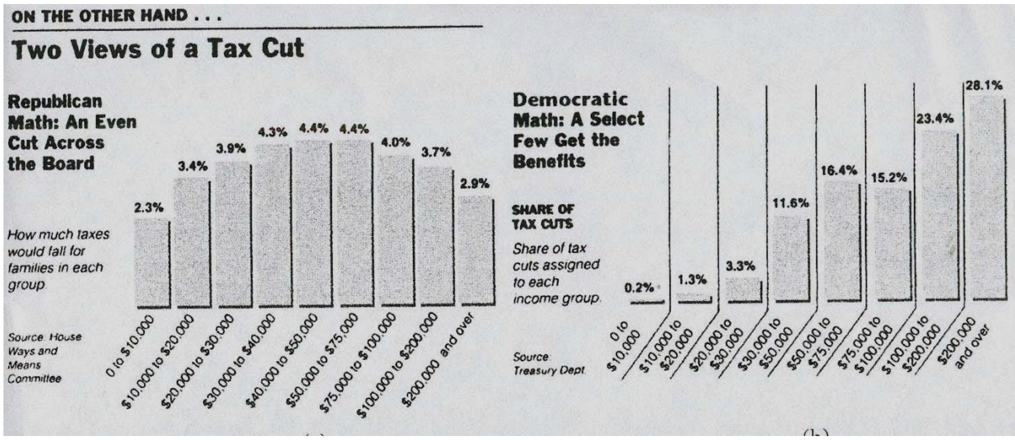
Arkansas Democrat-Gazette

June and July, 2003

1. **Create an organized list for how the tax rates for $5,000 taxes on $30,000 income and $53,000 taxes on $200,000 income are stated in the seven letters under consideration**
2. **Which of the stated rates are correct and which are incorrect? Be sure to support your conclusions with appropriate quantitative analysis.**
3. **Which, if any, of the letters dispute the amounts of tax cited by Mr. Massery: $5,000 on $30,000 and $53,000 on $200,000?**
4. **What is the mistake that Mr. Massery probably made in computing the tax rates?**
5. **Which of the six letters responding to Massery have errors, and what are those errors?**
6. **Which letter would you choose as the most appropriate rebuttal to Mr. Massery's letter? Write 2 or 3 sentences supporting your choice.**

Figure 13. Case Study 2.1 (Madison et al., 2012, pps. 28-31).

Two graphs representing the 1995 tax cut, *New York Times*, April 7, 1995.



Study Questions for Case Study 5.3

“Two Views of a Tax Cut”
New York Times
April 7, 1995

1. **Describe what the data in these two graphs represent.**
2. **Can both of these be correct? Explain why or why not.**
3. **Why is the percent decrease in taxes for families in the $10,000 to $20,000 income bracket 3.4% while these families receive only 1.3% of the tax cut? Give an example that shows this is possible.**
4. **Can the amount of the tax cut be determined from the information given in the two graphs? Why or why not?**
5. **If one assumes the tax cut is $245 billion, how much are taxes cut (in dollars) for families in the $200,000 and over income bracket? How much are taxes cut (in dollars) for families in the $20,000 to $30,000 income bracket.**
6. **Assuming the tax cut of $245 billion, what is the total amount of taxes paid by the families in the $200,000 and over income bracket before the tax cut? What is the total amount of taxes paid by families in the $20,000 to $30,000 income bracket before the tax cut?**
7. **A common argument that arises from issues related to tax cuts (as indicated by the titles of the two graphs) proceeds as follows: Republicans allude to the idea that taxes are cut approximately uniformly for all income brackets, while Democrats point to the fact that the bulk of the tax cuts go to the wealthy. How is this so? Why did people in the $20,000-$30,000 bracket get a larger (percentage) tax cut than the $200,000 and over bracket (3.9% to 2.9%) yet the people in the $200,000 and over bracket get a larger share of the tax cuts?**

Figure 14. Case Study 5.3 (Madison et al., 2012, pp. 128-130).

These analyses show the increasing contextual complexity in course texts for MAT 114, MATH 1313 and MATH 2183 respectively. Further, the way that contextual themes may be developed from exercises in the texts for MATH 1313 and MATH 2183 reflects the integrative conceptualization of QR. Mathematical topics are explored as a way to better understand compelling, real-world contexts. The MAT 114 text reflects the mathematical conceptualization of QR: contexts are added to compel the more rigid mathematical developments central to topic progression in MAT 114.

ARTICLE ANALYSIS – APPLICATION OF THE QLAR

The PI chose three chapters of *Common Sense Mathematics* (Bolker and Mast, 2016) and 5 modules from *Quantitative Reasoning* (Fahy and Nabours, 2016) to code exercises for core competencies as per Boersma et al. (2011). Chapters 5, 8 and 12 from *Common Sense Mathematics* (Bolker and Mast, 2016)) and modules 2, 3, 5, 6 and 8 from *Quantitative Reasoning* (Fahy and Nabours, 2016) were chosen because of their overlapping mathematical content. Chapter 5 from *Common Sense Mathematics* (Bolker and Mast, 2016), average values, overlaps with modules 2 and 3 from *Quantitative Reasoning* (Fahy and Nabours, 2016), which covers data visualization, average values, normal distributions and standard deviation. Chapter 8 from *Common Sense Mathematics* (Bolker and Mast, 2016), climate change – linear models, overlaps with module 8 from *Quantitative Reasoning* (Fahy and Nabours, 2016), which also covers linear models. Chapter 12 from *Common Sense Mathematics* (Bolker and Mast, 2016), break the bank – independent events, overlaps with modules 5 and 6 from *Quantitative Reasoning* (Fahy and Nabours, 2016), which covers basic probability and independent events.

In each of the chapters from *Common Sense Mathematics* (Bolker and Mast, 2016) all exercises were coded for core competencies. In-class exercises, in-class activities, homework exercises and looking forward exercises from previous modules were coded for core competencies in the modules from *Quantitative Reasoning* (Fahy and Nabours, 2016); quiz activities from *Quantitative Reasoning* (Fahy and Nabours, 2016) were not coded for core competencies since no quizzes were coded from MATH 1313, the course using *Common Sense Mathematics* (Bolker and Mast, 2016), were coded for core competencies. Data from Boersma et al. (2011) were used to help the PI correctly code for competencies as well as give data about core competencies in the course text for MATH 2183, *Case Studies for Quantitative Reasoning:*

A Casebook of Media Articles (3rd ed.) (Boersma et al., 2012) (QRCW). Note that Boersma et al. (2011) coded the second (Custom) edition of the QRCW; however, their data for the second edition is used for convenience.

Coding in this context identifies core competencies – *interpretation, representation, calculation, analysis/synthesis, assumptions* and *communication* – necessary to answer a question fully. In the event that a question consists of multiple parts, the PI treats each part as a separate question. No aspect of this coding indicates the difficulty of a problem. The PI honed his ability to code for these competencies as close to Boersma et al. (2011) as possible by first practicing on QRCW questions and comparing results with Boersma et al. (2011) and also through email conversations with one of the aforementioned paper's authors, Bernard Madison.

Interpretation, defined by Boersma et al. (2011) as the "(a)bility to glean and explain mathematical information presented in various forms (e.g. equations, graphs, diagrams, tables, words)" (p. 5), can be difficult to code in short word problems. The PI decided to code this competency in cases where a respondent must determine relevant information from irrelevant information, pull information from a graph or diagram, or interpret information in the context of other course materials. For example, Figure 15 shows exercise 5.7.7 from *Common Sense Mathematics* (Bolker and Mast, 2016). Exercise 5.7.7 meets the criteria for *interpretation* since a correct answer to this exercise requires the respondent to correctly interpret the information from the article snippet in the context of a weighted average as well as delineate relevant information from irrelevant information. Exercise 5.2 (a) from *Quantitative Reasoning* (Fahy and Nabours, 2016), seen in Figure 16, does not meet this criteria since all of the information provided is relevant and the question only asks students to re-organize the information.

Exercise 5.7.7. [S][Section 5.3][Goal 5.1][Goal 5.2] Five million unemployed.
In *The Hightower Lowdown* (Volume 12, Number 5, May 2010) you could read

- **5 MILLION PEOPLE** (about 10% of the workforce are out of work).
- **UNEMPLOYMENT IS HEAVILY SKEWED BY CLASS.** Among the **wealthiest 10%** of American families (incomes above $150,000), only **3% are unemployed** – a jobless rate that rises as you go down the income scale. Among the **bottom 10%, more than 30% are out of work.** [R155]

What average unemployment rate for the middle 80% of families fits with the given values for the top and bottom 10% to work out to the overall (weighted) average unemployment rate of 10%?

Figure 15. Exercise 5.7.7 from *Common Sense Mathematics* (Bolker and Mast, 2016, p. 107)

Several NAU graduates were honored at an Alumni Banquet as part of Homecoming festivities. The banquet program contained the following information about the honorees:

twelve earned Business degrees from NAU; seven of these 12 have gone on to earn a higher degree while the other five have started their own businesses;

eight earned Forestry degrees from NAU; six of these eight have gone on to earn a higher degree;

ten earned Psychology degrees from NAU; three of these 10 have started their own businesses;

five earned Hotel and Restaurant Management degrees from NAU; four of these five have started their own businesses.

Before calculating probabilities within this context, it may be helpful to organize the information given above in a table:

[Additional parts OMITTED]

Figure 16. Exercise 5.2 (a) from *Quantitative Reasoning* (Fahy and Nabours, 2016, p. 63)

Representation, defined by Boersma et al. (2011) as the "(a)bility to convert information from one mathematical form (e.g. equations, graphs, diagrams, tables, words) into another", can be difficult to code in situations where converting mathematical forms is the best solution to a problem, but it is not actually required. For example, homework question 2 from module 8 of *Quantitative Reasoning* (Fahy and Nabours, 2016), viewable in Figure 17, is not coded for representation even though many students will use a linear model to solve it; the question is still solvable as a repeated simple calculation without the need for a linear model. Further, questions receive the *representation* code any time they require the respondent to input data into a spreadsheet and create a graphic, or any time a question requires the respondent to reorganize information in a table, histogram, etc….

> Over the past several years, the total number of master's degree recipients at Northern Arizona University (NAU) has increased by about 12 recipients each year. If there were 196 master's degree recipients in 1997 and this trend continues, how many master's degree recipients will be expected in 2016?

Figure 17. Homework question 2 from module 8 of *Quantitative Reasoning* (Fahy and Nabours, 2016, p. 105)

Calculation, defined by Boersma et al. (2011) as the "(a)bility to perform arithmetical and mathematical calculations" (p. 5) is fairly straightforward to code. Initially, the PI incorrectly coded problems that require the use of a spreadsheet as *calculation*, however, the PI revised this practice, and generally questions of this type that were miscoded as *calculation* were re-coded as *representation.* Figure 17 contains an example of a question that is coded for *calculation.*

Analysis/synthesis, defined by Boersma et al. (2011) as the "(a)bility to make and draw conclusions based on quantitative analysis" is generally indicated by a question requiring the responded to make some choice or conclusion based on a quantitative analysis, or can be indicated by a question requiring the respondent to summarize the results of several quantitative analyses and make some generalization from this. Figure 18 is a simple example of a question that codes as *analysis/synthesis* from *Quantitative Reasoning* (Fahy and Nabours, 2016). This question requires a respondent to make a decision about whether or not two events are independent.

> If two fair six-sided dice are rolled simultaneously, would the following pairs of events be disjoint?
>
> *rolling at least one three* and *a sum of three* ________________
>
> *a sum greater than 10* and *rolling two of the same number* ________________
>
> *rolling exactly one odd number* and *a sum less than four* ________________

Figure 18. Exercise 5.1 from *Quantitative Reasoning* (Fahy and Nabours, 2016, p. 62)

Assumptions, defined by Boersma et al. (2011) as the "(a)bility to make and evaluate important assumptions in estimation, modeling, and data analysis" is only difficult to code in situations where it is not clear whether the assumptions required by the question need to be evaluated. Figure 19 is an example of an exercise from *Common Sense Mathematics* (Bolker and Mast, 2016) that the PI originally coded for *assumptions* but later revised as *analysis/synthesis* since the required assumptions from the respondent are not evaluated specifically. A full response to the exercise in Figure 19 requires the respondent to make an informed assumption that might explain an observed correlation, hence why it is coded as *analysis/synthesis*. Figure 10 contains an example of an exercise from *Common Sense Mathematics* (Bolker and Mast, 2016) that falls under the core competency *assumptions*. The question in Figure 20 requires the respondent make assumptions about grades for Alice and Bob, and it requires the respondent evaluate these assumptions to make sure they actually fit the given scenario.

Exercise 8.5.15. [S][Section 8.4][Goal 8.1][Goal 8.3] Watch TV! Live Longer!

The data in the spreadsheet TVData.xlsx show the life expectancy in years for several countries, along with the number of people per television set in those countries. (The idea (and the data) for this problem come from the article www.amstat.org/publications/jse/ v2n2/datasets.rossman.html.)

(f) What else could be going on here? Why might life expectancy be strongly correlated with a low ratio of people per tv set?

Figure 19. Exercise 8.5.15 (f) from *Common Sense Mathematics* (Bolker and Mast, 2016, p. 193)

Exercise 5.7.8. [S][A][W][Section 5.3][Goal 5.1][Goal 5.3] Who wins?

Alice and Bob are both students at ESU. In September they start a friendly competition. In June they compare transcripts. Alice had a higher GPA for both the fall and spring semesters. Bob had a higher GPA for the full year.

(a) Explain how this can happen, by imagining their transcripts – number of credits and GPA for each, for the two semesters and for the full year, as in this table:

	fall credits	fall GPA	spring credits	spring GPA	year GPA
Alice					
Bob					

Figure 20. Exercise 5.6.8 (a) from *Common Sense Mathematics* (Bolker and Mast, 2016, p. 108)

Communication, defined by Boersma et al. (2011) as the "(a)bility to explain thoughts and processes in terms of what evidence is used, how it is organized, presented, and contextualized" is often indicated through the key-words, *explain* or *why*. Further, the explanation should involve justification with quantitative reasoning. For example, the question "(w)hich most accurately and efficiently captures the important features of this data set? Why?" from the module 3 homework in *Quantitative Reasoning* (Fahy and Nabours, 2016, p. 41) requires the competency *communication* since the why must be justified with appropriate quantitative reasoning.

In Boersma et al. (2011) the QRCW Casebook (Cusom edition) was similarly coded. The results of their analyses can be seen in Table 3. Tables 4 and 5 provide the same summary for chapters 5, 8 and 12 exercises from *Common Sense Mathematics* (Bolker and Mast, 2016) and modules 2, 3, 5, 6 and 8 from *Quantitative Reasoning* (Fahy and Nabours, 2016) respectively. One hundred and sixty-six questions from *Common Sense Mathematics* (Bolker and Mast, 2016) and 190 questions from *Quantitative Reasoning* (Fahy and Nabours, 2016) were coded in this analysis.

Table 3. Frequency and Prevalence of Core Competencies in the QRCW Casebook (Boersma et al., 2011, p. 9)

Frequency and Prevalence of Core Competencies in the QRCW Casebook				
Competency	**Number of questions**	**Percent of questions**	**Number of case studies (*N*=24) which have at least one question mapped to competency**	**Number of case studies which have 50% or more questions mapped to competency**
Interpretation	152	65	24	18
Representation	68	29	19	6
Calculation	101	43	21	9
Analysis/Synthesis	79	34	20	7
Assumptions	14	6	10	0
Communication	53	23	18	4

Table 4. Frequency and Prevalence of Core Competencies in Chapters 5, 8 and 12 Exercises from *Common Sense Mathematics* (Bolker and Mast, 2016)

Frequency and Prevalence of Core Competencies in the Chapters 5, 8 and 12 Exercises from *Common Sense Mathematics* (Bolker and Mast, 2016)		
Competency	**Number of questions**	**Percent of questions**
Interpretation	51	31
Representation	39	23
Calculation	79	48
Analysis/Synthesis	35	21
Assumptions	5	3
Communication	18	11

Table 5. Frequency and Prevalence of Core Competencies in Modules 2, 3, 5, 6 and 8 Exercises, Activities and Previous Module "Looking Forward" Exercises in *Quantitative Reasoning* (Fahy and Nabours, 2016)

Frequency and Prevalence of Core Competencies in Modules 2, 3, 5, 6 and 8 Exercises, Activities and Previous Module "Looking Forward" Exercises in *Quantitative Reasoning* (Fahy and Nabours, 2016)		
Competency	**Number of questions**	**Percent of questions**
Interpretation	29	15
Representation	32	17
Calculation	105	55
Analysis/Synthesis	34	18
Assumptions	2	1
Communication	12	6

From the above tables it is clear all core competencies other than *calculation* are required most often in the QRCW Casebook questions and more often in the sample questions from *Common Sense Mathematics* (Bolker and Mast, 2016) than in the sample questions from *Quantitative Reasoning* (Fahy and Nabours, 2016). The reverse order is indicated for the competency *calculation.* Further, the competencies *assumptions* and *communication* are relatively underrepresented in all 3 analyses. This analysis reflects the philosophies behind each of the course texts. *Quantitative Reasoning* (Fahy and Nabours, 2016) is a text designed around a strict mathematical content progression with real-world context added in where it fits the mathematical content of the section. It is not surprising that *Quantitative Reasoning* (Fahy and Nabours, 2016) exercises show the largest representation of the core competency *calculation* compared to the other two texts. *Common Sense Mathematics* (Bolker and Mast, 2016), while also organized around a mathematical content progression, is marked by a great effort on the part Bolker and Mast to ground every exercise and explanation in an authentic, real-world context. Unsurprisingly we see a greater variety in the competencies required by its exercises when compared to *Quantitative Reasoning* (Fahy and Nabours, 2016). The QRCW Casebook represents an added level of quantitative abstraction compared to *Common Sense Mathematics* (Bolker and Mast, 2016) in that case studies are designed around mostly full, authentic media articles. This is especially evident in its much larger share of questions that are coded for the competency *interpretation*. What this analysis does not show is the added sophistication in applying *interpretation* to QRCW Casebook problems; it is generally more difficult to apply *interpretation* in the QRCW Casebook than in the other two texts because respondents often have an entire article, as opposed to a blurb or word problem, from which they need to glean quantitative data.

INTERVIEW ANALYSIS – FIRST PORTION

The PI conducted 15 student interviews with 14 different students (one student was interviewed twice). Nine students were enrolled in MAT 114 at the time of their interview with one to two weeks left in the semester, five students had completed MATH 1313 at the time of their interview and one student was enrolled in MATH 2183 at the time of his interview with less than one week left in the semester. Of the five students who had completed MATH 1313, three of them were enrolled in a summer section of MATH 2183; two of the three interviews were held within the first five days of MATH 2183, and one of the three was held on the 11th day of classes. Of these three students, one was later interviewed again in the last week of the summer session in which he was enrolled in MATH 2183.

The first portion of the interviews has interview subjects take the prompt-less instrument (PLHOM) from Boersma and Klyve (2013). Subjects who completed the online version of the PLHOM prior to the interview were given the same article they had during the online assessment with one exception where the student mistakenly received a different article from her online version of the PLHOM. Eleven subjects who completed the online version of the PLHOM prior to interviewing received the same article as they had seen during their PLHOM online assessment. Nine of these 11 students interviewed roughly three months after taking the online version of the PLHOM, and 2 of these 11 interviewed within a week of taking the online version of the PLHOM; two subjects never completed an online version of the PLHOM. This first portion of the interviews seeks to inform what habits of mind QR students demonstrate and to investigate similarities and differences in habits of mind among and between students. Part of this section intends to inform both how and where students apply a HoM as well as identify obstacles students face in doing so. Specifically, the goals of the first portion of these interviews were to:

a) Document the student's age, year in school, major, previous coursework in QR, possible future coursework in QR and previous quantitatively relevant coursework.
b) Observe student interaction with the prompt-less instrument (PLHOM) from Boersma and Klyve (2013).
c) Establish a base-line score on the PLHOM for the student on the day of his or her interview.
d) Attempt to elicit higher scoring responses on the PLHOM through prompting.
e) Investigate difficulties the student had in interpreting quantitative information from the article.
f) Investigate the student's impressions concerning reliability of the data/sources.

Due to time constraints, not all of these goals were met in every interview. Goals (a), (b), (c), and (d) were met during every interview. Since individual students had difficulty with different parts of the articles, and these difficulties were only investigated whenever they became apparent from student responses to the PLHOM, they were not investigated in a consistent manner for all interview subjects. Usually the students' impressions concerning reliability of the data and sources from the articles were apparent from responses to the PLHOM and discussion during the interview; however, goal (f) was not met in all interviews.

