FROM NO TO YES: THE IMPACT OF AN INTERVENTION ON THE PERSISTENCE OF ALGEBRAIC MISCONCEPTIONS AMONG SECONDARY SCHOOL ALGEBRA STUDENTS

A dissertation presented
by
Susan F. Zielinski
to
The College of Professional Studies
In partial fulfillment of the requirements
for the degree of Doctor of Education
In the field of
Education
Northeastern University
Boston, Massachusetts
December 2017
Abstract

Many students enter high school with persistent algebraic misconceptions that limit their success in mathematics and, by extension, limit potential educational attainment and future earnings.

The purpose of this study was to assess the effectiveness of a warm conceptual change based intervention on remediating algebraic misconceptions held by students at a secondary school. The study used a quasi-experimental, pre-test post-test, control group design as well as student focus groups. In this eight-week classroom intervention, students first took a diagnostic pre-test to identify their misconceptions. For each of seven different misconception categories, students watched a video, took an online practice quiz, and then completed a series of short paper practice slips in class. Students then took a post-test and changes were analyzed. The videos were designed using augmented activation and refutational text to create cognitive conflict. The theorized causal relationship between engagement and conceptual change was explored through analysis of usage statistics, such as the number of times a video was watched. The intervention also explored how interleaved practice and process mnemonics could improve long-term learning retention. A repeated measures ANOVA showed that students in the intervention group made significantly fewer misconception errors than the control group, while regression analysis indicated that deeper engagement led to significantly greater conceptual change. Mann-Whitney U tests showed that the intervention was significantly effective at reducing errors in some misconception categories but not in others. Finally, student focus groups explored how students experienced the intervention. The findings gave strong support for the effectiveness of the intervention at reducing the number of misconception errors students made.

*Keywords:* warm conceptual change, misconceptions, mathematics, refutational text, intervention, engagement, spaced practice, interleaved practice, process mnemonics
Acknowledgements

While in the end we doctoral candidates write alone, it’s the people around us who help us survive and thrive. I’d like to thank the people who inspired, supported, and encouraged me over the three plus years of this doctoral journey.

I would never have finished this project without supportive family and friends. It’s impossible to overestimate the help my husband Michael gave me as both an understanding spouse and an eager participant in the research. His enthusiasm for the project and for teaching high school kids inspires both me and his students. I’m lucky to have him in my life. Fortunately, my grown children didn't miss me that much over the past few years or question me too deeply about why I was going back to school yet again. My friend Stephanie and my sister Chris were always there when I needed to recharge by escaping from SCHOOL in all its forms.

I appreciate the support of my school, my colleagues who participated in the intervention, and the students who lived it. A special shout-out to Dr. Morse, who not only participated in the intervention but also served as my 3rd reader. His expertise and perspective were crucial to the project’s success. This study never would have seen the light of day without such a nurturing environment in which to experiment with best practices in teaching.

I’d also like to thank Dr. Conn (my steadfast adviser), Dr. Unger (2nd reader and genuinely inspiring educator), Dr. Qian (finally, numbers!), and all the faculty who helped me understand the softer side of science and that qualitative research is indeed legitimate. Thanks also to the cohort of students keeping me company online through this journey.

Finally, I’d also like to recognize my father, who passed away while I was writing this dissertation. Though he never went to college himself, he worked hard to make sure his kids did. He inspired me more than he knew.
Table of Contents

Abstract ............................................................................................................................................. 2
Acknowledgements .......................................................................................................................... 3
List of Tables ..................................................................................................................................... 7
List of Figures ................................................................................................................................... 8
List of Abbreviations ....................................................................................................................... 10
Chapter 1: Introduction .................................................................................................................. 11
  Statement of the Problem .............................................................................................................. 11
  Significance of the Problem .......................................................................................................... 14
  Researcher Positionality ............................................................................................................... 18
  Research Questions and Hypotheses ........................................................................................... 23
  Theoretical Framework .................................................................................................................. 27
  Conclusion .................................................................................................................................... 33
Chapter 2: A Review of the Literature .......................................................................................... 34
  Diagnosing Misconceptions ......................................................................................................... 34
  Types of Mathematical Misconceptions ....................................................................................... 35
  Conceptual Change Models ......................................................................................................... 41
  Mathematics Classroom Interventions ....................................................................................... 46
  Spaced and Interleaved Learning ................................................................................................. 51
  Process Mnemonics ..................................................................................................................... 53
  Conclusion .................................................................................................................................... 55
  Implications for this research ...................................................................................................... 55
Chapter 3: Research Design

Research Questions and Hypotheses

Research Design

Population and Sampling

Description of the No-to-Yes Intervention

Data Collection

Procedures

Data Analysis

Validity, Reliability and Generalizability

Protection of Human Subjects

Conclusion

Chapter 4: Research Findings

Research Question 1 (RQ1)

Research Question 2 (RQ2)

Research Question 3 (RQ3)

Research Question 4 (RQ4)

Conclusion

Chapter 5: Discussion of the Research Findings

Revisiting the Problem of Practice

Review of the Methodology

Discussion of Key Findings

Discussion of Findings in Relation to the Theoretical Framework

Discussion of Findings in Relation to the Literature Review
Implications for Current Practice ........................................................................................................ 156
Limitations of the Study ...................................................................................................................... 156
Significance of the Study ..................................................................................................................... 159
Conclusion ........................................................................................................................................... 160
Future Work ........................................................................................................................................ 161
References .......................................................................................................................................... 165