Interviewed students ranged in age from 18 to 38 years old; eight interviewees were 18 years old at the time of the interview, one was 21, one was 22, one was 26, one was 28, one was 36 and one was 38. There were eight freshmen, one sophomore, no juniors, four seniors, and one fifth-year senior. The students came from a variety of majors with a few students reporting a double major: English literature (1), political science (2), criminology and criminal justice (2), hotel and restaurant management (2), communication studies (2), African-American studies (1), sports and recreation management (1), photography (1), environmental sustainability (1), graphic design (1), nursing (1) and studio art (1). Only one student indicated having taken previous course in QR, however, the course was taken in high school. One student indicated having taken international bachelorette (IB) mathematics studies in high school, two took advanced placement (AP) statistics and one of these two also took IB calculus. Several students mentioned having taken high school chemistry, physics, biology or some combination of these. One student indicated having taken (and failed) college algebra. Sports and recreation business courses, economics, geology, human anthropology, communication research methods and information systems management are other courses students indicated taking that may be relevant to QR. Only one student indicated possibly taking another course in QR, as a part of a graduate school curriculum in political science.

The student who was interviewed twice had a unique article with his PLHOM during his second interview so that the interview would not influence his responses whenever he responded to the online version of the PLHOM. His second interview data are not included in the following discussion because of this abnormality. Students spent an average of 10 and 1/2 minutes responding to the PLHOM during interviews. Notably, the student with the highest score on the instrument, who will be referred to as Art, spent the longest amount of time filling out the PLHOM; Art spent 18 minutes and 45 seconds reading and filling out his responses and received a total score of 9 out of 12. The student who spent the least amount of time filling out the PLHOM, who will be referred to as Cory, spent only four minutes and 20 seconds reading and filling out the PLHOM; Cory's response was 1 of 2 responses during interviews that received a total score of 1 out of 12, the lowest score observed during interviews.

There are several possible explanations for this, and it is probably not the case that simply spending more time filling out the PLHOM will result in a higher score. However, the student who took the longest and received the highest score on the instrument took his time and often looked back through the article to check what he had read. Art underlined bits of information in the article, would begin a response, refer back to the article and, in one instance, revise what he had written. Interestingly, Art interviewed only two days after filling out the online version of the PLHOM with the exact same article. It can be argued that this gave Art an advantage, and it likely did; his score increased from an 8 out of 12 on the online version to a 9 out of 12 during the interview. However, it is surprising that Art still took the time to be thorough in his reading and critical reasoning when responding to the PLHOM during the interview.

Cory, on the other hand, breezed through the article in a little over a minute, but he did seem to apply serious thought to the questions, and indicated that *[he] takes everything [he does] seriously*. He went on to state that *I'm not kidding anymore. I don't play games. I'm not trying to waste my time, or anybody else's time. So… I figured if I was gonna participate in this, I need to be, you know, honest and forthcoming*. From this we can presume that Cory's low PLHOM score was not due to a lack of effort or seriousness in responding to the PLHOM. Cory, a criminology and criminal justice major, was very interested in the article and even noted that he *like(d) playing detective*. The cases of Art and Cory make for a good comparison. They were both the oldest subjects interviewed; Art was 38 and Cory 36 at the time of their interviews. Art had just completed MATH 1313, and Cory was in his last week of regular classes for MAT 114; both of these classes are introductory level QR courses.

If we take everything Art said and wrote during the interview, he produced level 3 responses to every question from the PLHOM, and was able to do this with fairly minimal prompting. For example, consider Art's initial response to question 3, "(w)hat facts did the author use to support the main points": *Lead abatement started in the mid 1970s and continued through the 1980s. This abatement would have affected young people representing high crime groups in the 1990s and 2000s. Studies have shown exposure to lead can increase violent behavior*. This level 2 response is missing two things: quantitative information from the article that supports a correlation between lead abatement and violent crime, and quantitative information that disputes the correlation.

The PI asked Art to provide one or two more facts, and Art responded by writing: *(a)verage preschool blood lead peaked around 1970 and were lowest in late 1980s. NCVS and homicide stats don't support the argument while UCR stats do support the argument*. At this point, the PI moved on with the interview. Together, these two responses represent a borderline level 3 response. The statement that the *NCVS and homicide stats don't support the argument while UCR stats do* is certainly true, however the reason why this is true is also provided in the article and mainly supported by the second graph. While this information is absent in these responses, consider Art's initial response to question 5, "(d)oes the graph help interpret the numerical information found in the text": *(t)he second graph is easier to interpret as it shows a correlation between lead exposure and UCR crime stats and a deviation between lead exposure and the NCVS and Homicide Stats*. If we include this response in Art's responses to question 3, then these combined responses represent a Level 3 response. Further, level 3 responses to the PLHOM typically require a fairly extensive exposition and, in the experience of the PI, are quite rare. Art had the necessary tools and habits to show the highest PLHOM-measurable level of HoM, he simply needed a little prodding to produce them.

If we take everything Cory said and wrote during the interview, he produced responses that result in a total score of 7 out of 12 on the PLHOM, an increase from his initial score by 6 points, the maximum observed increase through prompting in PLHOM score. It is curious that Art only needed a small amount of prompting in order to reach his full potential on the PLHOM, whereas, the PI practically had to drag the higher-level responses out of Cory. Consider Cory's initial, level 0 response to question 4, "(w)ere there any particular strengths or weaknesses in how these facts were reported": (*n)ot really, at least none that really held any weight. However, the argument within the article is quite interesting.* Here, it is not immediately clear if Cory is speaking about strengths, weaknesses, or both when he says *none that really held any weight*. The PI responds to this by asking Cory to identify a specific strength or weakness in how the information in the article is presented. Cory responds by writing,

> *I agree with the Journal of Quantitative Criminology Report, in the sense that even with the reduction of lead, crime rates have gone down but have not dissipated and much more serious crimes are still taking place at an alarming rate.*

In this response, it seems as though Cory is pointing out a strength, but the particular strength is still not clear. The PI responds by saying,

> *Okay, so… um, let me see how I can phrase this… try and, try and just give it one more shot, but focus more on, like, how [the facts] were reported in the article. Like, in, in the absence of um… the actual facts, I guess. But as far as, um… …what am I trying to say? Not just the facts, but how [the authors'] evidence is presented, basically. And so, if you can say one thing [where] you think the facts were presented clearly and the evidence was presented clearly, or you think that maybe the facts or the evidence um... could have been presented better? Or maybe, [the facts] could have been presented in a misleading manner, or something like that?*

Cory responds to this by writing, *(t)he authors present good data to argue their respective side but fail to bring in supporting evidence*. Here, we finally see Cory point out a valid weakness in how the facts of the article are presented (level 1 response), however, he presents the conflicting claims that *(t)he authors present good data* but *fail to bring in supporting evidence.* This leads to the following exchange:

> PI: Okay, so you're saying that you think that the supporting data is good… … so what kind of supporting evidence would you want to see brought in?
>
> [Silence]
>
> Cory: *They, they did a good job of bringing, you know, uh... the one autho brought in FBI data, which is good and helpful for their argument. However… I feel that they could have brought in more... what am I trying to say? More…*
>
> [Silence]
>
> Cory: *More statistical evidence, I guess?*
>
> PI: So… um… anything from 114? The kind of statistics that you would… deal with in there?
>
> Cory: *Yeah, I guess. Um…*
>
> PI: Can you name, like, one specifically?
>
> Cory: *That's what I was trying to think of... hmm, this is gonna sound ridiculous, but [the] first thing that comes to mind is, is like a stem-and-leaf plot.*
>
> PI: Okay
>
> Cory: *Put together statistics showing, you know, these are, these are alarming rates of, of crimes… and at the same time, during that period the lead exposure was significant as well. Um… boxplots would, would work, um… I don't recall if we use this or not, but for some reason, I'm, I'm thinking a cat-and-whisker plot?*

Here we see Cory grasping at concepts from MAT 114, but they do not seem particularly relevant to the prompt. Perhaps the PI was too leading when asking about MAT 114, so we cannot read too far into this exchange. What is interesting is that Cory has the instinct to know that there are data relevant to the authors' argument that are absent, but he is unable to nail down the kind of information that would satisfy him.

Art's initial response to question 4 from the PLHOM starkly contrasts with the above exchange during Cory's interview:

> *The statement that studies suggest a correlation between lead exposure and violent behavior should be backed up with evidence such as the statistics, names of the studies, etc…. The statement that other developed countries experienced crime drops should have been backed by specific evidence such as name of country, crime rates before and after, lead exposure before and after*.

Art, like Cory, felt that relevant supporting evidence for the authors' points was absent, and, unlike Cory, elaborates on exactly the type of evidence that would better support the argument. He adeptly and without any prompting makes note of this after careful thought and reflection on the information provided in the article. One possibility for Art's ability to fluently and accurately explain weaknesses in how the information in the article is presented is that Art's experiences in MATH 1313, a course that incorporates exercises that specifically ask students to critique claims from media articles, primed him to produce specific, detailed responses on the PLHOM. Cory, who genuinely wanted to make a relevant and correct analysis of the article, seemed confused by how he should do so. Without more data from MATH 1313 students, this possibility is difficult to follow up on. Another explanation has to do with Art and Cory's respective majors and standing in school. At the time of his interview Art was a senior communications major and mentioned he had completed communication research methods. Cory, on the other hand, was a sophomore criminology and criminal justice major, which explains his interest in the article, but does not inform us about why he was unable to specifically and accurately explain what information was missing from the article.

Another possibility exists that could help explain why Cory was so vague in his initial response to question 4: (*n)ot really, at least none that really held any weight. However, the argument within the article is quite interesting.* Cory may have had difficulty identifying a strength or weakness in the argument simply because he didn't properly understand the argument. Specifically, we can see this in his follow up response to prompting: *(t)he authors present good data to argue their respective side but fail to bring in supporting evidence.* The key phrase being *argue their respective side*; Cory seemed to believe that the argument the authors wanted to make was that reduced lead exposure fully or mostly explains a drop in crime rates, and he was not alone in this misconception. The authors' true point is to discuss the theory that reduced lead exposure in pre-school later caused a drop in violent crime in individuals at peak criminal age (23 years later). The authors give both sides of the argument and finish by reporting that there was probably some effect, but then cite a Brennan Center analysis that was unable to quantify the impact. So, at most, the authors are in weak support of the theory, but their goal is not to convince the reader that pre-school lead exposure is the primary factor in violent crime reduction.

Similarly, the incarceration article presents the theory that increased incarceration rates contributed to a drop in violent crime, but the authors' point is to give both sides of the argument; they argue both for incarceration's effect in reducing violent crime and for the limitations of this effect. One problem with investigating misconceptions about the authors' point is that question 2 from the PLHOM, "(w)hat was the main point(s) of the article", was absent during interviews. Because this question is not formally scored according to the PLHOM rubric, it was removed to save time in the interviews. However, based on responses to questions 3, 4 and 5, and dialogue from the interview transcripts, it was fairly easy to determine interview subjects' impression of the authors' points. For example, the response, *(h)e contrasts his main theory that mass incarceration lead to lower crime rates, but then gives evidence that mass incarceration didn't help lower crime rates* was a reported weakness for the incarceration article.

It seems clear that this response indicates confusion about the authors' point. In fact, 7 of the 14 interview subjects seemed to have trouble with what the main point of their article was. This indicates a problem with reading comprehension or perhaps a tendency to view and discuss complex issues from a binary perspective. Reading comprehension overlaps with the core competency *interpretation*, and binary descriptions could theoretically stem from a problem with *communication* in writing; nuanced perspectives are more difficult to explain than objective facts.

The online version of the PLHOM included question 2. The PI revisited the online responses in order to evaluate how well the QR students who responded to the PLHOM understood the articles. It was not possible to evaluate exactly how well students understood the article they read since many students gave vague responses such as *lead exposure, crime*. The PI only counted cases where the student made it clear they had a misconception about the points of the article. Sometimes students are explicit in their misconception and one reported, for example, that they *(r)eally don't know* the main points of the article. Others indicated their misconception indirectly. For example, *(t)he main points were about mass incarceration and comparing them to the years overtime* clearly indicates this student misunderstood the points of the article. While incarceration over time was discussed in the incarceration article, the main points of the article revolve around the theory that mass incarceration reduced rates of violent crime. Also, the PI considered reported weaknesses such as *lead exposure may not explain the entire crime drop* to indicate a misunderstanding of the article's main points. Roughly 38% of the 98 responses to the online PLHOM indicated a misunderstanding of the main points of the article. If we remove the 13 cases where responses were not clear enough to determine if the student had a misconception about the main points of the article, this jumps to roughly 45% of 85 responses indicating a misunderstanding of the main points of the article. This is a strong indication that reading comprehension and/or written communication is a significant obstacle for QR students. If students are unable to determine a written point, then they would be hard pressed to critique, reflect upon, and reason quantitatively about the points and supporting evidence; if improvements in the latter skillset is a desired outcome for QR students, it follows that practice with critiquing, reflecting upon, and reasoning quantitatively about the points and supporting evidence of articles should be directly practiced by QR students as a part of their QR course curriculum. This type of activity is mostly absent in MAT 114, modestly infused into MATH 1313, and prolific in MATH 2183.

Students who complete MATH 2183 should, in theory, increase their ability to critique a media article through a quantitative lens. Further, we hope this translates into an increased tendency to critique media articles through a quantitative lens. There was a single interview with a student, referred to as Adam, who had nearly completed a summer section of MATH 2183, and this student was also interviewed 3 weeks prior (2 weeks into the MATH 2183 summer session). The article Adam received during his final interview can be viewed in Appendix C. Any expectation of large, significant changes in HoM over this time frame is likely delusional; however, the interviews show subtle, positive shifts in Adam's HoM. Adam received a 2 out of 12 on two separate online versions of the PLHOM, and in both interviews Adam's initial responses to the PLHOM also scored a 2 out of 12. In all four assessments Adam received his 2 points through 1 point scores on questions 3 and 4.

The clearest change in Adam's thought processes during the interviews is observed in his responses to follow-up prompting for question 5. Adam was asked how well the information in the article matches the information in the graph(s) in both interviews. During the first interview

Adam answered with *honestly, I think it matches pretty well* and elaborated that *it has to, right, because it's there. I think if I actually look at it. If I take the time to go through it, I'm sure it matches.* This response indicates Adam did not think to check if the information in the graphs was consistent with the information in the article even when asked directly, and he further took the graphs' existence as proof to their credibility. This response was not followed-up with any further prompting about the graphs used in the article. When asked the same thing during the second interview, Adam responded:

> Adam: *2000… 2000…. Is this the full graph? Cause this says 2000 Medicaid has increased from 34 to 54 million people. When it starts at 2009.*
>
> PI: That's what was in the article, I didn't make it up.
>
> Adam: *It doesn't, so I guess it really doesn't match. I mean I guess we could see if we look at 2011, we could see if that matches what they say in there. 54 million but it really… (laughs). I don't think this matches!*

Adam's reasoning here is incorrect since the graph is labelled to show the number of Americans receiving some form of federal welfare other than Medicaid or Social Security. However, the important deviation from Adam's first interview is that Adam thought to check the information in the article with the information in the graph at all, and even went as far as believing he found a discrepancy. Searching articles for these kinds of discrepancies is practiced directly in MATH 2183 and it seems likely that, at the very least, Adam's predisposition to blind faith in the accuracy of reported information was productively challenged during the three weeks between these two interviews.

The idea behind the PLHOM is to measure students' HoM toward QR. The instrument is not truly prompt-less; however, the prompts are quantitatively neutral. Additional prompting avoided using quantitative qualifiers, but there were a few times where quantitative qualifiers were used. Also, students were likely predisposed to focus on quantitative information during the interviews since they were recruited from a QR course. Any time a student asked if they should focus on quantitative information in their responses they were told that it did not matter. In one case a student was asked to focus on quantitative information in follow-up prompting to question 1. In three cases students were asked not to focus so heavily on prior knowledge from previous courses; two of these cases involved students bringing in prior knowledge about criminology and the other involved the student's focus on the format of citations. In many cases students were asked to explain a graph. The follow-up prompting was generally successful in eliciting higher scores on the PLHOM.

On average, student scores rose by 3.33 points ($N = 15$) as a result of follow-up prompting. Interestingly, follow up prompting was not as effective in the group of students who had completed MATH 1313 as it was in the group of students who had completed MAT 114. Adam's second interview is not included in the following analysis; including two samples from the same interview subject could skew the results, especially since the sample sizes are already so low. Follow-up prompting resulted in a PLHOM average increase of 2.4 points in the MATH 1313 cohort ($N = 5$) and 4.11 points in the MAT 114 cohort ($N = 9$). Further, differences in mean PLHOM score increases due to follow-up prompting between these two cohorts was statistically significant at the 0.10 level ($p = 0.09$). However, the sample sizes are low enough to cause serious concern about the validity of this result. In the case that follow-up prompting really is less effective in the MATH 1313 cohort, one theory to explain this is that MATH 1313 results in increases in HoM up to limitations in mathematical or intellectual ability. Follow-up prompting generally failed at the point where a misunderstanding on the part of the student prevented him or her from increasing their score on the PLHOM further. Students in the MAT 114 cohort tended to have more room in their ability to critique the articles through a quantitative lens than their initial score would suggest, but they were less likely to apply those skills at first.

Follow-up prompting varied in effectiveness. There were two types of prompting that proved most effective, but this was question specific. On question 3 of the PLHOM, “What facts did the author use to support the main point(s)”, asking students to be more specific in their responses resulted in higher level responses (increase in 1 point on the question) five out of the six times it was asked. Asking students to find another fact was not as successful; this prompt resulted in an increase by 1 point on question 3 two of the four times it was used. Asking a student to elaborate or explain something he or she had written or said on question 3 was even less successful; this prompt resulted in an increase by 1 point on question 3 one of the five times it was used.

Typically, students were too general or vague in their responses to question 3. Consider, for example, this level 1 response to question 3: *(t)hat incarcerating violent people has lowered crime rates and that increasing incarceration has reduced crime rates all over*. Here, the student alludes to quantitative information in the article, but the response is too vague and the correctness is unable to be determined. When asked to be more specific, the student says *(i)n Steven Levitt’s paper written in 2004, he shared that 58% of the drop in violent crimes during the 1990’s was due to incarceration*. This level 2 (when viewed alongside the initial response) response is specific, it involves quantitative information and it is relevant to the authors’ point. In the cases where students were asked to explain or elaborate on something said in response to question 3, students did not generally add any more supporting evidence from the article. Though, resulting explanations sometimes uncovered student misconceptions. For example, one student responded to this follow-up prompt with *(t)he graph shows that MV theft and burglary* *are some examples of crime that have decreased, as well as overall crime but robbery and homicide are crimes that have oppositely increased.* The graph the student refers to represents percent decreases in crime by type; however, the student's explanation shows that they do not understand this. One take-away from looking at the results of follow-up prompting to question 3 is to consider adding "specific" as a qualifier to "facts" in the question. Changing "(w)hat facts did the author use to support the main point(s)?" to "(w)hat specific facts did the author use to support the main point(s)?" maintains the prompt's quantitatively neutral status, and could result in better student responses.

The most convincing pattern identified in student responses to question 4, "(w)ere there any particular strengths or weaknesses in how these facts were reported?", is that students showed difficulty in identifying strengths or weaknesses whenever they misunderstood the point of the authors. Further, asking students to elaborate or explain something they wrote or said in response to question 4 resulted in an increase by 1 point only in students who seemed to understand the point of the article. Of the six students who were asked to elaborate or explain something they wrote or said in response to question 4, the three students who seemed to understand the point of the article increased their score by 1 point, whereas the three students who did not seem to understand the point of the article saw no improvement in score as a result of this prompt.

One student, referred to as Lora, who seemed to grasp the authors' points said, *I felt the main points were obviously the theory that the study is trying to present and the main points that go into supporting that theory are both counter arguments and arguments that support what the theory is*. Lora's initial (level 1) response to question 4, *(t)he evidence that is presented needs to be more in depth and have citations. The way the info is organized makes it easy to understand*

and follow, is not wrong, but it is not specific enough to determine its correctness. When asked to elaborate on this response, Lora responded, *(s)ome of the info seems like speculation so having actual numbers with citations in the article that align with the graphs would make it more reliable.* This level 2 response points out the more specific weakness that the information in the article and the information in the graphs do not support each other in a verifiable way. Had Lora listed specific figures or statistics the author could use to rectify this weakness, her response would have reached the highest level of 3.