Appendices .......................................................................................................................................... 184
Appendix A: Teacher Recruitment Email ......................................................................................... 184
Appendix B: Teacher Signed Consent Form – Experimental Group .............................................. 185
Appendix C: Focus Group Protocol .................................................................................................. 188
Appendix D: Student Instructions ....................................................................................................... 189
Appendix E: NTY Test Instrument (Pre- and Post- Test) ................................................................. 190
Appendix F: Instructions for administering the No-To-Yes Diagnostic ......................................... 191
Appendix G: Sample of the NTY Test Grading Rubric .................................................................. 192
Appendix H: Screen Shots from Video ............................................................................................. 193
Appendix I: Process Mnemonics in Online Videos ........................................................................ 194
Appendix J: Sample Online Quiz Question and Feedback ............................................................. 195
Appendix K: Examples of In-Class Practice Slips ........................................................................... 196
Appendix L: Sample Wall Chart ........................................................................................................ 197
Appendix M: Assumption Checking for Hypothesis 1 .................................................................... 198
Appendix N: Assumption Checking for Hypothesis 3B .................................................................. 202
List of Tables

Table 1  Sample Misconception Questions and Answers .......................................................... 69
Table 2  Correlation Between Diagnostic Score and Misconception Errors on the Pre-test ... 78
Table 3  NTY Intervention Timeline ......................................................................................... 80
Table 4  Repeated Measures ANOVA for MES by Group ......................................................... 98
Table 5  Pairwise Differences for MES in the Repeated Measures ANOVA ............................ 98
Table 6  Pre-Post Test MES Comparison by MC for Students Included in RQ2 ...................... 103
Table 7  Mann-Whitney U-Test for MC1 .................................................................................. 105
Table 8  Mann-Whitney U-Test for MC2 .................................................................................. 106
Table 9  Mann-Whitney U-Test for MC3 .................................................................................. 108
Table 10 Mann-Whitney U-Test for MC4 ................................................................................. 110
Table 11 Mann-Whitney U-Test for MC5 ............................................................................... 111
Table 12 Mann-Whitney U-Test for MC6 ............................................................................... 113
Table 13 Mann-Whitney U-Test for MC7 ............................................................................... 114
Table 14 Descriptive Statistics for Variables in RQ3 ................................................................. 116
Table 15 Simple Regression Results for RQ3 ......................................................................... 118
Table 16 Results for the Multiple Linear Regression Model Predicting Change in MES from
Pre-Test to Post-Test .................................................................................................................. 120
Table 17 Coefficients for the Multiple Linear Regression Model Predicting Change in MES
from Pre-Test to Post-Test ........................................................................................................ 120
Table 18 Key Findings ................................................................................................................ 142
List of Figures

Figure 1: Dole and Sinatra's (1998) Cognitive Reconstruction of Knowledge Model .................. 28
Figure 2: Distribution of grade levels across the treatment groups ........................................ 64
Figure 3: Mean diagnostic score by class, with bars showing one standard deviation ............... 65
Figure 4: Distribution of scores on the algebra diagnostic test across the treatment groups ...... 65
Figure 5: Distribution of fall term test scores across treatment groups ................................. 66
Figure 6: Example of an embedded Edpuzzle pop-up question ............................................. 70
Figure 7: Practice slips organized into boxes ................................................................. 72
Figure 8: Sample of a graded practice slip ........................................................................ 72
Figure 9: Boxplots of pre-test misconceptions error scores by group ..................................... 95
Figure 10: Boxplots of post-test misconception error scores by group .................................. 96
Figure 11: Mean pre-test and post-test misconception errors scores by group ....................... 96
Figure 12: Profile plot showing estimated marginal mean number of errors by group ............. 97
Figure 13: Distribution of pre-test misconception errors by MC and group ......................... 100
Figure 14: Distribution of post-test misconception errors by MC and group with total percent
decrease from pre-test below each bar ................................................................................ 101
Figure 15: Percent decrease in misconception errors from pre-test to post-test by MC and group.
........................................................................................................................................... 102
Figure 16: Side-by-side histograms for the change in MES by group for MC1 ....................... 104
Figure 17: Side-by-side histograms for the change in MES by group for MC2 ....................... 106
Figure 18: Side-by-side histograms for the change in MES by group for MC3 ....................... 108
Figure 19: Side-by-side histograms for the change in MES by group for MC4 ....................... 109
Figure 20: Side-by-side histograms for the change in MES by group for MC5 ....................... 111
Figure 21: Side-by-side histograms for the change in MES by group for MC6......................... 112
Figure 22: Side-by-side histograms for the change in MES by group for MC7......................... 114
Figure 23: Scatterplot of total engagement vs. change in MES....................................... 117
Figure 24: Scatterplot of predicted vs. residual values of MES and Normal P-P plot ............ 117
List of Abbreviations

AA: Augmented activation, a conceptual change technique designed to activate prior misconceptions and alert students to a possible error in thinking.

ANOVA: Analysis of Variance

CCT: Conceptual Change Theory

CRKM: Cognitive Reconstruction of Knowledge Model, the warm conceptual change model propose by Dole and Sinatra (1998).

MC: Misconception Category

MES: Misconception Error Score, which counts the number of errors made on a student’s pre- or post-test directly attributable to a particular misconception.

MLR: Multiple Linear Regression

NTY: No-to-Yes, the name of the intervention

RQ: Research Question

RT: Refutational text, a conceptual change technique in which a text is designed to explicitly draw attention to a misconception and then refute it.

WCCT: Warm Conceptual Change Theory
Chapter 1: Introduction

Many high school math students make this very common error:

\[
\frac{a + b}{a} = b
\]

For any student with a true understanding of variables, this random canceling of letters gives a clearly false answer. However, despite repeated discussions and practice, algebra students continue to make this same error, in many ways, year after year. This problem appears in an article written in 1988 that lists twenty-two common errors that many math teachers in 2015 would also immediately recognize (Marquis, 1988). Looking even farther back, there are similar complaints dating from the early 1900s (Smith, 1946; The Mathematics Teacher, 1910; Webber, 1929). In the intervening century, little has changed. Students today are making the same errors their parents and grandparents made in high school, and math educators struggle to solve the problem. Merlin (2008) concurred, stating that “the universality of these errors, the striking identicalness of these errors across settings and decades…and the persistence of these errors among students taking higher-level math courses all suggest that affect and motivation alone do not account for them” (p. 101). Overall, “it is tempting to describe high school algebra as it is unveiled in our research as a disaster area” (L. Lee & Wheeler, 1989, p. 53). Since such algebraic misconceptions limit students’ success in math, thereby limiting their success in school and beyond, there is a pressing need to figure out how to help students change their misconceptions.

Statement of the Problem

The topic. As shown by the publication of articles on common errors in algebra over the decades, many students enter high school with persistent algebraic misconceptions that limit their success in mathematics and, by extension, limit potential educational attainment and future
earnings. They consistently make fundamental errors based on these persistent misconceptions, which are defined as deeply rooted, erroneous beliefs. To change these misconceptions, math teachers must acknowledge that traditional methods have not worked and that something different is needed. While the source of this problem is rooted in middle and elementary school math experiences, the focus of this research is intervention in a high school context, specifically focusing on the promise of warm conceptual change theory in remediating these deeply held misconceptions.

**Justification for the research problem.** There are many compelling reasons why it is important to learn math well. First, rigorous study of math develops problem solving skills, logical thinking, persistence, discipline, and other habits of mind that serve students in all areas of their lives. Mathematics is “a very broad and multidimensional subject that requires reasoning, creativity, connection making, and interpretation of methods; it is a set of ideas that illuminate the world; and it is constantly changing” (Boaler, 2016, p. xii).

Further, when students hold misconceptions about fundamental algebraic concepts, their future success in math classes and STEM-related careers is in jeopardy. When students do not reach their full potential in math, they have limited their ability to successfully complete college, obtain fulfilling and high paying jobs, contribute fully to society, and help the U.S. compete globally (Rose & Betts, 2001).

Just taking more math classes is not enough to cure students of their misconceptions. Students in high school have more misconceptions than in middle school (Lucariello, Tine, & Ganley, 2014), the errors building on each other and preventing the proper acquisition of new knowledge. Despite (or perhaps because of) more students taking algebra earlier, data from the National Assessment of Educational Progress (NAEP) show that only 6.9% of 17-year-olds
scored at or above a proficient level in algebra (2008). Further, the U.S. ranked 27 out of 34 countries in an international assessment of mathematics performance (“Programme for International Student Assessment (PISA) results for PISA 2012,” n.d.). When students are passed along to the next math course before they have mastered the concepts, they are set up for failure and increased dislike of mathematics.

This problem is far more pressing today than it was in the early 20th century because there is now an expectation for success in algebra among all students, whether or not they are bound for college or are in advanced math programs (Silver, 1997). However, policies that push for “algebra for all” have had low success rates (Viadero, 2010), in part due to poor mathematical preparation before algebra. Some systems have seen increased dropout rates, especially among Latino and African-American students, as more math is required to graduate but students are not adequately prepared (“Unintended consequences: More high school math, science linked to more dropouts,” 2014). While universal access to 8th grade algebra may help close the achievement gap related to family socio-economic status (Spielhagen, 2006), pushing inadequately prepared students into 8th grade algebra has led to a drop in the level of rigor (Baker, 2013), preparing students inadequately for higher math, instilling misconceptions, and setting them up for future struggles. In the current climate, it is more crucial than ever that math educators determine how to eradicate fundamental misconceptions so that students can achieve success in secondary school mathematics.

**Deficiencies in the evidence.** The body of scholarly research on this topic is lacking in many ways. While there are have been many studies aimed at understanding and diagnosing students’ mathematical misconceptions, there have been very few studies that propose solutions to remediate them. There are few intervention studies set in math class, and fewer still that
directly reference conceptual change models in math settings. Though researchers often employ conceptual change theory in a science context, the theory also applies to the domain of mathematics, where true understanding often requires conceptual change (Vamvakoussi & Vosniadou, 2006).

This research was designed to address these deficiencies. The purpose of this study was to assess the effectiveness of a conceptual change based intervention on remediating algebraic misconceptions held by students at a private secondary school.

**Relating to audiences.** The primary audience for this study is secondary school math educators. By adopting a successful intervention, teachers can help their students overcome their misconceptions and hence be more successful in math and, by extension, life. Middle school math teachers could adopt a modified intervention as a way of preventing students from forming these misconceptions in the first place. Teachers in other subjects may also find the conceptual change work relevant to their own subjects, especially in science. Additionally, school administrators can use this study as justification for implementing the intervention in their schools. Researchers in conceptual change often work in laboratory settings, so they may be interested in this classroom-based intervention. Finally, those interested in closing the equity gap in mathematics may also be interested in how remediating persistent misconceptions can help achieve this goal.

**Significance of the Problem**

Why does learning math well matter? Research shows that students who successfully take math beyond algebra will be better able to reach their full potential to attend college, obtain fulfilling jobs, contribute to society, and help the U.S. compete globally (Rose & Betts, 2001).
**Impact on individuals.** Students often see math as hard, boring, and irrelevant. This attitude, coupled with poor algebra skills, can lead students to take fewer math courses. This, in turn, is related to a number of undesirable consequences.