Two of the three students who saw no increase in score on question 4 from an explain/elaborate style follow-up prompt simply struggled to frame their reasoning in a clear, coherent way. It was not clear that this stemmed directly from a misunderstanding of the article's point; however, it seems likely that their inability to properly determine and/or explain the point of the article could be related to their inability to explain their chosen strength or weakness in how the facts of the article were reported. One of the three, referred to as Chayla, made it clear in her explanation that her low score stems directly from her misunderstanding of the article's point. In Chayla's original response – a reported weakness – to question 4, she makes her misunderstanding clear:

> *He contrasts his main theory that mass incarceration lead to lower crime rates, but then gives evidence that mass incarceration didn't help lower crime rates. He gives the fact that people who were locked up and causing the high crime rates were already put away thus lower rates, and putting more people away that didn't contribute to the high rates.*

When asked to elaborate on this, Chayla doubled down on her claim that the incarceration theory is false. What follows is a sample from Chayla's explanation:

> *Okay, so people who were put into like mass incarceration like during that period especially like during the 80s, they concluded that crime rates were like dropping because of all these people put away, but like what I was saying in the part that you underlined is that um, especially according to this graph is that crime rates had already been like, like going down because people had already been put away, so whenever*

> *you're trying to get tough on crime and you're putting away like non-violent offenders they're like suggesting that crime rate is going down because we are putting away all like these millions and millions of people, but the correlation is not there because the violent crime offenders had already been put away.*

Here, Chayla uses information from the authors' own exposition about diminishing returns as evidence that the incarceration theory is false. Further, she misrepresents what the graph she refers to actually shows. Chayla's responses make an interesting case. Chayla, an African-American studies and political science major, interviewed after completing MATH 1313. She also indicated she would not go on to take MATH 2183. Chayla's interest in the article's subject resulted in detailed, assertive responses. This places her responses close to higher level responses, but her misinterpretation of the authors' points and the graph in the article were her limiting factors on the PLHOM. Further, it seems plausible that her interest in the article's subject – including a bias against mass incarceration – simultaneously helped and hindered her scores on the PLHOM; although, it is not clear whether Chayla lacked the reading comprehension required to understand the authors' points or if her biases clouded her better judgement. It would be interesting to investigate the effects of MATH 2183 on students like Chayla. In any case, asking Chayla to elaborate on her response to question 4 did not harm her score and resulted in a more detailed response.

Asking for an explanation is a quantitatively neutral prompt; the above suggests that adding "explain" as a qualifier to "particular strengths or weaknesses" could improve scores on question 4 of the PLHOM in more quantitatively literate individuals without affecting scores for those limited by their ability to reason quantitatively. The prompt "(e)xplain any particular strengths or weaknesses in how these facts were reported?" could be more suitable for the PLHOM than "(w)here there any particular strengths or weaknesses in how these facts were reported?".

Recall that question 5, “Do the graphs from the images help interpret the numerical information found in the text? Explain your thoughts” is intended by Boersma and Klyve (2013) to be scored in two separate ways. The first rubric, referred to as 5a, assesses the extent to which figures or information written in the article are checked against information in the article’s graph(s). The second scoring, referred to as 5b, assesses reported strengths or weaknesses in how the graph(s) support(s) the authors’ point(s). It was difficult to disentangle follow-up prompting as it applied to these two scoring rubrics separately. In all but one interview the PI applied follow-up prompting as it is written in the interview protocol in response to scoring from the rubric for 5a and ignored the follow-up prompting as it is written in the interview protocol in response to scoring from the rubric for 5b. Typically, any increase in score observed with respect to the rubric for 5a also resulted in an increase in score with respect to the rubric for 5b.

Out of all 15 interviews, nine initial responses to question 5 scored at the 0 level for both rubrics, 5a and 5b. Responses that were too general characterized seven of these nine responses, and the remaining two responses were just admissions that the students found graphs to be confusing. *Yes, the graphs help to understand the information because you can visually see how incarceration helps/doesn’t help reduce crime* is an example of a response that is too general to have any points awarded via the rubrics 5a and 5b. *No, the graphs are confusing. The text better help me understand the number* is an example that indicates the student found graphs confusing. In all nine cases students were asked how well the information written in the article matched the information presented in the graphs. Three responses to this follow-up prompt received no increase in score for 5a or 5b, five responses received one additional point on 5a (with all but one receiving an additional point on 5b as well) and one response received two additional points on

5a and 5b. In all five cases where responses received one additional point on 5a, the response was deemed to have faulty quantitative reasoning.

At the heart of the generally low scores on question 5 seems to be a difficulty in interpreting graphs. Further, since a significant amount of the quantitative information in each article comes exclusively from the graphs, misunderstandings about the graphs caused lower scores on questions 3 and 4 as well. Consider the first graph (Figure 21) in the incarceration article:

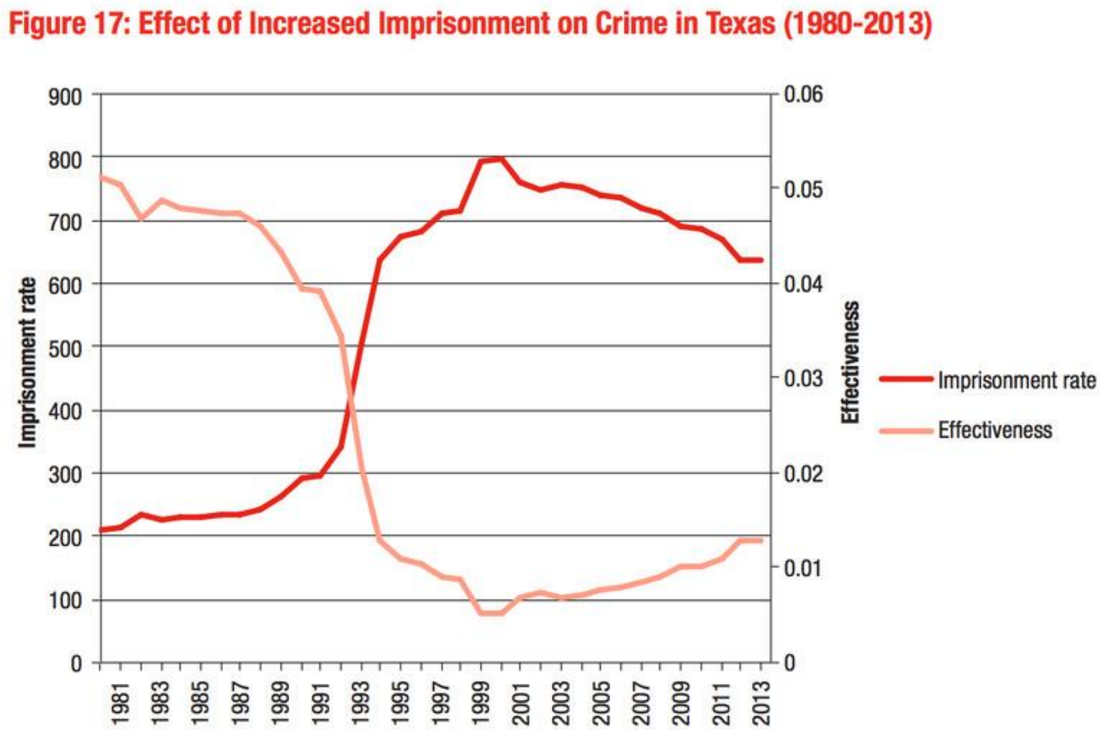


Figure 21. Incarceration article; first graph (Lopez and Lynn, 2016)

This graph appeared on the PLHOM in seven student interviews. Only one student gave a description of this graph that is in line with what the graph attempts to show: *(t)he first graph support[s] "the case against" because it shows that as imprisonment rates went up, it was actually less effective*. In the other six interviews students communicated incorrect interpretations of this graph, below are several examples of this:

- *The first graph shows the effect of increased imprisonment on crime, and shows that at some point, as imprisonment increases, crime decreased, then both seem to level out at end/change directions which supports theory.*
- *In Figure 17 it shares that no matter the increasing rate of incarceration gets, it doesn't affect the effectiveness.*
- *This one was just simply the years and then the rates and then like two lines and how they obviously went up and down, like totally different. And so for me, he basically like says in words like that graph, but having the graph there it's just like another like confirmation.*
- *in a chart that was used in Texas to examine the imprisonment rate compared to effectiveness [it] showed that specifically in the 90s that the lower the imprisonment rate had very little effectiveness.*

While three students mentioned the graph only includes information from Texas and indicated there could be a problem generalizing the reported results to the rest of the nation, there were no instances where a student complained that the measure of *effectiveness* is never explored or explained in the article; the units on *effectiveness* were never directly called into question either. The students bumped into these issues in some of their explanations. One student, while trying to determine how well the information written in the article matches the information in the graphs, explained:

> *The only problem is I don't really know what these numbers mean, so it could be the opposite way, but here, when there's more people imprisoned um the, the effectiveness is going lower and either that means there's less crimes or more again. The numbers here don't… [they] kind of confuse me.*

The previous statement shows this student is close to realizing a key weakness in the article's use of this graph, however, she still seems to feel that it is her own misunderstanding and not the fault of the authors.

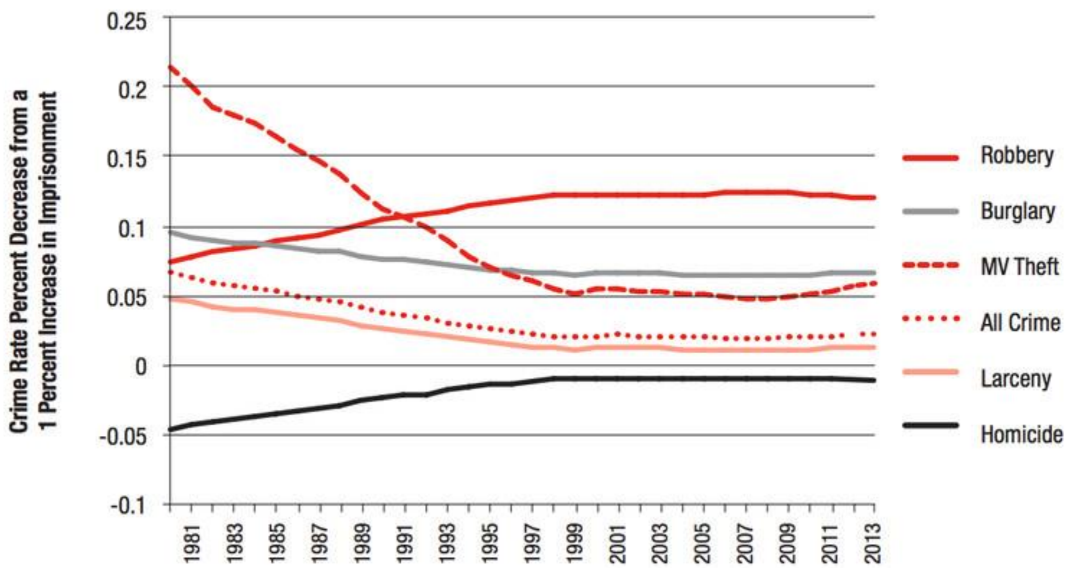


Figure 22. Incarceration article; second graph (Lopez and Lynn, 2016)

The second graph (Figure 22) from the incarceration article was even more difficult for students to interpret. Initially, no student seemed able to correctly interpret this graph, however, one student pointed out a valid weakness in its use in the article:

> *I think the use of a graph with specific crimes is a weakness since the information needing support does not refer to specific types of crime… … I think if the specific types of crime were touched on in the preceding paragraph, it would offer more impactful support to theory.*

Students generally showed hesitation in discussing this graph at all, as seen in several initial responses to question 5:

- *The first graph helps interpret that the higher the incarceration rate gets, the less it matters if you increase it more, but the second graph didn't help.*
- *I didn't even look at the second graph because it didn't seem relevant.*

- *Graphs tend to confuse me and don't help me as much as words do.*

When pressed to describe the information presented in this graph, four students showed a complete inability to correctly interpret the graph and one showed a correct interpretation (after a lengthy discussion). Two students were not pressed to describe the information in this graph since they admitted to not understanding it in their initial responses. Consider the following exchange with a student who will be referred to as Zoe:

> PI: Okay. And so then [the] last one about this, um, could, last thing, and you can just do this verbally, I'd like you to sort of explain what this, um, is illustrating.
>
> Zoe: *Okay. Um, don't really get the y-axis that much. Um, the crime rate percent decrease from one percent increase in imprisonment.*
>
> PI: And feel free to think out loud.
>
> Zoe: *I guess what I think it's showing is, um, five different specific types of crime along with all crime and it's showing the, uh, how crime rate is decreasing, um, as imprisonment increases?*
>
> PI: Okay.
>
> Zoe: *But it doesn't really show us, I mean it says one percent increase in imprisonment but it doesn't really show us, I don't know. I think that's, I don't, I think that's just what it's trying to say, that crime, it's trying to show the crime decrease? Crime rate decrease for each specific type of crime?*
>
> PI: Okay. And so then, if you were making the argument that increases in imprisonment, uh, lead to decreased crime-
>
> Zoe: *Mhm.*
>
> PI: Then what would you expect to see in like the overall crime or something?
>
> Zoe: *You expect to see it decrease.*
>
> PI: So like the line though. Just like describe it and like what the graph would look like.
>
> Zoe: *Um.*
>
> PI: Like if-if imprisonment has like a really large effect
>
> Zoe: *On crime?*
>
> PI: on, yeah, on decreasing crime.
>
> Zoe: *Then as imprisonment would increase the crime would go down. It-the lines would go all down.*

PI: They'd all go down?

Zoe: *Yeah.*

It was very difficult for students to realize correlating increases in incarceration with a greater decrease in crime rates results in lines that trend upwards in this graph. In the following exchange a student who will be referred to as Lisa came closer to this realization than Zoe:

PI: Mhm. Okay, and so, let's see. Do you think, uh, could you explain to me like what's being measured on this axis?

Lisa: *Mm, the crime rate percent decrease. So how much the crime rate has decreased from one percent increase in imprisonment? So for every one percent how much it decreases in crime rate?*

PI: Okay, so, um, so then, would we want to see higher or lower levels here if we want cri-, if we want to say, like, crime is going down?

Lisa: *Um, I think lower*.

PI: Lower?

Lisa: *Yeah.*

PI: Okay, so then, um, what are you laughin' about?

Lisa: *I'm so confused with-with graphs, that's just why.*

PI: You're just confused about, what's-what's confusing about this graph?

Lisa: Um, I think just wording. Wording trips me up sometime, yeah.

… …

PI: Okay, but, so you did mention though, so when you read this to me you said that this is, um, percent decrease in crime.

Lisa: *For every one percent.*

PI: for, right. Okay, so then in 1981, let's look at the MV theft, so motor vehicle theft, that's, like, this dashed line right? So then in 1981, um, and then also these are-these are probably percentages right?

Lisa: *Yeah*.

PI: So, interpret that for me that in 1981, if we increased imprisonment by one percent, what happened to motor vehicle theft?

Lisa: *Decreased by point 2?*

PI: By point 2 percent. Okay, and so then what about in 2013?

Lisa: *Um.*

PI: Same thing.

Lisa: *It decreased by point five percent? Oh wait, down point five percent I think that's…*

PI: Yeah, it's not quite there, it's, yeah, it's,

Lisa: *Point five something.*

PI: Yeah something like that. Yeah, yeah, yeah, close enough. Um, so then, now let's revisit that last thing that we said. Would we want to see these lines going, like, if we're wanting to argue that increases in imprisonment like lead to larger decreases in crime, then would we want see these lines going down or going up?

Lisa: *Um, I think going down.*

PI: Going down still?

Lisa: *Yeah*.

… …

PI: I want you to do your-your best to explain to me why, uh, why you, you would want to see these lines trending downward.

Lisa: *Um*.

PI: And like what that would mean.

[Silence]

Lisa: *I guess the concept that seeing a graph, like, decrease and in this case for like imprisonment and crime rate, it would be, like, it would be a-a good thing? I don't know.*

PI: What are you wondering about right now?

Lisa: *I'm wondering if that like the higher or lower it is, if it's worse or if it's like if it's good because it says for every one percent of imprisonment so I don't know if like the lower it is, that means that the imprisonment is decreasing and maybe the crime rate is increasing.*

PI: So, I think the-

Lisa: *That's what I'm confused about.*

Here, we see that Lisa is very close to understanding the graph, however, her belief that decreases should be represented by downward trends in a graph is so strong it overrides her ability to reason properly about the graph. Lisa is not alone in this belief. In another interview the student and the PI had a nearly identical conversation; they discussed the measure along the *y*-

axis (percent decreases in crime due to a one percent increase in incarceration), the student recognized that larger decreases in crime are represented by higher values along the *y*-axis, yet the student still believed that the graphs should trend downward to represent greater decreases in crime rate due to increased incarceration. In the following exchange we see a student, referred to as Monica, who comes to understand the graph fairly well:

> PI: Okay, so let's see, so then maybe you want to explain to me um how well you think the other graphic is sort of lining up with what's in the text?
>
> [Silence]
>
> Monica: *Mmm, this graph is a little bit more confusing to me. Uh it says crime rate percent decrease from a 1% increase imprisonment, imprisonment so, so the graph is showing the [decrease] of crime rate when there's an increase of imprisonment.*
>
> … …
>
> PI: The graph is like, what you said, it's like per an increase in imprisonment, so if uh, if there were large drops in crime due to imprisonment what would you expect to see?
>
> Monica: *On the graph, not in like life?*
>
> PI: Yeah, on the graph.
>
> Monica: *Um, I guess I would see the, like for all crime, I would see it start somewhere and then rise.*
>
> … …
>
> PI: And so why would there be a rise?
>
> Monica: *Um, [inaudible]. Um, because the graph is showing the crime rate percent decrease when there's an increase in imprisonment so the graph is just kind of depicting that um…*
>
> PI: So then like higher values on the graph would be higher amounts of…
>
> Monica: *Yeah, I guess so.*
>
> PI: Of, of what it's measuring, right, which is, what's it measuring again?
>
> Monica: *Crime rate percent decrease.*
>
> PI: Right.
>
> Monica: *Um, [inaudible] I, it might like, it might start at a high number and then go down, like as if the crime rate is decreasing, and decreasing normally is represented going downwards.*
>
> PI: Right, but the graph's measuring – is measuring, like, its units are in decreases right?

Monica: *Yeah.*

PI: So if we wanted more decrease than on this graph, we would want to see it…

Monica: *Yeah.*

PI: So you got to pick one. Do you want to see it go up or down?

Monica: *Um, I guess up.*

… …

PI: So if you had to pick one of those where the, one of those crimes where um the imprisonment is showing like an impact on the decreasing crime what would you say?

Monica: *Um, robbery and homicide.*

PI: Because you see those doing what?

Monica: *Um, they're going more up.*

Monica came the closest out of the seven interview subjects given the incarceration article to communicating an understanding of this graph. Notably, she still stumbled on her interpretation of the graph, at one point reverting back to thinking decreases in crime rates should be represented with a downward trend in the graph, and she showed no indication that she understood what the negative values along the *y*-axis represent.

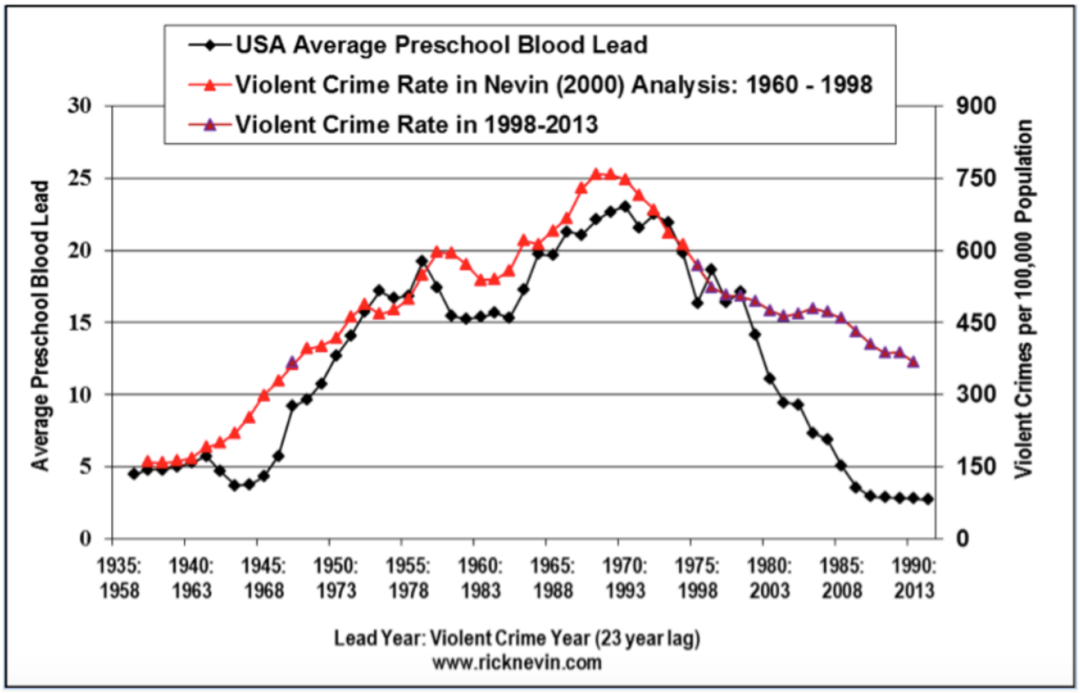


Figure 23. Lead article; first graph

The graphs from the lead article were generally better understood by the seven interview subjects who received the lead article in their version of the PLHOM. The first graph in the lead article (Figure 23) was generally well understood by the interview subjects.