Students with weak math backgrounds have poorer outcomes related to college. First, these students are less likely to attend college. One study reported that 83% of students who took Algebra 1 and Geometry enrolled in college within two years of their high school graduation, compared with 36% of those who did not take those courses (Department of Education, 1997). Further, students with weak math backgrounds are less likely to be admitted to selective colleges or be offered scholarships, in part due to lower math SAT scores (Dreyfus & Salomon - Fernandez, 2015). Once students are admitted to college, those with weak skills are often required to take remedial math classes. Nearly 75% of students entering a two-year college must take at least one remedial course in math (Bonham & Boylan, 2011). Unfortunately, students who take remedial classes often do not complete them. Nearly 40% of remedial students in community colleges never complete their remedial courses, and of those that do, fewer than 25% go on to complete college level English and math courses (Complete College America, 2012). Other reports are even more dire, estimating that 75% of all remedial math students in community colleges do not finish their remedial courses, and of those who do, 81.5% do not complete a credential or degree (Bahr, 2008). Finally, students with weak math skills have lower rates of college graduation. Students taking Algebra 2 or higher are 18% more likely to attend college and three times more likely to graduate (40% vs. 13%) than students taking at most Algebra 1 (James, 2013).

Students who take fewer math courses also earn less once they finish high school. One extra course in algebra or geometry is associated with 6.3% higher earnings, whether or not the student
attended college (Rose & Betts, 2001). Many fast-growing, high-paying careers, such as those in STEM fields, require a strong math background. In the United Kingdom, girls who have achieved two or more A levels in STEM subjects earn 33% higher salaries at mid-career, with early success in mathematics a key indicator (Morgan, n.d.). For black male students, each additional math course increases annual earnings by 8% (Goodman, 2012).

Finally, young adults who score higher on tests of math ability have lower unemployment rates, even if they did not go to college (Department of Education, 1997). Students with “low levels of math are 50% more likely to be unemployed than those with higher levels of math” (James, 2013, p. 3). Overall, for both high school dropouts and high school graduates, students with more advanced math skills have higher earnings and lower unemployment.

**Impact on schools.** High schools are also impacted by students’ lack of basic algebra skills. Teachers must continually reteach basic skills, which means loss of class time for new material or enrichment. For example, at the researcher’s high school in the fall of 2015, the incoming Algebra 2 class was given a diagnostic test of basic Algebra 1 skills. The median grade was 55% and a quarter of the students scored below 38%. In the fall of 2016, incoming students to non-Honors Algebra 2 had a mean score of 44%. These students simply don’t have the necessary Algebra 1 background to be successful in Algebra 2. As a result, the teachers at this school, like in many high schools, spend a great deal of time attempting to re-teach Algebra 1 skills, which weakens the entire math series leading up to Calculus. It also weakens the math-based science series.

**Impact on families and communities.** Families and communities are also impacted by student success in math. If students are not successful in college, they have limited their earning potential and hence are less able to support their families or contribute to the community. The
more math a person takes, the more he or she earns on average and the more likely he or she is to have a job, which translates into increased ability to support a family (“The surprising impact of high school math on job market outcomes,” n.d.). Further, lack of success in math can lead to dropping out of high school, and communities with higher dropout rates tend to have more crime (“Unintended consequences: More high school math, science linked to more dropouts,” 2014).

Impact on society. The U.S. is currently facing a shortage of STEM graduates in private industry and government, which has economic implications for our global competitiveness in technological fields (Xue & Larson, 2015). Some are calling for an additional one million STEM graduates in the next decade, in part by slashing the dropout rates from STEM college majors who are often underprepared in math (President’s Council of Advisors on Science and Technology, 2012). In 2008, the U.S was ranked 20th in the world in proportion of students earning a four-year degree in engineering or natural sciences. Parents, teachers, and schools must work together to ensure students are aware of the wide variety of STEM fields and have the preparation to succeed.

Overall, the research clearly indicates that success in math is good for both students and the nation. Evan, Gray, and Olchefske (n.d.) sum up the situation accurately:

If we want to dramatically increase the proportion of students graduating from high school with high-level, globally competitive skills, then we must dramatically increase the number of students who achieve proficiency in Algebra in their middle or early high school years as a gateway to the advanced high school coursework that is the driver of high school graduation, college readiness, and post-secondary completion rates (p. 2).
Helping students overcome their “persistent and pernicious” (J. L. Booth, Barbieri, Eyer, & Paré - Blagoev, 2015) algebraic misconceptions will remove a barrier to students’ success in mathematics.

**Extending the scholarly research.** There is a rich body of work that identifies, categorizes, and diagnoses mathematical misconceptions, as well as much research set in science classes that investigate the conceptual change theory used to attack such misconceptions. However, there is a distinct lack of research in applying that theory to mathematics in a practical way. The design and deployment of this intervention designed to remediate algebraic misconceptions adds to the body of scholarly knowledge about changing deeply held conceptions in mathematics.

**Improving practice.** By designing and testing an effective intervention, this work can immediately improve student outcomes at the local level. By sharing the research, this intervention can then be used by other high school teachers to improve the learning of students across the mathematical world. Further, and most excitingly, this work can be shared with middle school math teachers to help them prevent these misconceptions from taking root in the first place. By solving the problem at the source, students will be more likely to enjoy math classes, take more advanced math classes, and experience the many benefits discussed previously. Teachers will then have more time to focus on deeper problem-solving skills, projects, and enrichment, resulting in even better teaching and learning.