All seven of these interview subjects indicated understanding what this graph shows, a correlation between blood lead levels in pre-schoolers and violent crime rates. Three subjects indicated confusion about the time lag in the graph and two indicated concerns about the reliability of the source, www.ricknevin.com. Of the three students who showed confusion about the time lag, two did not at first understand the relevance of the time lag to the authors' argument, while the other showed difficulty in communicating the relevance of the time lag. For example, consider the following discussion with a student who will be referred to as Nina. Her initial response to question 3 of the PLHOM indicated she may have misunderstood the time lag in the first graph; the following is a follow-up discussion about her response. Nina begins by reading information relevant to her response from the article:

Nina: *The lead paint ban, removal [of] leaded gasoline from materials filling stations and lead abetment efforts which all described lead exposure, particularly in children born around 1975 to the late 1980, correlated strongly to criminal children who hit peak age in the 1990's and early 2000's… uh that data suggested that this specific cohort was less likely to get arrested for specific crimes.*

PI: So, do you think you would change anything about how you said that [her response]?

Nina: *Uh, maybe the wording in it?*

PI: Uh huh.

Nina: *Just that this is at the age of people associated to it, not necessarily the actual people involved in it.*

PI: Okay, so you said um that they both happened at the same time though…

Nina: *For…? Hmm, sorry. I'm really bad at explaining stuff on paper*.

[*Laughing*]

… …

PI: So, maybe explain to me what you're thinking?

Nina: *That having these people growing up in more of a time where there's less lead in common items such as gasoline and paint that we're exposed to all the time…*

PI: Mhmm.

Nina: *Um… that these are also people that are going to have certain amounts of crime associated with them once they hit a certain age.*

PI: Okay. And so, then the people that were exposed to lead…

Nina*: Mhmm, um… that they would then have… crap… the people exposed to lead would then have the uh... I'm confusing myself…*

It is the opinion of the PI that Nina more or less understood the significance of the time lag in the first graph, however, she had a lot of difficulty in communicating this. On the other hand, Art had no such difficulty and clearly explained the relevance of the time lag:

Um, well there's a 23 year lag, so I'm assuming that's um... maybe the average age or the uh... actually I should have looked at that more closely, but um, yeah so... so lead year, violent crime year, 23 year lag, so I would expect the first one to be the lead year, when you're exposed, and then um, again 23 years later you're a young adult and that's when they are taking the uh, the crime stats.

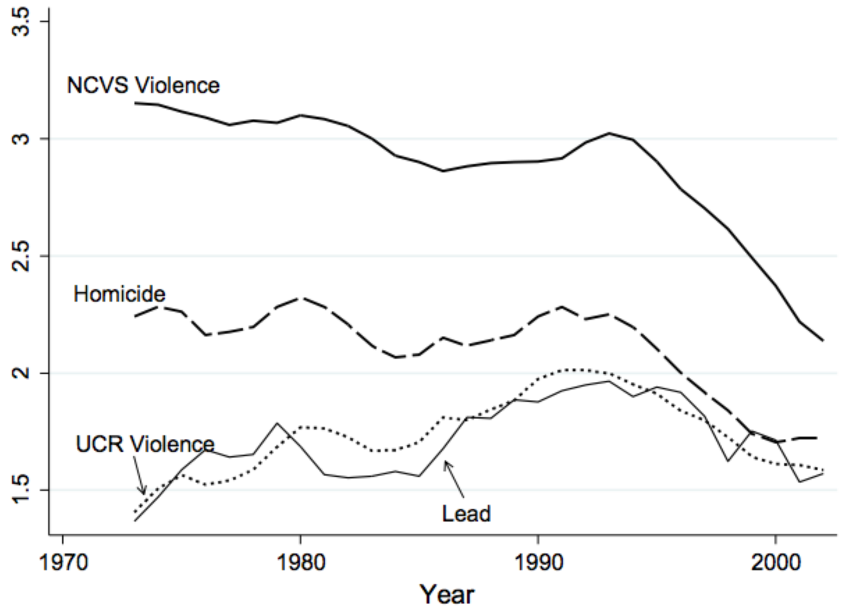


Fig. 3 Logged rates of gasoline lead exposure, homicide, and NCVS and UCR serious violence, 1973–2002. *Note* gasoline lead measure is lagged 23 years and rescaled to improve visual trend comparisons

Figure 24. Lead article; second graph

The second graph from the lead article (Figure 24) was equally well understood by interview subjects. Only one student gave a response to the PLHOM that indicated they misunderstood what this graph attempts to show, a correlation between violent crime rates and gasoline lead exposure (lagged 23 years) according to UCR reported rates and little to no correlation between violent crime rates and gasoline lead exposure (lagged 23 years) according to homicide rates or NCVS reported rates. This student included *(t)he author used the UCR to show the national crime rate over a certain period of time to show how the lead theory is not correct* in their response to question 3, however, they cleared up this misconception during follow-up questions; the student later said that the UCR statistics support the lead theory and that they meant to say it was the NCVS statistics that dispute the lead theory. When pressed about what is shown in this graph several students indicated some confusion, however, in every case they were able to communicate at least some understanding of the graph. Interestingly, only two students noticed that the *y*-axis on this graph is not labelled or explained in the article. Further, both students only realized this during follow-up questioning by the PI. Art realized this as a serious weakness in the use of the graph in this article. When Art was asked what the second graph shows, part of his response included:

> *those rates, uh, I mean this right here, I don't know if I should assume the same, like I don't know what these numbers are is it... crimes per 100,000? As it alluded to over here? or... or it could just be, you know, nothing, Shrute bucks.*

Shrute bucks is a reference to an imaginary currency in the popular television series *The Office*. Lora came to a similar realization, though she did not view this as significant of a problem as Art did:

> *Yeah, just cause one… the y-axis isn't labelled, so I don't know what that is... um, I almost think, I first thought, I thought it was um, like a stacked graph that we went over in MATH [MAT 114] but I can't exactly remember what a stacked graph is... so I'm like why are all these lines above each other and like...*

Lora, after puzzling over the graph and the article some more, indicated she understood what the graph shows and how it fits the narrative of the article.

From the interviews it seems clear that subjects had a harder time understanding the graphs from the incarceration article than in the lead article. Further, most subjects showed difficulty communicating the information shown by any graph, and, in the case of the incarceration article graphs, subjects had a lot of difficulty interpreting the graphs. It seems clear that QR students generally need more practice interpreting and communicating graphical data. Further, considering the difficulty students had in interpreting a graph of percent decreases, future implementations of the PLHOM should avoid pre/post-testing with articles where one contains a graph of this type and the other does not.

The extent to which the subjects viewed the relative reliability of the data/sources provided in the articles was not investigated in all interviews. Questions in the interview protocol that relate to this were only used when responses reached scores of a 2 or 3 on individual questions. A good base-line indicator of HoM in students is an initial response that receives a total score of 4 or higher (achievable with level 1 responses to all questions). Three of the five students who had completed MATH 1313 matched this criteria and two of the nine students enrolled in MAT 114 matched this criteria. Indicators of HoM potentially missed by the initial application of the PLHOM were cases where follow-up prompting resulted in a total score of 8 or higher (achievable with level 2 responses to all questions). Two of the three students who had completed MATH 1313 and scored at four or above on the initial application of the PLHOM matched this criteria, and one of the two students enrolled in MAT 114 and scored at four or above on the initial application of the PLHOM matched this criteria. One student enrolled in MAT 114 who did not score at four or above on the initial application of the PLHOM matched this criteria.

All three of the students who had completed MATH 1313 and scored at or above a 4 on the initial application of the PLHOM showed skepticism in the reliability of the sources used in the articles. One of the students enrolled in MAT 114 that scored at or above a 4 on the initial application of the PLHOM showed skepticism in the reliability of the sources/data, whereas, the other – Nina – seemed to believe the authors used reliable sources; Nina also noted that the narrative of her article (lead) matched what she had learned in her criminology and criminal justice courses. Nina, much like Chayla from before, might have let her biases influence her. Lora, the student enrolled in MAT 114 that did not score a 4 on the initial assessment but whose follow-up prompting resulted in a total score of 9 on the PLHOM, indicated a high degree of skepticism about the reliability of the sources/data in her article.

Further, the three students from MAT 114 with higher levels of PLHOM measured HOM all indicated different areas in their lives they drew on when responding to the PLHOM; Nina drew on her criminology and criminal justice background, another student said she likes to read academic articles about space and crime novels, and the last of these three said they regularly read and evaluate news from social media (Facebook, Twitter, etc…). The three students from MATH 1313 with higher levels of PLHOM measured HOM also indicated different areas in their lives they drew on when responding to the PLHOM. One student indicated she relied on her English literature background in evaluating sources and mentioned critiquing news articles in MATH 1313. Art said he reads the news regularly and that MATH 1313 helped him better understand and evaluate graphical information. The last of these three did not specifically say what they drew on while responding to the PLHOM, however he showed a tendency to try and explain things in hypothetical, real-world contexts, a focus of MATH 1313; for example, when this student was asked to describe what the curve for MV theft in the graphic seen in Figure 2 illustrates, he responded:

> *Okay, I guess it's just... I don't know... my guess would be just, I feel like we have better security on cars now than we did in 1981 I guess, so we have car alarms and you know people have, have really customized their cars to be a certain way. um... like they make em like bullet proof now, I guess they could back then, but, umm... yeah so I guess just like the increase in technology as well... uh, like I don't understand why the robbery has gone up like that with home security like that's gone up and then... I guess stayed at a certain level, but um, I guess I'm, I don't know, that's kind of tough... um... Oh I meant burglary, not robbery is, just home security.*

Regardless of the accuracy of this explanation – involving a misinterpretation of the second graph from the incarceration article –, it is interesting that he felt compelled to provide an explanation in real-world terms of what could have caused the behavior in the graph. Further, he was the only interview subject with this tendency. This student also made references to QR relevant anecdotes from professors in his major area of study (sports and recreation management). Of the students who did not meet either of the PLHOM criteria discussed above, three indicated drawing on their experience in criminology and criminal justice, three others indicated that they don't read news articles, Chayla indicated she drew on her experience as an African-American studies major and Cory, who attempted to draw on his experience in MAT 114, was unable to connect material from MAT 114 to his analysis on the PLHOM in a meaningful way.

The PI believes this illustrates the importance of grounding QR in authentic contexts. Students indicated drawing on life experiences situated in context while responding to the PLHOM: reading articles about space, reading crime novels, studies in their major area, and reading news articles in or out of the QR classroom. Art, the interviewee with the highest PLHOM scores, is a regular reader of the news and had relevant QR classroom experience (critiquing articles, interpreting context situated graphs). Nina and Chayla, interviewees at one point above the median PLHOM score and at the median PLHOM score respectively, gave responses infused with influence from their major area of study. Two of the three interviewees with the lowest PLHOM score said they do not read the news. Further, one of these students, when asked about her reported weakness (*(t)he facts may not have correct citing*), said she *didn't know what to put* and seemed lost. In addition, it seems likely that the PLHOM misses HoM not specific to media articles and could be highly influenced by the content of those articles.

INTERVIEW ANALYSIS – SECOND PORTION

The second portion of the interviews has interview subjects answer the two course-like material questions from Figures 1 and 2 in the Methodology. Students are then asked follow-up questions from page 6 of the interview protocol viewable in Appendix E. This portion of the interviews seeks to inform about similarities and differences between opportunities to develop habits of mind in MAT 114, MATH 1313 and MATH 2183 from the students' perspective. This portion of the interviews also seeks to inform about obstacles to HoM QR students face as well as their view on types of problems where reasoning transfers to a setting other than the QR classroom or academia in general. Specifically, the goals of this portion of the interviews were to:

a) Determine if the student views course-like materials in the same way as the PI.
b) Determine the relative perceived difficulty between the two course-like materials.
c) Identify difficulties the student had when responding to the course-like materials.
d) Identify how the students believe they think about the two problems.
e) Identify the student's views on the applicability of the two problems to his or her life outside of the classroom.

Note that this portion of the interview was not conducted in the single interview where the student had nearly completed MATH 2183. Nine interview subjects had nearly completed MAT 114 and the other five had already completed MATH 1313. The two problems will be referred to as the Harps problem and the bricks problem (Appendix D). Note that in *Quantitative Reasoning* (Fahy and Nabours, 2016), the course text for MAT 114, the problem on which the bricks problem is modelled does not ask the students to explain their responses; this was added by the PI.

The bricks problem was scored out of 5 points, with one point awarded for each correct match. Note that 7 and 7/40 are both considered correct responses to match with *statistic*. The Harps problem was scored on a scale of 0 to 2 with 0 representing complete confusion and/or no indication that any discrepancy in the Harps ad claim exists, 1 representing that the student recognizes some discrepancy in the Harps ad claim and justifies with QR, and 2 representing that the student recognizes a clear discrepancy in the Harps ad claim and justifies with correct QR. A student needs to point out that a literal interpretation of the ad results in negative sodium levels for Harps' chicken in order to score a 2.

Eight out of the nine students from MAT 114 identified the bricks problem as more like MAT 114 problems – one said neither – , and all five students from MATH 1313 identified the Harps problem as more like MATH 1313 problems. Monica from MAT 114, when describing the Harps problem, said she had *never seen anything like [it]*. Two students from MATH 1313 identified specifically the part of the bricks problem that asks for an explanation (modified by PI) as being like MATH 1313 problems. From this we can assume that the PI's classifications were correct. Overall, students identified neither problem as more difficult than the other. Six students identified the Harps problem as easier, six students identified the bricks problem as easier and one student said neither was more difficult than the other. Several students indicated that the open-endedness of the Harps problem made them slightly uncomfortable.

Students were more successful responding to the bricks problem than the Harps problem. Students from MAT 114 averaged 78% correct on the bricks problem and were all scored at the 1 level on the Harps problem. Students from MATH 1313 also received a sheet of definitions for the bricks problem since MATH 1313 typically does not cover all of the bricks problem's material. Students from MATH 1313 averaged 64% correct on the bricks problem, one scored no points on the Harps problem, one scored 2 points on the Harps problem and the other three scored 1 point on the Harps problem. Eight of the 15 students interpreted the Harps ad as meaning Harps chicken contains one fifth the sodium of its competitors. Art from MATH 1313 was the only student to identify that a literal interpretation of the Harps ad results in Harps' chicken containing negative sodium. The only student that seemed completely baffled by the ad and received a score of 0 on the Harps problem was also from MATH 1313. Cory, from MAT 114, claimed that a reduction from 100 mg to 80 mg was a 5% reduction, and Lisa, from MAT 114, interpreted the meaning of the ad as Harps chicken contains one fifth the sodium of its competitors but explained this as *for every mg of sodium in the competitor's chicken, Harps has 5 mg less*. A student from MATH 1313, who also took the one-fifth interpretation, later described this as 20% less. Here, we see that the language of arithmetic and percentages poses a significant challenge to QR students. Similar to interpreting a graph of percent decreases, the language of arithmetic and percentages represents a significant obstacle to a QR HoM. Interestingly, during follow-up questioning, two MAT 114 students indicated they considered the idea that the Harps ad implies Harps chicken should have negative sodium. They wrote the thought off as too ridiculous and defaulted to a one-fifth interpretation. This suggests a lack of mathematical confidence hindered their responses.

In response to *do either of these problems incorporate the type of reasoning you expect to use outside of a purely academic setting?*, 10 of the 14 subjects identified the Harps problem as most resembling the type of reasoning they expect to use outside of a purely academic setting and the other four identified both problems. The reason the PI asked this question stems from the assumption that course materials students identify as relevant outside of a purely academic setting are the most likely to support or stimulate a HoM. When asked directly about the reasoning used in responding to the two problems, student responses varied; however, certain themes in the responses are present. While all students indicated they thought the Harps problem involved reasoning they would expect to use outside of a purely academic setting, three students, while describing the reasoning they used on the Harps problem – before they were asked if problems incorporated reasoning they expect to use outside of class – went as far as to use a variation of the phrase *real-world*:

- *this one makes me think of like real life so I kind of just imagine myself in a grocery store and reading that sign... being like "oh, that sounds pretty good" like I want less sodium in my chicken so obviously I'm going to buy that one... so I definitely applied this more real life... uh... especially just because um, you're looking at less numbers and it's more word problems so I feel like that leads to more real life thinking whereas this one is... I still imagine myself at a lot, buying 1000 bricks, but this one has a lot more numbers, so I just feel like... I thought a lot more... realistically... (taps fingers) on the harps*
- *maybe because it's kind of real world*
- *I feel like this is just more um useful in real life than this one is and I don't really ever hear terms from this paper*

Further, four students directly indicated that they did not see the bricks problem as *real-world*:

- *I don't think you really need to know like population and all that, sample sizes, with everyday life no matter like what your profession may be*
- *I don't really ever hear terms from this paper. Of course population, sample maybe, but it's not really something that a lot of people would be thinking about a lot of the time*
- *just kind [of], not so much like [a] real life situation*
- *the bricks question you would definitely, obviously… … you could run into cracked bricks and that type of thing, but you wouldn't be thinking of it in terms of sample sizes and... statistic, though; what it's asking you to do isn't what you would do in normal life*

Of the four students who claimed they could see using reasoning from the bricks problem outside of the classroom, they all indicated the reasoning for the bricks problem could be useful to certain professions/professional activities (e.g. contractors, air force lieutenant, data collection, etc…).

When describing the reasoning they used on the two problems, two students mentioned that they were not really reasoning during the bricks problem until they were asked to explain their reasoning – an addition made by the PI. Further, eight students changed their answers as a result of trying to explain their reasoning. Lora, from MAT 114, draws a clear image of her experience while responding:

> *okay, so for that like, I did use more reasoning because I had to kind of figure out why I chose each answer that I did and figure out like, oh why did I choose that? And then there I was, standing in a lot... looking at a 1000 bricks in my head*

Further, three students mentioned that the bricks problem had a *definite answer*, three others said they simply tried to *remember* how they were supposed to respond, one student referred to the bricks problem as *pure math*, one student referred to the bricks problem as *textbook* and *clinical*, one student called the bricks problem more *mathy* than the Harps problem and one student said the bricks problem requires *critical thinking*. In contrast to this, two students described the Harps problem as *open-ended*, one described her response to the Harps problem as *creative*, one described the Harps problem as *conceptual*, one claimed her response to the Harps problem required *deep, analytical thinking* and six students described needing to *interpret* the ad in their response.

From these analyses it seems that some students view reasoning about real-world phenomena as fundamentally different from reasoning about something they see in a purely mathematical context. One student, while describing the Harps problem as more *real life*, said her reasoning on the bricks problem felt *opposite*. Further, the addition of *explain your reasoning* to the bricks problem generally spurred deeper reasoning and even made the problem more vivid, at least definitively in the case of Lora. The PI's impression during these interviews was that students saw real merit in connecting a quantitative analysis to real-world phenomena. Further, the Harps problem – course-like material for MATH 1313 and MATH 2183 - was generally viewed as more *real-world* than the bricks problem, and the described thought processes on this problem were more complex, varying and perhaps less *textbook*.

INTERVIEW ANALYSIS – THIRD PORTION

The third portion of the interviews consists of three primary questions:

1. Would you say that your QR course is different from a typical mathematics course? If yes, how so, and if no, could you describe what a typical mathematics course means to you?
2. How would you apply any of the techniques/reasoning you use in class to other areas of study or outside of an academic setting?
3. Could you describe quantitative reasoning in your own words?

This portion of the interviews seeks to inform about similarities and differences between opportunities to develop habits of mind in the three QR courses from the students' perspective. Further, this portion of the interviews seeks to inform about similarities and differences in the ways that students may apply a HoM. This portion of the interviews broadens the focus to student opinions about the entirety of the course taken. Note that both interviews from the student who interviewed twice, once for MATH 1313 and once for MATH 2183, are included in this analysis.

Students generally viewed their QR course as different from typical mathematics courses; all five students from MATH 1313 indicated MATH 1313 was different from a typical mathematics course, and eight of the nine students from MAT 114 indicated MAT 114 was different from a typical mathematics course. While describing how the courses differed from a typical mathematics course, 12 of the 15 responses included a reference to the applicability of course content to *life, real life* or *everyday life*. Lora's description about what made MAT 114 different is typical for the seven MAT 114 respondents who identify applicability to *life* as a primary difference: *I'd say there is a lot more reasoning behind it because there is a lot more* *word problems and like you're still looking at numbers and stuff but it is a lot easier to apply it to your life.* Art's description is similarly typical for the four MATH 1313 respondents who identify applicability to *life* as a primary difference:

> *I don't consider myself a big math person, although I honestly don't feel like what we were doing in there was really, mathematics? … … I think that there was just a larger component of what I alluded to before, um, logic and reasoning, the math that we did, a lot of it was just simple multiplication, division... um, things of that nature, but using simple functions to... solve problems of everyday, you know, everyday life, like, interest rates and um... gas usage, utility usage, things like that.*

Several students drew comparisons with other math courses:

- *Yeah it uses, like, algebra but I mean it's, like, it's different kind of, it's just kind of a different type of math than I've, like, ever taken before, where I normally just, like, taken, you know, math like, algebra one, algebra two. One's just like a lot of, simple algebra. Two is, gets into, some calculus stuff. So, I've, and a lot of when I was in high school I used to, like, complain, like, what is this used for? And so, and certain things that can be useful but, so I just thought it was different because just, like, you-you can actually see where you can use it.* (MAT 114)
- *I think College Algebra is more concrete. You know, you have the formulas. This equals this plus this. These reasoning courses [are] ambiguous… …it's open to interpretation. [They] can mean a lot of stuff. I think you see more of this, like mathematical reasoning, I think you see that more in everyday life than you would College Algebra class or Finite or something like that.* (MATH 1313)
- *I guess it's just like more real world applications of like how to use math which is something that I like more than just having to figure out like algebraic formulas all day, um, like with the-the loans and the interest and stuff that all has very like real world sort*

of like applications to it so I find it very useful and like good for me to learn essentially. (MAT 114)

Also, five MAT 114 students identified course structure (flipped) and the variety of topics as primary differences. Three of the students from MATH 1313 and the student from MATH 2183 said course instruction differed from typical mathematics courses; they highlighted a higher level of explanation, greater variety in examples, and group work in their descriptions.

In responses to the first and second questions from this portion of the interviews every student identified something from the course that they thought could apply to their lives outside of the QR classroom. All responses from the MAT 114 students identified financial mathematics topics – profit, savings, loans and interest – from MAT 114 as relevant to life outside the classroom. Most students who identified financial mathematics topics specifically mentioned a buying versus renting project that made heavy use of savings and loan formulas. Nina was especially detailed when describing how she applied compound interest formulas outside the classroom:

> Nina: *I was going through and just seeing like how many deposits I've made for work like and stuff and seeing like, okay if I deposit this, this, and this per each month how much can I have by next year so I can see like how many extra classes can I pay for, or um can I afford to fix this on my car? Things like that.*
>
> PI: Okay. Do you remember which formula you were using?
>
> Nina: *I, I think I used the savings and then the compound interest because I've got one savings account that I've had since I was a kid that my grandmother used to put money into, and then I've just got my other account that I use for just whatever I've earned at work.*
>
> PI: Okay.
>
> Nina: *And I was trying to figure out what will I need to draw from, from my original account in order to pay for college and do this kind of stuff.*
>
> PI: Okay, cool.

Nina: *And figuring out how much I could kind of pull from without like dropping it too much in the original account.*

PI: Right. Okay, and so do you think you would've been able to do that before 114, or did that actually really help you?

Nina: *That actually really helped me because I think before 114 I would have needed to get help from somebody else that knew what they were doing, or at least knew the equations that could help me with it, but once we learned those I could kind of figure out a general area about what I could do.*

Interestingly, financial mathematical topics were the only topics identified as applicable to everyday life in six out of the nine MAT 114 interviews, and financial topics where the only topics – other than three references to the bricks problem from earlier in the interview – described in any detail and with references to specific assignments. Several MAT 114 students claimed that everything in the course was applicable, but most of these students had difficulty remembering the other topics that could apply to life outside of the classroom. This suggests that other course topics – of which there are many; financial topics are the primary focus of only four of the 13 course modules – are less connected to authentic applications in the minds of these students. Further, four MAT 114 students indicated they saw little applicability in certain course topics. Two of these students specifically identified the graph theory as less applicable to their lives and one student said *when it came to using the formulas for mortgages and loans… … I feel like that's the only thing that made sense. The rest I feel like I didn't need to know.*

MAT 114 has a stronger mathematical focus than either MATH 1313 or MATH 2183, and this mathematical focus came up in several MAT 114 student interviews. Two students indicated that using finance formulas and understanding the mathematics behind the formulas was relevant to them; however, two other students wished there was more focus on real world applications. Monica, who received a B as her final grade, made this point clear:

I'd rather learn about what each of those things [savings, loans and interest] really are and like how, how significant they are in real life and like how to make smart decisions

> *about things like that, regarding like finances and loans and interest and you know whatever, buying cars and things like learning more about how to make decisions that make sense and are beneficial in life, and not so much focused on like exactly the amount of money that was saved, or exactly what the interest is, or in, in just example problems.*

Monica's comment about *example problems* indicates she sees little use in traditional, usually hypothetical, word problems – problems common in MAT 114. Monica also gave a fairly accurate description of QR: *applying mathematical concepts to abstract, like real-life scenarios*. Here, the PI believes Monica uses *abstract* to mean *complex*. When asked if she felt MAT 114 prepared her to apply mathematical concepts to abstract real-life scenarios, Monica responded, *not very well*. Though Monica's view seems to be an exception compared to most other MAT 114 interviewees, removal of some math-centric topics like graph theory could make room in the curriculum for more authentic applications.

Topics applicable to life outside the QR classroom identified by MATH 1313 students varied. Art identified interest rates, gas/utility usage, estimation, personal finance topics and depreciation models. Another student said she found gas/utility usage relevant, and added that she *[doesn't] know why but they made a lot of sense to [her];* this student also indicated she found applying QR to newspaper articles relevant to her life outside the QR classroom. Chayla said she learned a lot about taxes, changed the way she calculates tips and did a QR project on rainforest deforestation in the Amazon. Another student recalled his projects on the melting glaciers and coral bleaching, and also said he learned a lot about real estate taxes, home expenses and estimation; this student described learning about the finances involved in home ownership as *an explosion in [his] mind*. Adam, who interviewed for MATH 1313 and MATH 2183, discussed the applicability of budgeting, car loans, home loans and tax rates during his MATH 1313 interview.

Adam – who received a C and did not like his instructor in MATH 1313 – gave a convincing description of how relevant MATH 1313 is to life outside the classroom:

> Adam: *I think it's more outside of academic life, like in 1313, what [instructor] actually did was like at the very end of class, I'm trying to remember… … he had us, you know, like write down what type of job we [were hoping] to have, you know, how much we would make and then, car loans, you know, homeowner loans, and, you know, tax rates and stuff like that, he applied it to stuff we [were] going to have to go through when we graduate college.*
>
> PI: Yeah.
>
> Adam: *I think it would be applied more outside academia*.
>
> PI: And so, did you feel like you got much out of that?
>
> Adam: *Yeah. A lot.*
>
> PI: You liked that.
>
> Adam: *Yeah, because you know we don't think that, like in college, we don't think about things like that.*
>
> PI: Yeah. And did you feel like, there's a way for that to be relevant and there's a way for that to sort of be like, you know what I mean like, did it feel authentic, I guess is what I'm asking. Like did you feel like you were actually figuring that kind of stuff out or did it feel like it was just set up like a math class?
>
> Adam: *I think it was authentic. I think it was genuine.*

In Adam's second interview he discussed MATH 2183's focus on news media articles. Adam – a political science major – identified a specific case study from early in the course relating to mathematical fallacies in the media that he said had broad application outside the classroom. Interestingly, this corroborates Adam's observed shift in HoM discussed earlier; in Adam's first interview he took the existence of graphs in an article as proof of their credibility, and he showed a shift toward skepticism during the second interview.

In Adam's second interview the PI also asked him to compare MATH 1313 with MATH 2183. Adam said he felt that material in MATH 2183 was more *complex* and added:

> *1313 was the simple stuff, percent increase, you know, just small stuff like that, or compound interest. This stuff… …in 1313 we didn't focus on looking at case studies.*

> *There was no case studies. It was more pure math. 1313 was kind of leading up to this mathematical reasoning where you had to look at a case study and you know draw out the numbers and make sense of them.*

This fits the QLAR analyses as well; the text for MATH 2183, the QRCW Casebook, is richest with respect to the core competencies and particularly so with respect to *interpretation*. Adam also talked about his views of mathematics before MATH 1313 and after MATH 2183. Adam used strong language in his description:

> *In a sense, before I took these two math classes, math was something that I wanted to stay away from; to me, math was complicated, right?*
>
> *… and that's why I waited, I waited till my senior year, I put it off so much, you know, I waited till the last moment of my senior year to take it. But, you know that view, it changed, it switched once I took these two courses it changed.*
>
> *… I think it's not as complicated as I thought it was. [Because] I feel like I know math now. I feel like once you actually sit down and learn, it's not as complicated as it seems so... and I actually feel like I could take more math classes now.*

If true, this represents a dramatic shift in Adam's disposition toward mathematics and is quite remarkable. Now, Adam received a C in MATH 1313, but had an A in MATH 2183 at the time of his final interview. The PI believes that Adam's success in MATH 2183 and his interest in *authentic* or *genuine* applications influenced his dispositional shift. The PI further believes that Adam's success in MATH 2183 and his interest in *authentic* or *genuine* applications influenced each other. Adam also reported that he liked his instructor.

In all responses to the final portion of the interviews students indicated they saw value in real world application. Further, most students felt their QR course was rich with application, and the students who did not were unhappy with their course. Students were also better able to remember applications they connected with specific, real-world contexts, and showed difficulty remembering applications where this connection was artificial or unexplored. Students generally viewed their QR course as very different from traditional mathematics classes; most attribute this – at least in part – to their course's applicability to the real world. This analysis indicates the importance of authentic, student relevant context in QR course materials; authentic, student relevant context seems to promote application recall and it is continued practice with application that should promote a HoM among any individual.

In response to the third question, students gave varying descriptions of QR. Below is a sample of responses that best represent student descriptions of QR:

- *Logical reasoning and thinking… it's very, in my opinion, logical*. (MAT 114)
- *Applying mathematical concepts to [complex] like real-life scenarios*. (MAT 114)
- *I first saw it as statistics like when we first started out dealing with a lot of percentages and stuff.* (MAT 114)
- *Numbers that we see in everyday life*. (MATH 1313)
- *Taking numerical expressions and applying the process of reason to them… … You're using reasoning but you're, you actually have quantities in there, like mathematical like numbers* (MATH 1313)

The PI compared all responses to this question with corresponding scores on the PLHOM; however, no discernable pattern emerged. The only thing notable about responses is that students showed difficulty in responding to question three.

Several students used the end of the interview to vent their frustrations with the course. Two students from MAT 114 and two students from MATH 1313 voiced complaints about their instructor. Adam and Chayla complained about the same instructor; they said the instructor offered little guidance and the grading was often late. The two MAT 114 students complained about rigidity and a lack of explanation in the grading. Monica complained that she lost points on quizzes and projects but she didn't understand why. She was again descriptive with how she wished the course (MAT 114) could have gone:

> *With words you can kind of express, express yourself in many different ways and kind of get your point across, but with numbers it's like, you could always like, you can do one thing and the whole thing will be wrong. But with words it's like some part of what you say might be right. Some part of what you say might be wrong, but it doesn't necessarily mean if you say one thing the entire thing [is] completely wrong.*

In contrast to this, Joey from MATH 1313 – who claims to have *dyscalculia* and described coming up with solutions to MATH 1313 problems as *like digging through concrete with a spoon* – relays a much different experience with respect to grading:

> *I realized that it wasn't... ...he wasn't asking us to come up with what the book wanted us to do, so as long as we showed a process that made sense at all, he was okay with that, and we didn't have to get the right answer all the time as long as we showed a thorough thinking process so that helped me relax about it a lot, so... … as long as I showed a process, let's say that I didn't get the right answer, he might take a point off for the wrong answer but give me three points for showing a correct calculation even if it either had a small error or even if it wasn't the exact thing he wanted me to do. If what I did in the calculation was correct, even if it was the wrong thing to do sometimes.*

Now, this could easily be explained through differences in the particular instructor each student had. However, the PI has experience teaching both courses, and believes there is another influence to these opposing experiences. MAT 114 is heavily coordinated and has a lot of graded work; instructors are encouraged to grade fast but fairly, and final answers are where students earn the majority of their points. MATH 1313 is more loosely coordinated; however, in the experience of the PI, instructors are encouraged to grade written work, and look for relevant or partially correct reasoning in order to scaffold learning. Grading was not a widely reported issue from MAT 114 interviewees, but the stark contrast in the above student descriptions is interesting.

QUANTITATIVE ANALYSIS – PRE/POST TESTING

Pre/post testing in sections of MAT 114, MATH 1313, and MATH 2183 investigates habits of mind demonstrated by QR students and opportunities to develop a HoM in MAT 114. Students responded to the prompt-less instrument (PLHOM) designed by Boersma and Klyve (2013) in the pre-post assessments. The articles used in the online versions of the PLHOM are viewable in Appendices A and B. In total, there were 98 responses to the online version of the PLHOM. 83 of the responses came from students enrolled in MAT 114, 13 from students enrolled in MATH 2183, and 2 responses from a student enrolled in MATH 1313. Further, only 19 students filled out both a pre and post version of the instrument with the pre-test occurring within the first two weeks of classes and the post-test occurring within 2 weeks (before or after) of the final day of classes. Of those 19 students, 17 were enrolled in MAT 114, 1 was enrolled in MATH 1313, and 1 was enrolled in MATH 2183. Further, 2 of the 17 MAT 114 students who filled out both a pre- and post-test mistakenly used the link for the pre-test whenever they took their post-test, and so they filled out the PLHOM twice for the same article.

Using the PLHOM for pre- and post-testing is problematic because of the use of two different articles. It is possible that one article is, in a sense, easier than the other. To investigate this an unpaired, two-tailed *t*-test was conducted on pre-test total scores between the two articles. Table 6 illustrates the differences in pre-test scores between the 2 articles; note that a score of 12 is the maximum possible score on the instrument. The difference in means was not statistically significant at the $p = 0.10$ level, however, interview data indicates the Incarceration article contained a graphic that was especially problematic for students to interpret successfully. In theory, this should most affect scores on questions 5a and 5b from the PLHOM. Table 7 illustrates the differences in pre-test scores on the sum of scores for questions 5a and 5b between the two articles; note that a score of 6 is the maximum possible score for this sum. Again, the difference in means was not statistically significant at the $p = 0.10$ level.

Table 6. Pre-test mean total scores for PLHOM by article

Article	Mean	*N*
Lead	2.48	33
Incarceration	2.11	35

Table 7. Pre-test mean sum of scores for questions 5a and 5b by article

Article	Mean	*N*
Lead	0.97	33
Incarceration	0.77	35

Interestingly, including post-test data when comparing total scores between the two articles yields a statistically significant difference in means at the $p = 0.10$ level ($p = 0.056$). However, adding post-test data to the sum of scores on questions 5a and 5b does not result in a statistically significant difference between articles at the $p = 0.10$ level. It is notable that in all of these analyses, mean scores for the lead article are higher than mean scores for the incarceration article. Table 8 summarizes the results of this of this analysis.

One possible extraneous variable in this analysis is the variability of the student cohorts. In particular, only 3 of the responses to the incarceration article came from students enrolled in MATH 2183, whereas, 10 of the responses to the lead article came from students enrolled in MATH 2183; MATH 2183 is technically a higher level course than both MAT 114 and MATH 1313. A second analysis was conducted to compare mean scores between articles using only the data from students enrolled in MAT 114. Eliminating data from MATH 1313 and MATH 2183 yielded a non-statistically significant difference in mean total scores between articles at the p = 0.10 level. Table 8 summarizes the results of this analysis.

Table 8. Mean total scores by article and cohort, and mean sum scores of questions 5a and 5b by article

Article	**Mean**	***N***
Lead Total Score	2.73	55
Incarceration Total Score	2.02	43
Lead Sum of 5a and 5b	1.16	55
Incarceration Sum of 5a and 5b	0.77	43
Lead Total Score (MAT 114)	2.52	44
Incarceration Total Score (MAT114)	2.03	39

Overall, these analyses provide very weak evidence that the incarceration article may have been more difficult for students to interpret and critique than the lead article. Mean scores were consistently higher on the lead article, however, the lack of statistical significance in these mean scores in all but one, potentially flawed analysis, signals the difficulty of responding to the PLHOM between articles was not significantly different.

With this in mind, several analyses were conducted in order to investigate any possible gains in habits of mind made by students over the course of the semester in which they were enrolled in a QR course. A paired, two-tailed t-test was conducted on total scores for the 17 students who successfully completed a pre- and post- test assessment with the PLHOM. Table 9 summarizes the data from this analysis.

Table 9. Mean total pre- and post-test scores of students who completed both the pre- and post-test PLHOM

	Mean	***N***
Pre-test scores	2.82	17
Post-test scores	2.76	17

An unpaired, two-tailed *t*-test was conducted on total scores for all MAT 114 pre- and post-test results. Table 10 summarizes the data from this analysis. Neither of these analyses resulted in a statistically significant difference in means at the $p = 0.10$ level, indicating no significant changes in students' habits of mind. MATH 2183 student data ($N = 13$) were omitted from these analyses since including a small amount of student data from a higher-level course could skew the data. MATH 1313 student data ($N = 2$) was omitted because of the low response rate.

Table 10. Mean total pre- and post-test scores of MAT 114 responses PLHOM

	Mean	***N***
Pre-test scores	2.16	63
Post-test scores	2.70	20

Especially noticeable from these data are the relatively low scores on the instrument overall. The maximum score of 12 was not observed. No analysis resulted in a mean score of 3 or higher. Individually, three responses scored an 8, the highest observed score on the instrument, whereas, eight responses scored a 0, and another 29 responses scored a 1. Table 11 illustrates the frequency of each score from the online responses to the PLHOM.

Table 11. Frequency of scores from the 98 responses to the PLHOM

Score	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
N	8	29	21	20	6	9	1	1	3

CHAPTER 5: CONCLUSION

INTRODUCTION

The purpose of this research is to respond to the following two research questions:

1. What are similarities and differences between opportunities to develop habits of mind in three different QR courses?

   a. Curriculum and design

      i. Core competencies

      ii. Conceptualization of QR

2. What habits of mind do QR students demonstrate?

   b. What similarities and differences in habits of mind exist within and between students?

      i. Application

      ii. Obstacles

The PI investigated these questions through two main goals:

1. Identify opportunities and hindrances for QR students to develop a HoM toward QR in three different QR courses.
2. Identify where and/or how QR students apply a HoM, and identify obstacles students face in developing a HoM.

The first and second goals reflect on the first and second research questions respectively. Application of the QLAR rubric from Boersma et al. (2011) and a description of objective/possible content progressions in the course texts pursue the first goal. Qualitative analysis of student interview responses to the prompt-less instrument (PLHOM), student interview responses to follow-up prompting based on original responses to the PLHOM, student interview responses to course-like materials and student impressions of QR courses also pursue the first goal. Qualitative analysis of student interviews and a combination of quantitative and qualitative analyses of student responses to the online version of the PLHOM pursue the second goal. Further, the PLHOM – the primary instrument used throughout this research – was itself investigated through the qualitative analysis of student interviews.

COURSE TEXTS – ANALYSES TOWARD THE FIRST GOAL

Quantitative Reasoning (Fahy and Nabours, 2016) is a course pack that students fill out over the course of the semester in which they are enrolled in MAT 114. This course pack progresses linearly through the following mathematical content progressions:

- Basic statistics and the central limit theorem (BSCLT)
- Linear/exponential models and finance (LEMF)
- Graph theory and scheduling (GTS)

The mathematical content progressions include real-world applications, but primarily in the form of hypothetical word problems. The PI applied the QLAR rubric to 190 questions from a sample of 5 out of 13 modules. The frequency and prevalence of core competencies is summarized in Table 12.