**Researcher Positionality**

Because the problem of persistent algebraic misconceptions is frustrating and damaging to both students and teachers, I am passionate about this subject. First, students are stressed and frustrated, since it is painful for them to continually struggle to solve problems while repeatedly
making the same mistakes. They develop a deep dislike of math and an unshakable sense that they are hopelessly bad at it. These struggling students stop taking math as soon as they can, and are unlikely to pursue STEM-related fields. Classes move more slowly and are “dumbed down” as too much time is spent attempting to fill gaps and correct misconceptions rather than covering new material. As a result, about half of Algebra 2 has become a review of Algebra 1, which delays all the math classes that come next.

Even more frustrating, no matter how much time I have spent, I have rarely been successful helping students stop making these particular mistakes; I see my calculus students making the same common errors as Algebra 2 students. However, they can no longer fake their way through. “In stacking concept upon concept, calculus is the subject most likely to tip the balance, reveal the dry rot, and send the whole edifice crashing down” (Khan, 2012, p. 87).

I believe the problem is getting worse as students take watered-down Algebra 1 ever earlier, before they are ready for abstract thinking. In the preparatory school system, students are rushed into taking Algebra 1 in middle school so they can eventually take Calculus earlier, which may improve their chance of admission to elite colleges. Since their choice to pursue math through calculus is driven by college admission, not interest in mathematics, students are content to pass through without gaining true understanding or mastery.

My positionality relative to this problem interacts with many aspects of the problem. Dimensions of positionality “include one’s demographic positioning within society, one’s ideological positioning, and how one discursively positions the other and oneself” (Briscoe, 2005, p. 31). Briscoe advises us to look beyond demographic positioning since humans are too complex to sum up with just dry demographic data. Self-awareness and analysis are hard, and
“the researcher must be prepared for the personal and professional consequences for turning one’s gaze within” (Fennell & Arnot, 2008, p. 533).

**Privileged mathematical understanding.** My learning and teaching of mathematics spans more than forty years. As a student in public schools and an engineering major in college, I experienced sixteen years of math as a successful student. As a teacher, I have spent fifteen years crafting ways to both teach and assess mathematics. This experience adds up to a privileged position of mathematical understanding. I never struggled to learn math, which may mean I had amazing teachers, a strong work ethic, a natural tendency to think mathematically, or all of the above.

Knowing math as well as I do, with no history of difficulty, makes it challenging for me to understand why students have persistent misconceptions. I attempt to explain things in effective ways that will get to the root of a misconception in a way that has a lasting impact, but these approaches have not proven as successful as I would like. My intimate math understanding is a double-edged sword, as it not only allows me to teach effectively, but it also makes it more difficult for me to reach struggling students.

**Biases about students.** When my Algebra 2 students demonstrate fundamental misconceptions, I find my biases surfacing. While I do believe that all students can learn math and I work hard to foster a growth mindset in my students (Dweck, 2006), some of the lowest-achieving students push me to the edge of this belief. I also am inclined to believe they should “just try harder,” though that is akin to asking a blind person to try harder to see. I believe I treat my students fairly and without bias regarding race, gender, or class, but I also worry I may not see my own biases.
**Biases about middle school.** I believe that math teaching in many middle schools is substandard, even at some of the best middle schools in the nation. Middle school math teachers all too often fail to ensure their students learn some of the most basic arithmetic and algebraic concepts, perhaps because the teachers themselves don’t deeply understand the math. In 2007, 69% of US fifth- through eighth-grade students were being taught mathematics by teachers who did not possess a degree or certificate in mathematics (Augustine, 2007). Even worse, in a study of middle school math teachers in Michigan, only about half of the teachers felt academically very well prepared to teach some basic algebra skills. Even fewer teachers (38%) felt they had adequate content knowledge to teach other vital, basic concepts, such as slope (Schmidt & Mcknight, 2012). There is simply no possibility for students to learn math deeply and well from teachers who don’t understand mathematics. At the same time, despite the lack of mastery, students often earn A’s and feel good about their mathematical abilities. This is illustrated in the 2003 TIMSS, in which 55% of US 8th grade students had high self-confidence in math, while only 8% had high performance (Ginsburg, Cooke, Leinwand, Noell, & Pollock, 2005). It truly frustrates me to spend so much time futilely trying to correct students’ faulty mathematical education.

**Biases around demographic issues.** As a middle-class white girl growing up in suburbia, I never experienced racism or classism. My backpack was well and truly empty (Franklin, 2014) and I saw the country as a generally equal playing field for all races, classes, and sexes. I have since realized that I was contentedly oblivious to my own privileged positionality and blind to the systemic biases that exist in our country. However, though I have learned that racism and classism are far more pervasive and damaging than I can ever truly understand, I do not think race or class is relevant to this problem in my particular context. In
my experience, many of my students with the weakest skills are white. Since I rarely know the socioeconomic status of my students, I don’t know if class is relevant, though I believe the problem exists across all classes. Anecdotally, my lowest achieving students have been almost entirely female. While I certainly don’t believe that females are less talented in math, my experience show they often struggle the most with misconceptions. This may be in part due to societal norms such as “girls are bad at math” that young girls absorb or their tendency to avoid asking questions in class. My goal is to reach students of all races, classes, and genders with this intervention.

**Relationships with students.** My relationships with the students in my study was varied. Since none of my own classes were involved in the study, my teaching role was limited to introducing and monitoring the intervention. Though a handful of students lived in my dormitory or were on my sports team, this extra relationship dimension did not have an impact on the intervention.