Table 12. Frequency and Prevalence of Core Competencies in Modules 2, 3, 5, 6 and 8 Exercises, Activities and Previous Module "Looking Forward" Exercises in *Quantitative Reasoning* (Fahy and Nabours, 2016)

Frequency and Prevalence of Core Competencies in Modules 2, 3, 5, 6 and 8 Exercises, Activities and Previous Module "Looking Forward" Exercises in *Quantitative Reasoning* (Fahy and Nabours, 2016)		
Competency	**Number of questions**	**Percent of questions**
Interpretation	29	15
Representation	32	17
Calculation	105	55
Analysis/Synthesis	34	18
Assumptions	2	1
Communication	12	6

Common Sense Mathematics (Bolker and Mast, 2016) is the course text for MAT 1313. This text also follows a mathematical content progression, however, the text is designed around a context-oriented approach since all of its exercises are grounded in non-hypothetical, real-world contexts. The PI identified contextual themes in chapter exercises: personal finance, extrapolating data, fact checking, conservation (subtopic: hybrid vehicles), rates in the media, argument in writing, evaluating arguments, understanding large numbers in context, and facts and figures in public policy (subtopics: taxes, unemployment). The PI also applied the QLAR rubric to 166 questions from a sample of 3 out of 13 chapters. The frequency and prevalence of core competencies is summarized in Table 13.

Table 13. Frequency and Prevalence of Core Competencies in Chapters 5, 8 and 12 Exercises from *Common Sense Mathematics* (Bolker and Mast, 2016)

Frequency and Prevalence of Core Competencies in the Chapters 5, 8 and 12 Exercises from *Common Sense Mathematics* (Bolker and Mast, 2016)		
Competency	**Number of questions**	**Percent of questions**
Interpretation	51	31
Representation	39	23
Calculation	79	48
Analysis/Synthesis	35	21
Assumptions	5	3
Communication	18	11

Case Studies for Quantitative Reasoning: A Casebook of Media Articles (3rd ed.) (Boersma et al., 2012) is the course text for MATH 2183. This text loosely follows a mathematical content progression, but is designed around a media-out approach with nearly all of its material centered on case studies of media articles. The PI identified some possible contextual themes that can be developed through case study assignments throughout the course: understanding large numbers, units and conversions, extrapolating data, understanding and expressing rates/percentages in writing, expressing rates/percentages in writing, facts and figures in public policy (subtopic: taxes), critical summary, argument in writing, fact checking and HoM (subtopic: representing data). Boersma et al. (2011) applied the QLAR to all study questions in the second edition of *Case Studies for Quantitative Reasoning: A Casebook of Media Articles* (Dingman & Madison, 2008). The frequency and prevalence of core competencies is summarized in Table 14.

Table 14. Frequency and Prevalence of Core Competencies in the QRCW Casebook (Boersma et al., 2011, p. 9)

Frequency and Prevalence of Core Competencies in the QRCW Casebook				
Competency	**Number of questions**	**Percent of questions**	**Number of case studies (*N*=24) which have at least one question mapped to competency**	**Number of case studies which have 50% or more questions mapped to competency**
Interpretation	152	65	24	18
Representation	68	29	19	6
Calculation	101	43	21	9
Analysis/Synthesis	79	34	20	7
Assumptions	14	6	10	0
Communication	53	23	18	4

The QLAR analysis for each of these texts shows increasing complexity with respect to core competency representation from *Quantitative Reasoning* (Fahy and Nabours, 2016), *Common Sense Mathematics* (Bolker and Mast, 2016), and *Case Studies for Quantitative Reasoning: A Casebook of Media Articles* (3rd ed.) (Boersma et al., 2012) respectively. Further, the MATH 2183 text and the MAT 114 text contain the greatest and least contextual complexity respectively. These article analyses reflect the conceptualizations of QR under which each course was designed. MAT 114 was designed under a mathematical conceptualization of QR and both MATH 1313 and MATH 2183 were designed under an integrative conceptualization of QR.

STUDENT INTERVIEWS – ANALYSES TOWARD THE FIRST GOAL

The first portion interview analysis uncovered Adam's positive shift in HoM and the third portion analysis supports the conclusion that this resulted from Adam's experience fact checking media articles in MATH 2183. Further, the third portion analysis uncovered a strong positive shift in Adam's disposition toward mathematics as a result of completing MATH 1313 and MATH 2183. One trend among interviewees uncovered in the first portion analysis was that follow-up prompting was more successful among MAT 114 students than in MATH 1313 students. This supports the conclusion that MATH 1313 better prepared students to respond to the PLHOM – this study's primary measure for HoM in QR students.

In the second portion interview analysis 100% of interviewees identified the MATH 1313/2183 course-like problem as incorporating reasoning they would expect to use outside the QR classroom, and 29% of students identified the MAT 114 course-like problem as incorporating reasoning they would expect to use outside of the QR classroom. Further, three students identified the MATH 1313/2183 problem as *real-world* before asked about it specifically and four students identified the MAT 114 course-like problem as incorporating reasoning they would not expect to see outside of a QR classroom. Interview subjects described the MATH 1313/2183 course-like problem as *open-ended, conceptual, creative,* interpretive and requiring *deep, analytical thinking.* Interview subjects described the MAT 114 course-like problem as *mathy, textbook, clinical* and requiring *critical thinking.* This shows that student views on course-like problems align with the results of the article analysis.

The third portion interview analysis showed that students view QR courses as being different than traditional mathematics courses. Further, 80% of students attributed this difference primarily to the course's applicability to the real world. All interview subjects were able to identify specific topics from the courses that they found relevant to life outside of the QR classroom. Financial mathematics topics were the only topics students could describe in any detail when explaining MAT 114's applicability to real life, and one student expressed that even the financial mathematics topics were not presented in a way that she felt was relevant to her life. MATH 1313 students produced a long list of specific topics they found relevant to life outside the QR classroom: interest rates, gas/utility usage, evaluating news media, taxes, home taxes, budgeting, home expenses, estimation and specific conservation topics. Further, a student who was unhappy with his MATH 1313 instructor was still adamant about the course's *genuine* applicability to real life. Third portion analysis uncovered a case in MAT 114 where grading – as a likely result of course design – negatively impacted a student's disposition toward the course, and also uncovered a case in MATH 1313 where grading – as a likely result of course design – positively impacted a student's disposition toward the course. The students who used textbooks with greater contextual complexity were also better able to describe topics of study they found relevant to life outside the QR classroom.

STUDENT INTERVIEWS – ANALYSES TOWARD THE SECOND GOAL

First portion interview analysis identified experiences students drew on when responding to the PLHOM. Students drew on a combination of situations connected to a HoM: reading habits, major area of study, and experiences in QR courses. Art, the student with the most successful responses to the PLHOM, identified that he regularly reads news articles and claimed his experiences working with graphs from MATH 1313 aided his responses to the PLHOM. Nina and Chayla drew on knowledge gained through their major areas of study; this context-specific knowledge added detail to their responses, but also possibly biased their responses in a way that negatively impacted their scores on the PLHOM. Overall, students more successful in their responses to the PLHOM identified mainly personal interests and major areas of study from which they drew while responding to the PLHOM. One student showed a tendency to give hypothetical real-world explanations when interpreting information from his PLHOM article.

First portion analysis identified several obstacles students faced in applying a HoM. Half of the interview subjects showed a misunderstanding in the authors' points, and this negatively affected their ability to identify strengths and weaknesses in the authors' argument. Every interview subject who received the incarceration article showed difficulty interpreting the graph of percent decreases, and all but one showed difficulty interpreting a graph with an ambiguous measure. Cory showed difficulty incorporating MAT 114 topics into his analysis. Students who scored higher on the PLHOM also showed a more critical eye toward evaluating arguments. In several cases students showed difficulty communicating their understanding. These results suggest students fall back onto experiences situated in contexts important or meaningful to their lives when applying a HoM. The results also show that students have difficulty applying a HoM – especially when interpreting or communicating quantitative information – when either their context-specific experiences lack supporting QR or their content-specific experiences (QR) lack supporting context.

ONLINE PLHOM RESPONSES – ANALYSES TOWARD THE FIRST AND SECOND GOALS

Quantitative analysis of the online PLHOM scores illustrated two main points: overall PLHOM scores were very low among all QR students and MAT 114 had no effect on PLHOM measurable levels of HoM. Qualitative analysis of online PLHOM responses – discussed in the first portion interview analysis – further identifies interpretation of authors' main points as a significant obstacle in student application of a HoM. Forty-five percent of the responses indicated a misunderstanding of the authors' main points in cases where it was clear enough to determine whether or not a response showed a misunderstanding of the authors' main points.

DISCUSSION

Similarities between opportunities to develop a HoM in the three QR courses are shown in part through core competencies required by exercises from all three course texts. The distribution of core competencies required by each text points to differences in HoM opportunities, and through this the PI expects students encounter increasing opportunity to develop a HoM in MAT 114, MATH 1313 and MATH 2183 respectively. Another similarity in opportunity to develop a HoM is found in the three courses' use of real-world context and applicability in course exercises. The way that context is applied in the three courses varies according to course conceptualizations of QR. MAT 114 follows a mathematical conceptualization of QR, and both MATH 1313 and MATH 2183 follow an integrative conceptualization of QR. This is reflected in the content-out (MAT 114) and context-out (MATH 1313/2183) approaches described in the article analyses. Student interview data support the existence of differences in student opportunity to develop a HoM based on the amount/authenticity of context within curricular materials. Further research is necessary in order to quantify these impacts.

HoM demonstrated by students varied. The greatest similarity is found in the generally low levels of HoM exhibited by students. Students tended to draw on experiences from their major areas of study or direct experience consuming and evaluating media articles in their application of HoM. Further, students seemed to encounter similar obstacles in communicating and/or applying HoM; interpreting a written argument and graphical information (especially when involving percent decreases) were especially problematic for students. Differences in HoM demonstrated by students includes the level of skepticism shown when evaluating written arguments, the extent to which experience in their major area of study aided their application of HoM and the life-relevant topics from QR courses they were able to identify. Further research is necessary in order to identify solutions to overcoming these obstacles and ways to encourage application of HoM.

LIMITATIONS

A number of important limitations to this research exists. Sample sizes from the MATH 1313 and MATH 2183 cohorts for the online responses to the PLHOM – 1 and 13 total responses respectively – render any quantitative analysis involving those data useless. Further, sample sizes from the MATH 1313 and MATH 2183 cohorts for the in-person interviews – 5 and 1 respectively – should raise serious concern with the validity of any conclusions drawn about the effects of those courses. In addition, qualitative analyses inherently come with their own set of limitations: human recollection is notoriously faulty and influenced by personal bias and opinion, the perception of the researcher is similarly influenced by his or her personal bias and opinion, and personalities that conflict or agree with the researcher's personality can affect the outcomes of interviews.

In the QLAR article analysis and PLHOM scores there exist reliability issues. For the PLHOM scores, the PI used a process similar to grading coursework from the perspective of an instructor and took notes on trends in scoring and special cases in order to be as reliable as possible. However, it is possible that discrepancies in grading responses to the PLHOM went unidentified. For the QLAR article analyses the PI attempted to follow the rubric close to the way it was applied by Boersma et al. (2011), however, it is possible that discrepancies exist in how the QLAR was applied by the PI when compared with how it was applied by Boersma et al. (2011). Further, the QLAR rubric was only applied to sample portions of *Quantitative Reasoning* (Fahy and Nabours, 2016) and *Common Sense Mathematics* (Bolker and Mast, 2016), but these results were compared with data from Boersma et al. (2011) in which the QLAR rubric was applied to the entirety of *Case Studies for Quantitative Reasoning: A Casebook of Media Articles* (Custom ed.) (Dingman & Madison, 2008).

This research conflates HoM with PLHOM scores. Several of the analyses contained suggest the PLHOM misses aspects of a HoM due to phrasing issues and due to influences from the content of the article used. In some cases the PLHOM may overestimate HoM due to the content of the article used. The need for different articles in the application of the PLHOM introduces a variable that is difficult to control; this difficulty is supported by the differences observed in student ability to understand graphs between articles. One way to account for this in future research would be for researchers to write the articles themselves; although, this conflicts with the ideological standpoint that HoM should be measured against authentic, real-world media.

FUTURE RESEARCH

Through this research the PI identified several future avenues for follow-up research. The PI made several recommendations for changes to the PLHOM based on interview analyses. Possibly the most obvious continued research would be quantifying the effects in PLHOM scores based on recommended changes to the PLHOM. In fact, this could be done without a large cohort of QR students. Any suitably large and random set of responses to both the current version of the PLHOM and an augmented version should meet this goal, and the use of the same article for both would add validity to the results.

The effect of article content bias on PLHOM scores is another suitable avenue for continued research. This could be investigated through testing with several versions of the PLHOM that utilize articles chosen to elicit bias. For example, articles that involve common political or social issues would be good choices. Additional questions aimed at gathering data on respondents' political affiliation or opinion of article content would allow the researcher to quantify impacts on PLHOM scores due to content bias.

During student interviews several subjects described QR topics they found relevant to life outside the QR classroom as topics that made *sense*. The relationship between sense-making and topics that students find applicable to their lives interests the PI. A relevant investigation could be conducted through a more extensive qualitative analysis – with higher sample sizes – similar to the analyses from this research. Also, a suitable instrument could be developed for a quantitative study.

Future research that quantifies the effects of MATH 1313, MATH 2183, and both MATH 1313 and MATH 2183 on HoM would involve quantitative methods similar to those described in this research (pre/post testing with the PLHOM). However, additional measures would need to be taken in order to ensure a suitable sample size. Further, a longitudinal study could better quantify any lasting effects on HoM due to experience in these courses. The PI also identified several obstacles faced by students in applying a HoM. MATH 2183 involves extensive coursework in interpreting and evaluating written arguments in media articles. In theory, MATH 2183 should help students interpret and communicate an article's main point(s). Any effects

MATH 2183 has on students' ability in this regard could be investigated through pre/post testing with the PLHOM as well.

Further, there are case studies in *Case Studies for Quantitative Reasoning: A Casebook of Media Articles* (Boersma et al., 2012) that involve graphs of percent changes. So again, in theory, this course should help remediate student misconceptions in interpreting graphs of this type. A follow-up study that investigates the effect MATH 2183 has on student ability to interpret graphs of this type could point to suitable interventions that help remediate student misconceptions in interpreting graphs of this type. The PI believes that more extensive practice with interpreting graphs of percent changes – situated in a familiar context and in the presence of a helpful mentor – could help remediate student misconceptions involving graphs of percent changes. These kinds of interventions warrant further study.

LIST OF REFERENCES

Association of American Colleges and Universities. (2009). *Quantitative Literacy VALUE Rubric*. Retrieved from https://www.aacu.org/value/rubrics/quantitative-literacy.

Best, J. (2008). Beyond calculation: Quantitative literacy and critical thinking about public issues. In Madison, B. & Steen, L. A. (Ed.), *Calculation vs. context: Quantitative literacy and its implications for teacher education* (pp. 125-135). Washington, DC: Mathematical Association of America (Inc.).

Boersma, S.; Diefenderfer, C.; Dingman, S.; & Madison, B. (2011). "Quantitative reasoning in the contemporary world, 3: Assessing student learning", *Numeracy*: 4(2), Article 8. DOI: http://dx.doi.org/10.5038/1936-4660.4.2.8

Boersma, S., Diefenderfer, C., Dingman, S. & Madison, B. (2012). *Case studies for quantitative reasoning: A casebook of media articles* (3rd ed.). Boston, MA: Pearson Learning Solutions.

Boersma, S. & Klyve, D. (2013). "Measuring habits of mind: Toward a prompt-less instrument for assessing quantitative literacy", *Numeracy*: 6(1), Article 6. DOI: http://dx.doi.org/10.5038/1936-4660.6.1.6

Bolker, E. & Mast, M. (2016). *Common sense mathematics.* Washington, DC: Mathematical Association of America (Inc.).

Bressoud, D. (2004). Forward. In Steen, L. A. (Ed.), *Achieving quantitative literacy: An urgent challenge for higher education* (p. ix). Washington, DC: Mathematical Association of America (Inc.).

Cohen, P. (2001). The emergence of numeracy. In Steen, L. A. (Ed.), *Mathematics and democracy: The case for quantitative literacy* (pp. 23-29). Princeton, NJ: National Council on Education and the Disciplines.

Cole, R. (2003). Look for QL in major organizing questions, not in techniques. In Madison, B. & Steen, L. A. (Eds.), *Quantitative literacy: Why numeracy matters for schools and colleges* (pp. 252-253). Princeton, NJ: National Council on Education and the Disciplines.

Dingman, S. & Madison, B. (2008). *Case studies for quantitative reasoning: A casebook of media articles* (Custom ed.). Boston, MA: Pearson Custom Publishing.

Dingman, S. & Madison, B. (2010A). "Quantitative reasoning in the contemporary world, 1: The course and its challenges", *Numeracy*: 3(2), Article 4. DOI: http://dx.doi.org/10.5038/1936-4660.3.2.4

Dingman, S. & Madison, B. (2010B). "Quantitative reasoning in the contemporary world, 2: Focus questions for the numeracy community", *Numeracy*: 3(2), Article 5. DOI: http://dx.doi.org/10.5038/1936-4660.3.2.5

Fahy, M. & Nabours, G. (2016). *Quantitative reasoning*. Duboque, IA: Great River Learning.

Gal, I. (1997). Numeracy: Imperatives of a forgotten goal. In Steen, L. A. (Ed.), *Why numbers count: Quantitative literacy for tomorrow's America* (pp. 36-44). New York, NY: College Entrance Examination Board.

Grawe, N.; Lutsky, N.; & Tassava, C. (2010). "A rubric for assessing quantitative reasoning in written arguments", *Numeracy*: 3(1), Article 3. DOI:http://dx.doi.org/10.5038/1936-4660.3.1.3

Hallett, D. (2003). The role of mathematics courses in the development of quantitative literacy. In Madison, B. & Steen, L. A. (Eds.), *Quantitative literacy: Why numeracy matters for schools and colleges* (pp. 91-98). Princeton, NJ: National Council on Education and the Disciplines.

Halper, D. (2012, August 12). Over 100 million now receiving federal welfare. Retrieved from https://www.weeklystandard.com/daniel-halper/over-100-million-now-receiving-federal-welfare

Kennedy, D. (2001). The emperor's vanishing clothes. In Steen, L. A. (Ed.), *Mathematics and democracy: The case for quantitative literacy* (pp. 55-59). Princeton, NJ: National Council on Education and the Disciplines.

Leonhardt, D. (2007, January 17). What $1.2 trillion can buy. Retrieved from https://www.nytimes.com/2007/01/17/business/17leonhardt.html

Lopez, G. & Lynn, D. (2016, January 19). There's about half as much violent crime in the US as there was 25 years ago. Retrieved from https://www.vox.com/cards/crime-rate-drop/crime-rate

Lutsky, N. (2008). Arguing with numbers: Teaching quantitative reasoning through argument and writing. In Madison, B. & Steen, L. A. (Ed.), *Calculation vs. context: Quantitative literacy and its implications for teacher education* (pp. 59-74). Washington, DC: Mathematical Association of America (Inc.).

Madison, B. (2003). Articulation and quantitative literacy: A view from inside mathematics. In Madison, B. & Steen, L. A. (Eds.), *Quantitative literacy: Why numeracy matters for schools and colleges* (pp. 153-164). Princeton, NJ: National Council on Education and the Disciplines.

Madison, B. (2008). Planning a conversation about quantitative literacy and teacher education. In Madison, B. & Steen, L. A. (Eds.), *Calculation vs. context: Quantitative literacy and its implications for teacher education* (pp. 3-10). Washington, DC: Mathematical Association of America (Inc.).

Madison, B. & Steen, L. A. (Eds.). (2003). *Quantitative literacy: Why numeracy matters for schools and colleges.* Princeton, NJ: National Council on Education and the Disciplines.

Madison, B. & Steen, L. A. (Ed.). (2008). *Calculation vs. context: Quantitative literacy and its implications for teacher education.* Washington, DC: Mathematical Association of America (Inc.).

Maguire, T & O'Donoghue, J. (2002). Numeracy concept sophistication – An organizing framework: A useful thinking tool. In Maasz, J. & Schloeglmann, W. (Eds.), *Learning mathematics to live and work in our world: Proceedings of the 10th international conference on adults learning mathematics* (pp. 154-161). Ströbl, Austria: Adults Learning Mathematics.

Mathematical Association of America. (1994). *Quantitative reasoning for college graduates: A complement to the standards*. Retrieved from https://www.maa.org/programs/faculty-and-departments/curriculum-department-guidelines-recommendations/quantitative-literacy/quantitative-reasoning-college-graduates

Northern Arizona University. (1991). *1991-1993 Northern Arizona University undergraduate catalog.* Flagstaff, AZ: Northern Arizona University.

Northern Arizona University. (1999). *1999-2001 Northern Arizona University undergraduate catalog*: 99(4). Flagstaff, AZ: Northern Arizona University.