**Preserving neutrality.** Self-awareness of one’s positionality is the primary key to preserving neutrality. Machi & McEvoy (2012) advise us that “careful introspection can bring these personal views forward, where they can be identified as what they are” (p. 19). As an engineer and statistician, I look forward to analyzing my data, though I’m well aware of how statistics can be used to bias results, especially if one is unaware of one’s own biases. To mitigate this bias, I have asked others to review my work, both those who know me personally to address positionality issues and those who know me professionally to address technical and statistical issues. With these measures in place, I improved my ability to stay neutral while conducting my research.
My positionality predisposes me to reach conclusions that align with my own views. Given that I designed and piloted the No-to-Yes intervention over the last three years, I was predisposed to find that my intervention would successfully remediate persistent algebraic misconceptions held by my students. However, for my work to be valid, I needed to move beyond my own biases and hold consistently to rubrics and protocols to enhance the validity of the study. By deliberately factoring my positionality into my research, I moved closer to my goal of developing an intervention that helps students succeed in mathematics and beyond.

**Research Questions and Hypotheses**

In the No-to-Yes intervention, which will be discussed in greater detail in the Theoretical Framework section and Chapter 3, the presence of students’ algebraic misconceptions was measured by student scores on an identical pre-test and post-test. The Misconception Error Score (MES) represented the total number of errors attributable to misconceptions as opposed to a careless mistake. For example, suppose that of the 28 questions on the pre-test, a student answered 13 questions correctly and made 15 errors. Ten of the fifteen errors were directly attributable to the presence of a misconception, so the student has a Misconception Error Score (MES) of 10. The other five errors were due to arithmetic mistakes and hence not directly related to a misconception.

Students were assigned by intact classes to an intervention group or the control group. For students participating in the intervention, engagement was measured by statistics obtained from records of activity in the intervention, such as number of correct practice quizzes, number of video views, and number of practice slips correctly completed. Students in the control group took the pre- and post-test, but did not receive the intervention. Since there was very little daily class time involved in the intervention, the control group did not receive an alternate assignment.
Teachers in the control group were encouraged to deliver the intervention in the spring term to their students.

Four research questions were investigated. Research Questions 1 and 2 examined the differences in the results between the two intervention groups and the control group both overall and by misconception category (MC). Focusing on the two intervention groups, Research Question 3 explored the relationship between engagement and gain in test scores. Finally, Research Question 4 explored the experiences of the students in the intervention groups. For all RQs, the subjects were students in Algebra 2 classes at a private secondary school in New England.

**Research question 1 (RQ1).** To what extent were there significant differences over time in the MES on a test of algebraic misconceptions among the control group and two intervention groups?

The dependent, ratio level variables for this question were the MES on a pre- and post-test of algebra skills. The independent, categorical variables were the student group assignment (one of two intervention groups or the control group) and time. The corresponding null hypothesis was:

**Null hypothesis 1.** There was no interaction between MES over time among the control group and two intervention groups.

Given that the ultimate goal of the intervention was to increase student success in mathematics by decreasing the number of errors students make due to their persistent algebraic misconceptions, the first research question got at the heart of the study. Did students make fewer errors due to common misconceptions after the intervention than before?
Research question 2 (RQ2). For each MC, to what extent were there significant differences in the change in MES among students in the control group and successful students in the intervention group, after excluding students who earned a “Yes!” on the pre-test for that MC?

The dependent, ratio level variable for this study was the MC-specific MES on a pre- and post-test of algebra skills. The independent, categorical variable was the student group assignment (one of two intervention groups or the control group). There were seven corresponding null hypotheses, one for each MC:

Null hypotheses 2A-G. For MC1-7, there was no difference in the distribution of the change in MES among students in the control group and successful students in the intervention group, after excluding students who earned a “Yes!” on the pre-test in that MC.

The second question examined the data at the MC level to understand how the intervention impacted performance on each MC. Past experience and literature (Bush & Karp, 2013; Cangelosi, Madrid, Cooper, Olson, & Hartter, 2013; Pitta-Pantazi, Christou, & Zachariades, 2007) indicate that some misconceptions are “stickier” than others, requiring more effort to overcome. These misconceptions may not show significant improvement after the intervention, while other misconceptions respond quickly. Further, studies have found that incomplete conceptual change can confuse students and cause them to make more errors than before (Adams, 1998), which could lead to an increase in misconception errors rather than a decrease. By comparing p-values, it was possible to quantify for which MCs the intervention had the largest and smallest effect on the change in MES.

Research question 3 (RQ3). To what extent did engagement predict the depth of conceptual change, as measured by the change in MES in the two intervention groups?
The dependent, ratio level variable for this study was the change in MES, which was computed by subtracting the pre-test score from the post-test score, so that a negative number meant the student made fewer errors after the intervention. Two hypotheses were tested. The first examined the relationship between total engagement and change in MES, while the second examined three individual components of engagement for their relationship to the change in MES. The independent, ratio level variables were the individual intervention engagement statistics related to videos, quizzes, and in-class slips as well as the total. The corresponding null hypotheses were:

**Null hypothesis 3A.** There was no relationship between total engagement and the depth of conceptual change as measured by change in MES among students in the two intervention groups.

**Null hypothesis 3B.** There was no relationship between total engagement in videos, number of correct online quizzes, and number of correct in-class practice slips and the depth of conceptual change as measured by change in MES among students in the two intervention groups.

The third research question sought to support or refute the warm conceptual change theory (WCCT) idea that the interaction between the student and the message (the intervention) leads to depth of engagement, which in turn causes conceptual change (Dole & Sinatra, 1998). According to the theory, a higher level of engagement in the intervention should lead to the conceptual change necessary to decrease the number of misconception errors on the post-test.

**Research question 4 (RQ4).** How did participating students experience the intervention?

This qualitative question sought to understand the experience of the students in the intervention groups. The question was explored through focus groups with each of the
intervention group classes and augmented by interviews with teachers. Questions included what students liked and did not like, how they used the online component, and recommendations for improvement.