Northern Arizona University. (2018). *2017-2018 Northern Arizona University catalog.* Retrieved from https://nau.edu/cefns/natsci/math/lumberjack-mathematics-center/mat-114/

O'Donoghue, J. (2002). Numeracy and mathematics. In the Irish Mathematical Society bulletin 48 (pp. 47-55). Retrieved from https://www.maths.tcd.ie/pub/ims/bull48/index.html

Orrill, R. (2001). Preface. In Steen, L. A. (Ed.), *Mathematics and democracy: The case for quantitative literacy* (pp. xiii-xx). Princeton, NJ: National Council on Education and the Disciplines.

Pollak, H. O. (1997). Solving problems in the real world. In Steen, L. A. (Ed.), *Why numbers count: Quantitative literacy for tomorrow's America* (pp. 91-105). New York, NY: College Entrance Examination Board.

Packer, A. (2003). What mathematics should "everyone" know and be able to do? In Madison, B. & Steen, L. A. (Eds.), *Quantitative literacy: Why numeracy matters for schools and colleges* (pp. 33-42). Princeton, NJ: National Council on Education and the Disciplines.

Schoenfeld, A. (2001). Reflections on an impoverished education. In Steen, L. A. (Ed.), *Mathematics and democracy: The case for quantitative literacy* (pp. 49-54). Princeton, NJ: National Council on Education and the Disciplines.

Schield, M. (2008). Quantitative literacy and school mathematics: Percentages and fractions. In Madison, B. & Steen, L. A. (Eds.), *Calculation vs. context: Quantitative literacy and its implications for teacher education* (pp. 87-107). Washington, DC: Mathematical Association of America (Inc.).

Shavelson, R. J. (2008). Reflections on quantitative reasoning: An assessment perspective. In Madison, B. & Steen, L. A. (Eds.), *Calculation vs. context: Quantitative literacy and its implications for teacher education* (pp. 27-44). Washington, DC: Mathematical Association of America (Inc.).

Spradley, J. P. (1980). *Participant observation.* New York, NY: Holt, Rinehart and Winston.

Steen, L. A. (Ed.). (1997). *Why numbers count: Quantitative literacy for tomorrow's America.* New York, NY: College Entrance Examination Board.

Steen, L. A. (Ed.). (2001A). *Mathematics and democracy: The case for quantitative literacy*. Princeton, NJ: National Council on Education and the Disciplines.

Steen, L. A. (2001B). The case for quantitative literacy. In Steen, L. A. (Ed.), *Mathematics and democracy: The case for quantitative literacy* (pp. 1-22). Princeton, NJ: National Council on Education and the Disciplines.

Steen, L. A. (2004A). *Achieving quantitative literacy: An urgent challenge for higher education.* Washington DC: The Mathematical Association of America (Inc.).

Steen, L. A. (2004B). Preface. In Steen, L. A. (Ed.), *Achieving quantitative literacy: An urgent challenge for higher education* (pp. xi-xiii). Washington, DC: The Mathematical Association of America (Inc.).

Steen, L. A. (2004C). The Third R: Quantitative Literacy. In Steen, L. A. (Ed.), *Achieving quantitative literacy: An urgent challenge for higher education* (pp. 3-11). Washington, DC: The Mathematical Association of America (Inc.).

Steen, L. A. (2004D). Quantitative literacy in higher education. In Steen, L. A. (Ed.), *Achieving quantitative literacy: An urgent challenge for higher education* (pp. 13-19). Washington, DC: The Mathematical Association of America (Inc.).

Steen, L. A. (2004E). Educational goals for quantitative literacy. In Steen, L. A. (Ed.), *Achieving quantitative literacy: An urgent challenge for higher education* (pp. 21-26). Washington, DC: The Mathematical Association of America (Inc.).

Steen, L. A. (2008). Reflections on wingspread workshop. In Madison, B. & Steen, L. A. (Eds.), *Calculation vs. context: Quantitative literacy and its implications for teacher education* (pp. 11-23). Washington, DC: The Mathematical Association of America (Inc.).

Taylor, C. (2008). Preparing students for the business of the real (and highly quantitative) world. In Madison, B. & Steen, L. A. (Eds.), *Calculation vs. context: Quantitative literacy and its implications for teacher education* (pp. 109-124). Washington, DC: The Mathematical Association of America (Inc.).

Wiggins, G. (2003). "Get Real!" Assessing for Quantitative Literacy. In Madison, B. & Steen, L. A. (Eds.), *Quantitative Literacy: Why numeracy matters for schools and colleges* (pp. 121-143). Princeton, NJ: The National Council on Education and the Disciplines.

Van Groenestijn, M. (2003). Numeracy: A challenge for adult education. In Madison, B. & Steen, L. A. (Eds.), *Quantitative literacy: Why numeracy matters for schools and colleges* (pp. 229- 234). Princeton, NJ: The National Council on Education and the Disciplines.

APPENDIX A – FIRST ARTICLE

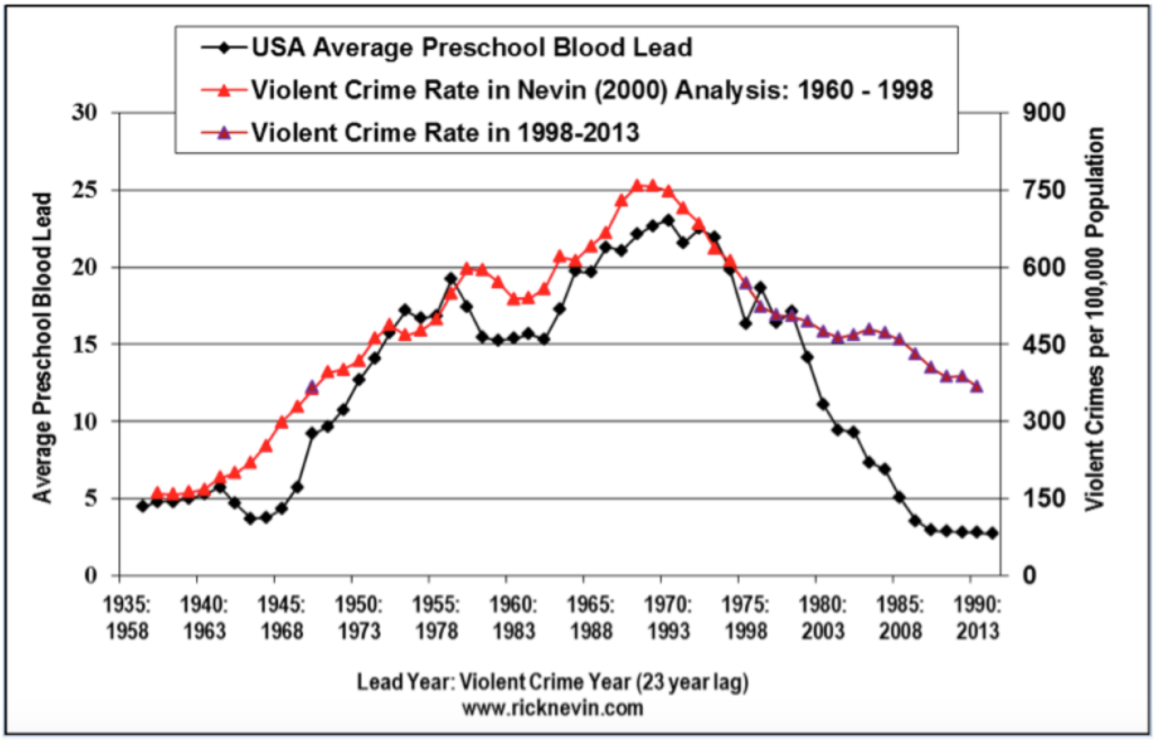
There's about half as much violent crime in the US as there was 25 years ago. Why?

Some theories (like mass incarceration) seemed pretty solid in the 1990s, but have been called into question as more data has come in. Meanwhile, some new theories — like lead getting taken out of gasoline — have gotten popular. But everyone agrees there's no one answer.

The theory: lead exposure caused crime, and lead abatement efforts reduced it

The case for: This is another newly popular theory, in part because of coverage from Kevin Drum at Mother Jones and others. The lead paint ban, removal of leaded gasoline from America's filling stations, and lead abatement efforts — which all decreased lead exposure particularly among children born from around 1975 to the late 1980s — correlates strongly to the cohort of children who hit peak criminal age in the 1990s and 2000s. And the data suggests that these specific cohorts were less likely to get arrested for crime. Given that there's a body of psychological research tying lead exposure to more aggressive behavior, it's likely reduced lead exposure played a role in reducing arrests and crime.

Unlike some of the other theories, evidence for the lead theory also comes from other developed countries, which also experienced a crime drop in the past few decades. "Put all this together and you have an astonishing body of evidence," Drum wrote. "We now have studies at the international level, the national level, the state level, the city level, and even the individual level."



lead and crime

(Rick Nevin)

The case against: The lead theory has the same problem as the abortion theory: In the 1990s, even people who had been exposed to lead as children started committing fewer crimes, albeit not to the same extent as younger cohorts. That indicates that while reduced lead exposure may have been a factor, even a big one, it may not explain the entire crime drop.

One study, published in the Journal of Quantitative Criminology, also found that the correlation between lead and crime seems to be based on faulty data from the FBI's Uniform Crime Report, which collects reports from police departments around the country. The study suggested that when you look at more reliable National Crime Victimization Survey and homicide data, there is little to no correlation between lead exposure and crime. So, at the very least, lead doesn't seem to explain the full crime rise and drop.

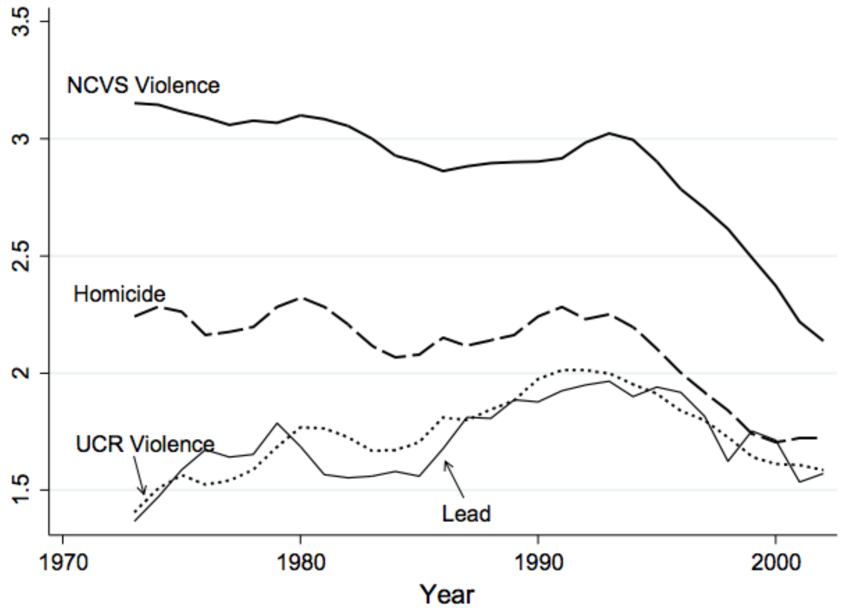


Fig. 3 Logged rates of gasoline lead exposure, homicide, and NCVS and UCR serious violence, 1973–2002. *Note* gasoline lead measure is lagged 23 years and rescaled to improve visual trend comparisons

This chart suggests that UCR data closely correlates with the rise and fall of lead exposure, but homicide and NCVS data do not. (Journal of Quantitative Criminology)

The bottom line: At least some effect. A 2015 Brennan Center analysis, which attempted to quantify the effect of several potential causes of the crime decline, didn't have enough data to quantify lead's impact. But past research makes a good case that it had at least some effect, particularly in the 1990s and going through the 2000s.

Edited by Dara Lynn and German Lopez (www.vox.com)

APPENDIX B – SECOND ARTICLE

There's about half as much violent crime in the US as there was 25 years ago. Why?

Some theories (like mass incarceration) seemed pretty solid in the 1990s, but have been called into question as more data has come in. Meanwhile, some new theories — like lead getting taken out of gasoline — have gotten popular. But everyone agrees there's no one answer.

The theory: putting more people in prison helped reduce crime

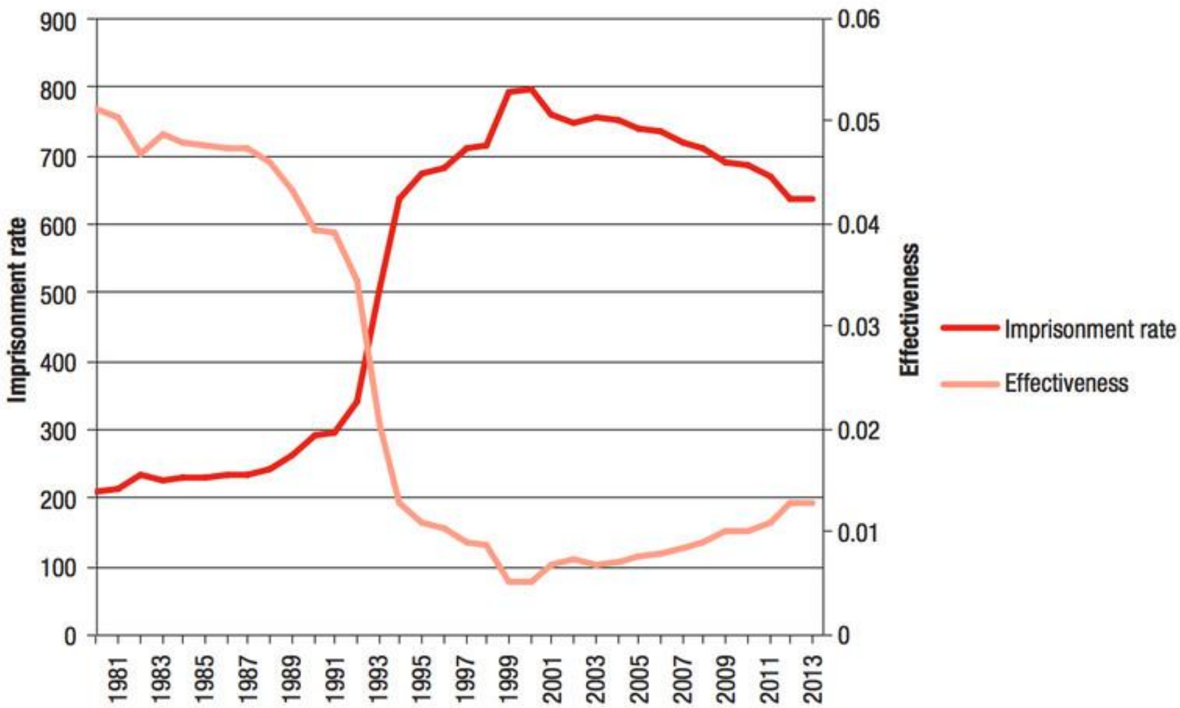
The case for: It seems intuitive. The incarceration rate's been rising; the crime rate's been falling. Surely this is because people are being locked up who'd otherwise be committing crimes out on the streets.

Several academic studies have found that increased incarceration had a big impact on reducing crime. In particular, Steven Levitt (of Freakonomics fame) wrote a paper in 2004 that concluded that 58 percent of the drop in violent crime during the 1990s was due to incarceration.

The case against: These studies were based on older data that only included a few years of the crime decline. Levitt acknowledged he couldn't account for the point of diminishing returns: There are only so many serious criminals out there, and after a certain point the people getting put in prison aren't people who'd be committing crime after crime on the street. The higher the incarceration rate gets, the less it matters if you increase that rate even more. Studies that examine more recent data, after the point of diminishing returns has been hit, find that incarceration wasn't nearly as influential.

"Incarcerating violent people has a big effect on violence," John Roman, senior fellow at the Urban Institute's Justice Policy Center, said. "But most people we incarcerate aren't violent."

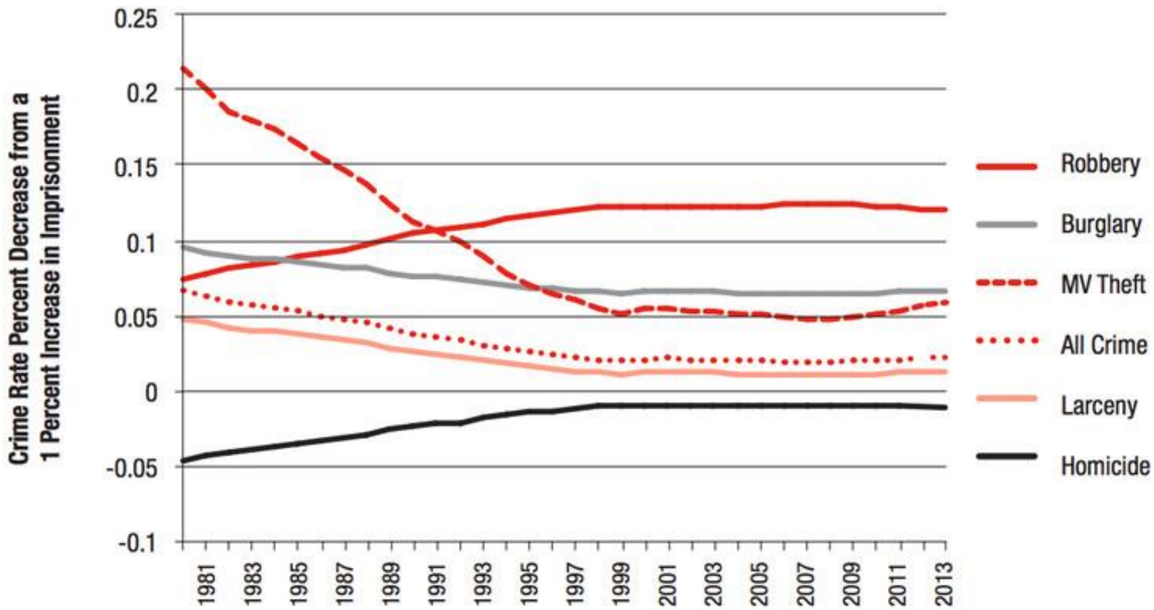
Figure 17: Effect of Increased Imprisonment on Crime in Texas (1980-2013)



(Brennan Center for Justice)

The diminishing returns aren't only about who's being put in prison, but about how long people remain there. The research suggests that people age out of crime, so letting them out of prison 10 or 20 years down the line — instead of the longer sentences applied today — might not pose a threat to public safety. "Crime is a young man's endeavor," Brian Elderbroom, senior fellow at the Urban Institute's Justice Policy Center, said in December. "It's not surprising that someone who commits a crime at a young age would be a completely different person by the time they're in their 30s."

The other problem with this theory is that incarceration rates were increasing for years before crime started going down.



(Brennan Center for Justice)

The bottom line: Some effect. Criminologists now tend to believe that incarceration accounts for a fraction of the drop in crime (say, 25 percent), but no more. A 2015 Brennan Center for Justice report estimates that incarceration played even less of a role than that, especially when it came to violent crime. The Brennan Center concluded that the rising incarceration rates through the 1980s had already locked up the truly violent criminals, and the point of diminishing returns was hit even before the crime rate started to fall.

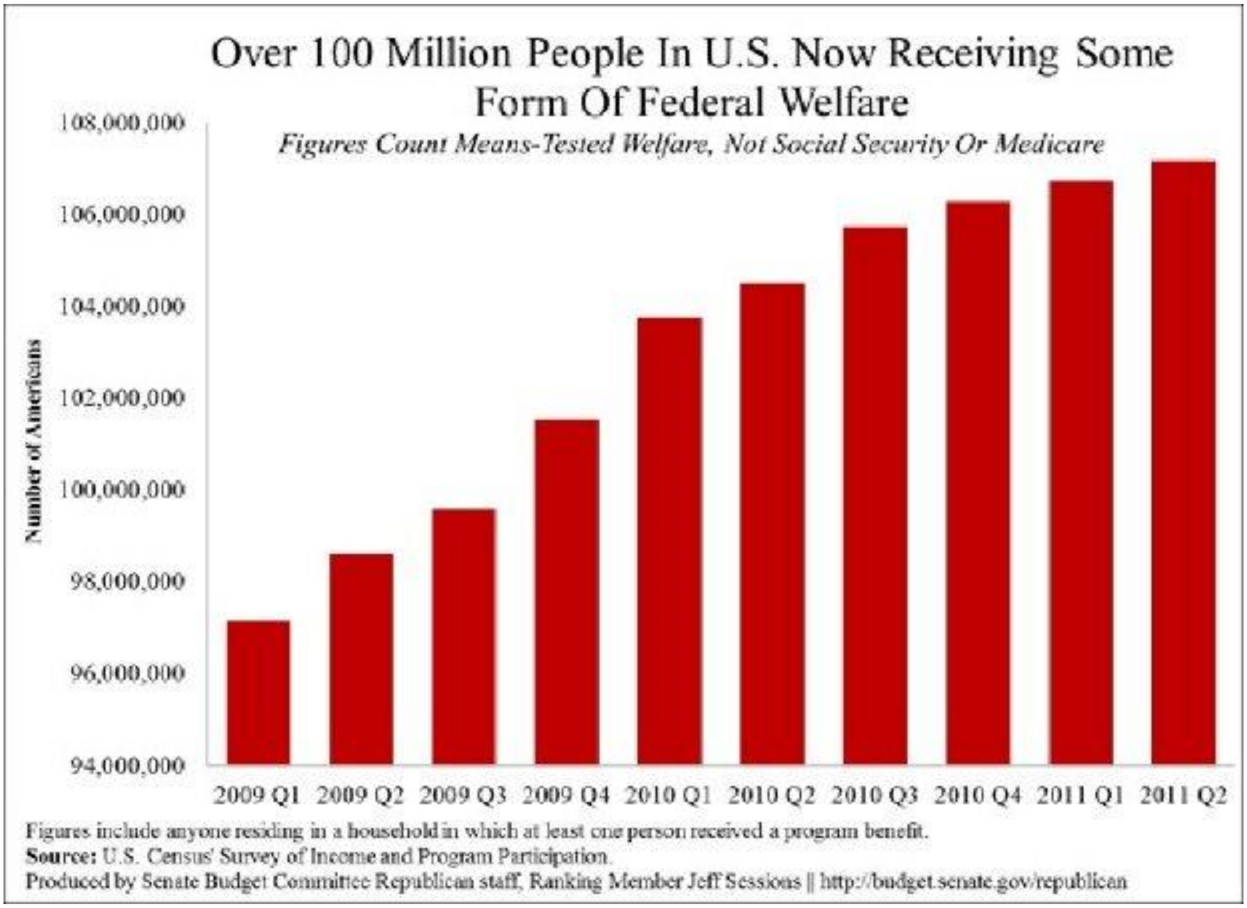
Edited by Dara Lynn and German Lopez (www.vox.com)

APPENDIX C – THIRD ARTICLE

Over 100 Million Now Receiving Federal Welfare

1:40 PM, AUG 08, 2012 | By DANIEL HALPER

A new chart set to be released later today by the Republican side of the Senate Budget Committee details a startling statistic: "Over 100 Million People in U.S. Now Receiving Some Form Of Federal Welfare."



"The federal government administers nearly 80 different overlapping federal means-tested welfare programs," the Senate Budget Committee notes. However, the committee states, the figures used in the chart do not include those who are only benefiting from Social Security and/or Medicare.

Food stamps and Medicaid make up a large--and growing--chunk of the more than 100 million recipients. "Among the major means tested welfare programs, since 2000 Medicaid has increased from 34 million people to 54 million in 2011 and the Supplemental Nutrition Assistance Program (SNAP, or food stamps) from 17 million to 45 million in 2011," says the Senate Budget Committee. "Spending on food stamps alone is projected to reach $800 billion over the next decade."

The data come "from the U.S. Census’s Survey of Income and Program Participation shows that over 110,000 million individuals received a welfare benefit in 2011. (These figures do not include other means-tested benefits such as the Earned Income Tax Credit or the health insurance premium subsidies included in the President’s health care law. CBO estimates that the premium subsidies, scheduled to begin in 2014, will cover at least 25 million individuals by the end of the decade.)"

This is not just Americans, however. "These figures include not only citizens, but non-citizens as well," according to the committee.

APPENDIX D – COURSE-LIKE MATERIALS

Harps Food Stores, a grocer with stores in Arkansas, Missouri, and Oklahoma, used the slogan in the poster to indicate that the chicken they marketed had less sodium than that marketed by Harps' competitors.

1. The poster states, "Our chicken contains up to 5 times less sodium [than the chicken of our competitor]." Explain carefully what this might mean.

2. Using your reasoning from Question 1, if Harps' competitor's chicken has 100 mg of sodium per serving, how much sodium does Harps' chicken have, assuming the ad is correct?

3. Explain what Harps probably meant by the phrase "five times less" and restate the ad slogan with this interpretation. Discuss whether or not you think the new slogan is as effective as the one in the original ad.

Figure 1. Course-like materials from interview protocol (1)

A contractor has the opportunity to purchase a lot of 1,000 used bricks at an auction, some of which are cracked and therefore unusable. The contractor wants to estimate what percentage of the bricks are cracked, but can't inspect all 1,000, so he instead looks through the 40 bricks on top and finds 7 which are cracked.

Match each vocabulary word below with its value or description on the right:

Population	a) 7/1,000
Sample	b) 40
Parameter	c) 7/40
Statistic	d) Entire lot of bricks
Sample Size	e) Percentage of entire lot which are cracked
	f) Bricks which were inspected
	g) Range of values in which 95% of values fall
	h) 7
	i) 33

Figure 2. Course-like materials from interview protocol (2)

APPENDIX E – INTERVIEW PROTOCOL

Interviewee (Name):

Course and Instructor:

INTRODUCTION:

Thank you for taking the time to complete this interview. This interview intends to shed light on how quantitative reasoning students respond to quantitative information as well as how coursework in quantitative reasoning might influence responses to quantitative information. I, David Deville, am the primary researcher and I will maintain your anonymity. Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or as required by law or University policy. Confidentiality will be maintained. Audio recordings will be transcribed without identifiers and original recordings will be destroyed upon completion of the research project.

You were selected based on your responses to the survey you completed at the beginning of this semester and your willingness to participate in this research study. Upon completion of this interview, you will receive a $20 gift card as a token of appreciation. Your cooperation in this study relies on your honest responses to the interview questions. This study does not intend to evaluate your personal abilities in mathematics, nor does it intend to evaluate the abilities of your instructor.

Interviewee Background

AGE:

MAJOR AND YEAR IN SCHOOL:

PREVIOUS COURSEWORK IN QUANTITATIVE REASONING: YES/NO

FUTURE COURSEWORK IN QUANTITATIVE REASONING: YES/NO/MAYBE

PREVIOUS STEM COURSES:

PREVIOUS COURSES THAT INVOLVED REASONING WITH NUMBERS:

1. Habit of mind instrument introduction: The same article and questions from the student's pre-test for habit of mind are re-introduced to the student. The student is given several minutes to look over the article and the questions.

    1. Do you have any questions about the article or questions that I have presented to you?

    2. Could you please answer the last 3 questions on the sheet? (This may take several minutes)

2. Habit of mind follow-up: The student's responses to the habit of mind instrument are quickly graded (discretely) and followed with prompts based on the level (0-3) of the responses to each of the 3 questions. The prompts are followed in ascending order with the goal of eliciting a level 2 response. Students are asked level 3 prompts only if they provide a level 3 response.

    Question: "What facts did the author use to support the main point(s)?"

    Level/Follow-up prompt:

    0. Follow-up: Were there any other facts the author uses to support the main point?

    Were there any quantitative facts the author uses to support the main point?

    1. Follow-up: Could you be more specific?

    In the case that the information from the student is incorrect:

    Could you read me the section from the article where the author provides this fact?

Are you sure that supports the author's point?

I think the article is confusing here and what the author really means is ___________________ (provide explanation to the student).

2. Follow-up: Can you identify any other facts the author uses to support his point?

3. Follow-up: Excellent, do you usually tend to pick out quantitative information from articles?

Was there anything particular about the information that caught your attention?

Question: "Were there any particular strengths or weaknesses in how these facts were reported?"

0. (No strength/weakness identified) Could you please do your best to provide one strength and one weakness?

(Strength/weakness identified without support) Could you explain to me why you think this is a strength or weakness?

1. (Reasoning not specific) Could you elaborate on that?

(Reasoning incorrect) Probe for the misunderstanding; e.g. Are you sure that is what that percentage represents?

2. Could you add anything to that?

Probe for support of any unsubstantiated claims.

3. Excellent, are you used to looking for flaws or strengths in how quantities are expressed in written media?

Question: “Does the graph help interpret the numerical information found in the text? Explain your thoughts” –Focus on reconciliation of information in graph with information in article

0. Did you check to see if the information written in the article matches the information presented in the graph?

Can you explain to me how the information in the graph supports the written information in the article or point out any discrepancies between the information in the article and the graph?

1. (no justification) Could you elaborate on why you think that?

(incorrect justification) Probe for misunderstanding; e.g. are you sure that is what is represented by the graph?

2. Did you notice any discrepancies between the graph and information in the article?

3. Excellent, do you usually check for inconsistencies in information presented in articles?

Question: “Does the graph help interpret the numerical information found in the text? Explain your thoughts” – Focus on strengths/weaknesses in how the graph supports the author’s point.

0. Could you explain how the graph supports or fails to support the author’s point?

1. Probe for misunderstanding

2. Could you elaborate on that?

So, I feel like you haven’t completely convinced me of your point. Could you try to be more specific about why you think that?

3. Excellent, do you usually look for graphical information in articles?

3. Course-like materials: Interviewee is presented two entry-level problems typical of coursework from MAT 114 and the courses at the University of Arkansas. The student is instructed to attempt both problems in the order of their choice.

   1. NAU (MAT 114) problem
      a. Attempted 1st/2nd?
      b. Solved correctly?
   2. UofA problem
      a. Attempted 1st/2nd?
      b. Solved correctly?
   3. Follow-up questions:
      a. What made you decide to do __________ problem first?
      b. Which of these did you feel was easier to complete? (Probe for why)
      c. Which of these problems was more like the problems you encounter during your quantitative reasoning course?
         i. Could you describe any differences in how you had to think about each problem?

ii. Do either of these problems incorporate the type of reasoning you expect to use outside of a purely academic setting? (probe for why)

4. General impressions and comments: The interview finishes with three questions about the student's impression of their QR course.

    1. Have you found that you apply any of the techniques/reasoning you use in class to other areas of study or outside of an academic setting?

    2. Could you describe quantitative reasoning in your own words?

    3. Would you say that your QR course is different from a typical mathematics course? If yes, how so, and if no, could you describe what a "typical mathematics course" means to you?

General Comments:

APPENDIX F – APPLYING THE RUBRIC FOR THE PLHOM

Applying the rubric for the prompt-less instrument from Boersma and Klyve (2013) proved difficult in several ways. In order to reliably score responses to the instrument, the PI used the following protocol. Once the pre-test scores for the instrument were gathered and prepared into an excel spreadsheet, the PI did a quick scoring for all of the responses. During the initial scoring, the PI often doubled back to check that responses were scored similarly. Then, during student interviews, the PI scored responses on the fly in order to respond with appropriate prompting. Upon completion of the interviews, the PI went back through the interview transcripts alongside notes and scores on the interview protocol form. In several instances, the PI determined his initial scoring of student responses (during interviews) was faulty. Once the PI felt he had reliably scored written responses obtained from students during the interviews, the PI went through all of the online responses obtained from pre/post-testing. During this final scoring of the pre/post-test responses, the PI again crosschecked responses to promote reliability and noted where any difficulties in scoring arose. Most of the difficulty arose from either determining when a low level response reached a suitable level 1 response or in separating contextually intelligent responses from quantitatively specific responses. Below is the list of questions from the prompt-less instrument.

1. Did you understand the article?
2. What was the main point(s) of the article?
3. What facts did the author use to support the main point(s)?
4. Were there any particular strengths or weaknesses in how these facts were reported?
5. Does the graph help interpret the numerical information found in the text? Explain your thoughts (Boersma and Klyve, 2013, p. 7).

Determining when low-level responses reached a suitable level 1 response was problematic for several reasons. There were 25 responses to the online prompt-less instrument initially scored at the 0-level that were revised to level 1 responses. Below is the list of scoring criteria for 0- and 1-level responses on questions 3 through 5 from Boersma and Klyve (2013):

3. (0) No quantitative information given or alluded to.

   (1) Some relevant quantitative information is identified (or alluded to), but none is correct (or specific enough to be judged correct or incorrect).

4. (0) No strength or weakness identified or, if identified, not supported with quantitative reasoning.

   (1) A strength or weakness is identified but is supported with incorrect quantitative reasoning (or the reasoning is not specific enough to be able to judge correctness).

5a. (0) No indication that the numbers in the article were checked against their representation in the graph.

   (1) Claims, with no justification, or incorrect justification, that the graph does or does not accurately present the numerical information in the article.

5b. (0) No strength or weakness of the graph identified or, if identified, not supported with quantitative reasoning.

   (1) A strength or weakness of the graph is identified but is supported with incorrect quantitative reasoning (or the reasoning is not specific enough to be able to judge correctness) (pp.7-9).

Initially, 8 responses to question 3 that did not refer to specific figures from the article where scored at the 0-level; however, the articles used in application of the instrument did not report many figures, and the PI decided to use certain key words as indicators of quantitative information. For example, *the graphs*, an undeniably weak response to question 3, was originally scored a 0-level response, however, the response alludes to quantitative information (contained in the graphs) and the final decision of the PI was to score this as a level 1 response. The intention of question 3 in the context of the instrument is to identify whether or not a student identifies quantitative evidence in support of the author's argument. The PI understands a level 1 response to this question signals the response identifies quantitative information supporting the author's point. *Based their data off the FBI's Uniform Crime Report* is another example of a response that was originally scored a 0, and later upgraded to a level-1 response; here, the key-word *data* alludes to quantitative information in the article. Other key words that tipped low-level responses from a score of 0 to a score of 1 include: *statistics, rates, charts, correlation, decrease,* and *increase.* Further, several students included information in their response to question 2 that would have improved their score on question 3. Rather than include responses to question 2 in the scoring of question 3, the PI relied on key words to indicate references to quantitative information.

Six responses to question 4, originally scored a 0, were later revised to a score of 1. The PI originally scored these responses at the 0-level because he did not agree that the responses identified a true strength or weakness of the article. However, what constitutes a strength or weakness is influenced by opinion; what one person views as a strength may be viewed as a weakness or neither by another individual. The PI made the decision that any claims made in response to question 4 that could conceivably be supported with quantitative reasoning warrant a score of 1. For example, *Use of graphs was good* was originally scored a 0 and later revised to a score of 1. This response refers to the article correlating pre-school blood lead level with drops in crime rate. The PI considers the graphs used in the article as a weakness, since the data set behind the first graph is never discussed and the second graph has neither labels for crime rate units nor blood lead level units. *A strength being how well the facts correlate with one another* is another example of a response that was later upgraded from level 0 to level 1; since half of the article disputes the correlation discussed in the first half, the PI originally viewed this as neither a strength nor a weakness. However, they later concluded that this claim could be argued with quantitative reasoning, but it would likely be deemed incorrect (another indicator of a level 1 response). The 4 responses scored at the 0-level in 5b that were later revised to level 1 are explained similarly.

Seven responses scored according to the rubric for 5a were revised from a level 0 to a level 1 response. Boersma and Klyve (2013) included this version of the rubric for question 5 to identify whether or not students checked claims made by the author against information presented in a graphical form. This is generally difficult to determine for low-level responses. The PI decided to err on the side of leniency in scoring this question. For example, *yes, they are clear and concise to understand the data* is likely the lowest-level response scored as a 1 for the protocol in 5a. Here, the response refers vaguely to *the data*, however, the same student responds to question 2 with *the graphs helped me understand the data from the author's claims*; it is the impression of the PI that *the data* could refer to claims made by the author in this case. Responses to question 5 scored by the protocol for 5a that made any reference or allusion to how the graph fit within the author's claims were scored as a level 1 response. The PI feels these responses, at the very least, indicate that the student considered the graphs within the context of the argument. Further, scores according to protocol 5a can be delineated from scores according to protocol 5b by the mandate that level 1 responses in 5b include some quantitative reasoning.

Responses originally scored at a level 2 or 3 were generally less problematic, and required few changes upon revision. This was due in part to the low number of responses scored at this level, but also due to the clarity in communication this level of response requires; for example, in all cases a level 2 response includes claims that are correct or at least justified correctly. Below are the scoring criteria for level 2 and 3 responses to questions 3 through 5.

3. (2) Some relevant and correct information is identified, but not all.

   (3) All relevant quantitative information is correctly identified.

4. (2) A strength or weakness is identified and is supported with quantitative reasoning, but the reasoning is incomplete (e.g., it contains unsubstantiated claims).

   (3) A strength or weakness is identified and supported with correct and complete quantitative reasoning.

5a. (2) Claims, with justification, that the graph does or does not accurately present the numerical information found in the text.

   (3) Correctly points out a specific discrepancy between the graphical presentation and the quantitative information found in the text.

5b. (2) A strength or weakness of the graph is identified and is supported with quantitative reasoning, but the reasoning is incomplete (e.g. it contains unsubstantiated claims).

   (3) A strength or weakness of the graph is identified and supported with correct and complete quantitative reasoning. (Boersma and Klyve, 2013, pp. 7-9).

In one case a response, scored at the 1-level (during in-person interviews), was revised to a level 2 response. Similarly, a single level 2 response (during in-person interviews) was later revised to a level 3 response. These responses were upgraded based on review of interview transcripts.

There were 2 responses scored at a level 3 (during in-person interviews) that were later revised to level 2 responses, 4 responses scored as level 2 that were later revised to level 1 responses and a single level 2 response that was later revised to a level 0 response. The confusion in scoring these responses was due to a conflation (in the PI's mind) of contextually intelligent responses with quantitatively specific responses. The clearest example of this is highlighted by the response to question 4 that was dropped from a level 2 to a level 0 response;

> *The researcher didn't explain what the abortion theory is, leaving the reader to assume and try and figure it out for themselves. Personally, I didn't see any reason to include the abortion theory, as the mentioning of it confused me and led the main point to be put off track.*

Initially, the PI scored this as a level 2 response to question 4. The response is well written, and it identifies a valid weakness of the article. However, the response is not supported with quantitative reasoning.

> *They used the facts that the FBI UCR is a rather faulty system of calculating crime increases and decreases (this is typically because police departments may have incorrect paper trails for cases which would alter the way that the UCR determines statistics of crime) as well as facts that the NCVS shows that crime really didn't reduce during the time where lead usage dropped.*

Above is an example of a well-reasoned response to question 3 that draws on this criminology and criminal justice major's apparent knowledge of common crime statistics. This response was later revised to a level 1 response from a level 2 response based on the small, but relevant mistake in interpreting quantitative information from the article. They key error is found in the phrase *during the time where lead usage dropped.* The article discusses correlations between pre-school blood lead levels and crime rates, but with a 23 year lag in order to correlate pre-school blood lead levels with criminal activity during the time where the pre-schoolers hit peak criminal age. The following response to question 4 was later revised from a level 3 response to a level 2 response. *Some of this info seems like speculation. Having actual numbers with citations in the* *article that align with the graphs would make it more reliable.* Here, the student correctly identifies a weakness in the article's use of graphs, however, the reasoning was deemed not specific enough to warrant a level 3 response.

APPENDIX G: IRB APPROVAL LETTER

UNIVERSITY OF ARKANSAS

Office of Research Compliance
Institutional Review Board

January 11, 2017

MEMORANDUM

TO:	David Deville Bernard Madison
FROM:	Ro Windwalker IRB Coordinator
RE:	New Protocol Approval
IRB Protocol #:	16-11-204
Protocol Title:	*Quantitative Reasoning Course Inquiry*
Review Type:	☐ EXEMPT ☒ EXPEDITED ☐ FULL IRB
Approved Project Period:	Start Date: 01/06/2017 Expiration Date: 01/05/2018

Your protocol has been approved by the IRB. Protocols are approved for a maximum period of one year. If you wish to continue the project past the approved project period (see above), you must submit a request, using the form *Continuing Review for IRB Approved Projects*, prior to the expiration date. This form is available from the IRB Coordinator or on the Research Compliance website (https://vpred.uark.edu/units/rscp/index.php). As a courtesy, you will be sent a reminder two months in advance of that date. However, failure to receive a reminder does not negate your obligation to make the request in sufficient time for review and approval. Federal regulations prohibit retroactive approval of continuation. Failure to receive approval to continue the project prior to the expiration date will result in Termination of the protocol approval. The IRB Coordinator can give you guidance on submission times.

This protocol has been approved for 80 participants. If you wish to make *any* modifications in the approved protocol, including enrolling more than this number, you must seek approval *prior to* implementing those changes. All modifications should be requested in writing (email is acceptable) and must provide sufficient detail to assess the impact of the change.

If you have questions or need any assistance from the IRB, please contact me at 109 MLKG Building, 5-2208, or irb@uark.edu.

109 MLKG • 1 University of Arkansas • Fayetteville, AR 72701-1201 • (479) 575-2208 • Fax (479) 575-6527 • Email irb@uark.edu