**Theoretical Framework**

The study is grounded in “warm” conceptual change theory (WCCT), which adds motivational, affective, and contextual constructs to purely cognitive or “cold” models of conceptual change (Sinatra, 2005). WCCT acknowledges that emotions matter and that learning is much more than just an objectively rational endeavor. A warm model accepts that learners have different individual feelings and motivations, such as their response to social pressure or need to achieve, that affect how deeply the learner will engage in the process of changing their conceptions.

Dole and Sinatra’s (1998) Cognitive Reconstruction of Knowledge Model (CRKM), shown in Figure 1, was used in this study to model how a learner and a message interact to create the cognitive engagement that leads to conceptual change. The intervention’s videos and practice quizzes were carefully designed to induce a state of cognitive conflict, or CRKM’s “dissatisfaction” (Alvermann & Hague, 1989; Gill, Ashton, & Algina, 2004; C. R. Hynd, 2001; Mason, 2002), in order to encourage engagement. These ideas are discussed in greater depth below.
Figure 1: Dole and Sinatra's (1998) Cognitive Reconstruction of Knowledge Model.

Conceptual change theory. The term conceptual change is used to “characterize the kind of learning required when the new information to be learned comes in conflict with the learners’ prior knowledge usually acquired on the basis of everyday experiences” (Vosniadou & Verschaffel, 2004a, p. 445). This kind of learning is not just adding something new to current knowledge, but instead requires learners to reorganize what they already believe to be true. Since deeply rooted beliefs are notoriously difficult to change (Disessa & Sherin, 1999), it is simply not enough just to tell someone they should change a conception; instead, there has to be a way to activate this conceptual change, usually through causing a state of cognitive conflict (Guzzetti, Snyder, Glass, & Gamas, 1993). While most research in CCT has been in the realm of science, especially physics, which is acutely prone to misconception, there is no apparent reason
CCT cannot be applied successfully to mathematics, where gaining deep understanding often requires conceptual change (Vamvakoussi and Vosniadou, 2006).

Posner, Strike, Hewson and Gertzog’s (1982) seminal work on conceptual change advocated making beliefs explicit, reflecting on them, then refuting misconceptions through sound, plausible arguments. Behind Posner’s work lie the ideas of Piaget (1977), who wrote of accommodation to new conceptions, Vygotsky (1986), who advocated increased reflection on mental processes to transform implicit theories into explicit ones, and Kuhn (1970), who wrote of the paradigm shifts necessary to change conceptions.

As discussed further below, seminal authors in conceptual change theory have deemed augmented activation (AA) and refutational text (RT) as effective methods for creating cognitive conflict, which in turn creates conceptual change (Posner et al., 1982; Strike & Posner, 1992). AA makes implicit theories and misconceptions explicit and then calls into question the usefulness or truthfulness of such beliefs, while RT induces cognitive conflict by having subjects read a text designed to introduce incongruity or dissonance (Guzzetti et al., 1993; C. Hynd, Alvermann, & Qian, 1997).

More recently, models such as Posner’s (1982) have been criticized as being too “cold,” or exclusively cognitively based, with more modern models taking into account “warm” features such as affective, situational, and motivational factors (Pintrich, Marx, & Boyle, 1993; Sinatra, 2005; Strike & Posner, 1992). These warm models, based on Pintrich’s (1993) seminal work, theorize that individual learner differences and contexts lead to different levels of engagement with the ideas being presented, which in turn leads to different degrees of conceptual change.

Two warm models in the literature are Dole and Sinatra’s (1998) Cognitive Reconstruction of Knowledge Model (CRKM) and Gregoire’s (2003) Cognitive-Affective
Model of Conceptual Change (CAMCC), which is targeted at changing teachers’ conceptions of reform messages. Both CRKM and CAMCC theorize that depth of engagement is crucial to conceptual change. When a person is confronted with a persuasive message, he or she either processes it via a central route (deliberate, systematic, and deeply engaged) or a peripheral route (quick heuristic at a low level of engagement) (Eagly & Chaiken, 1993). The goal is to activate the central route and provide tasks at the right challenge level to encourage deep engagement rather than easy heuristic processing of new information.

The major criticisms of CRKM come from the author of the competing model, who criticizes CRKM for failing to specify how peripheral cues affect conceptual change, neglecting the role of anxiety, and ignoring the role of automatic or instinctual evaluations in perceiving a message as a threat or an opportunity (Gregoire, 2003). While the third criticism is not as important for this study involving students as with teachers facing school reform, math anxiety certainly plays an important role in how students learn in math class and should be included in a useful model.

**Tenets of WCCT.** The tenets of CRKM (Dole & Sinatra, 1998) include characteristics of the message, learner, and peripheral cues, as well as how their interaction creates deep engagement that leads to change.

**The message.** The CRKM specifies that the message must be comprehensible, coherent, plausible, and rhetorically compelling. These reasonable requirements ensure the learner is willing to engage with the message rather than rejecting it for being confusing, unbelievable, or uninteresting. Material should be written at the learners’ level of understanding and maturity to inspire engagement.
**The learner.** Many characteristics determine how the learner will interact with the message and hence how deeply they will engage. These characteristics include background knowledge, motivation, dissatisfaction (cognitive conflict), personal relevance, social context, and need for cognition. It can be difficult to foster engagement in all learners with the same message, since learners vary widely in terms of these characteristics.

**The peripheral cue.** Instead of having conceptions changed by a compelling argument, students could be persuaded to change their mind via a peripheral cue, which could be an attractive or credible source or an easily understandable, simple message such as a slogan or mnemonic (Eagly & Chaiken, 1993). Such cues may lead to quick, heuristic, shallow engagement and a temporary change in beliefs. While a conceptual change induced this way is often superficial, the CRKM leverages the peripheral cue to draw the learner into high engagement with the message to create long lasting conceptual change (Dole & Sinatra, 1998).

If an exciting, simple peripheral cue can help students engage in the message, it makes sense to include one.

**High engagement.** This fundamental tenet posits that high engagement is a necessary condition for successful conceptual change. Without engagement, no change is possible. Further, the greater the engagement, the higher the likelihood of conceptual change occurring.

**Impact of WCCT on the study.** The grounding in WCCT and specifically the CRKM made this study richer in many ways. First, CRKM and WCCT influenced the design of the message and the inducement of dissatisfaction (cognitive conflict) in the learner. A peripheral cue was added in the form of process mnemonics to spark interest and aid long-term retention. RQ3 specifically targeted the WCCT tenet that the level of engagement influenced the degree of conceptual change. Finally, some of the constructs in the CRKM were investigated in the focus
groups, including motivation and social context. These proved to be quite powerful in explaining the depth of engagement and resulting conceptual change.

**The No-to-Yes intervention.** A pilot intervention has been in development by the researcher and her colleagues for the last three years. The impetus for the project was the frustration of math teachers at this school over a set of errors that all agreed were pervasive, difficult to eradicate, and ultimately damaging to student success. The pilot intervention started with a pre-test covering seven fundamental algebraic misconception categories (MCs). If students did not initially answer all the questions related to an MC correctly, they practiced on paper slips in class each day until they could answer a similar slip correctly on three different days. At the end of the intervention, each student took an identical post-test to see if improvement had occurred. There was no online component and no explicit incorporation of WCCT. The results from the pilot showed marked improvement in post-test scores. Of the 10 students that completed the intervention in the researcher’s Algebra 2 class during the winter term of 2016, the mean number of questions correct increased from 17.7 to 25.1, a 42% increase. Even more striking, the number of MCs earned increased 174% from 1.9 to 5.2.

For this study, the pilot intervention was redesigned to align with the CRKM (Dole & Sinatra, 1998) by adding an online tutorial component and addressing specific characteristics of the learner and the message (see Chapter 3 for more details). First, learner constructs related to motivation were addressed. Dissatisfaction was induced through augmented activation and refutational text in the online videos, which was intended to provoke students to change their misconceptions in a deep and long-lasting way. The social context of the classroom and wall chart provided an environment that may have motivated some students, while a grade for completing the intervention may have motivated others to find the intervention personally
relevant. Students with a high need for cognition should have responded to the videos helping them finally make sense of each MC. The message (online videos and practice quizzes) were designed to be comprehensible, coherent, plausible, and rhetorically compelling by combining multiple sources, worked examples, and embedded interactivity in an engaging delivery platform. Measuring engagement through activity statistics gave insight into the theoretical mechanism causing conceptual change.

Conclusion

For students to reach their full potential academically and professionally, a strong grounding in mathematics is essential. Unfortunately, many students have persistent algebraic misconceptions that prevent them from achieving this potential. To remediate these misconceptions, an intervention was created based on warm conceptual change theory, specifically the CRKM (Dole & Sinatra, 1998). The CRKM argues that conceptual change occurs due to student engagement created by the interaction between learner and message. By analyzing pre-test vs. post-test results and engagement, the current study sought to determine if the intervention decreased errors and if depth of engagement played a role in that decrease. Given this background, the next chapter reviews the literature in depth.
Chapter 2: A Review of the Literature

A literature review was conducted to better understand the existing knowledge base related to the existence and persistence of algebraic misconceptions that limit the success of secondary school students in mathematics. This review will first examine ways to diagnose students’ algebraic misconceptions and then describe the general types of misconceptions that have been identified in the literature. Next, it will explore how conceptual change theory, specifically the “warm” version, can facilitate the difficult process of changing deeply rooted beliefs. The review will then report on the results of interventions designed to root out misconceptions in mathematics classrooms and evaluate the success of these interventions. Following the exploration of misconceptions, the review will explore two ideas that informed the design of the intervention: spaced/interleaved learning and the use of process mnemonics as a peripheral cue to increase long-term retention.

Diagnosing Misconceptions

Researchers have developed many different diagnostic methods and instruments to diagnose a student’s misconceptions. An accurate diagnosis is crucial prerequisite to remediation, so an efficient and effective diagnostic is essential. In some studies, students were given a general algebra skills test with open-ended questions (Akyüz, 2015; J. L. Booth et al., 2015; Poon & Leung, 2010) or carefully designed multiple choice questions that included a misconception response along with the correct response and two distractors (Lucariello et al., 2014; Russell, O’Dwyer, & Miranda, 2009). Yet other diagnostics were based on interviews alone (Vamvakoussi & Vosniadou, 2006; Vlassis, 2008) or a combination of tests and interviews (Almog & Ilany, 2012; Cangelosi et al., 2013; Lim, 2010; Pitta-Pantazi et al., 2007).
Each method has its benefits and drawbacks. Diagnosis of misconceptions is difficult because an error could be made due to a number of causes: a student might misapply a rule, misunderstand a rule, make a careless error, or just guess. Further, understanding a student’s thought process is vital to diagnosing the misconception in play. While multiple-choice tests are the easiest to grade and can be administered efficiently to the greatest number of students, they do not show insight into student thinking. On the other extreme, interviews give great insight into a student’s thinking, but are time-consuming and cannot be administered to many students quickly. In the middle, open-ended tests give some insight into student thinking in a timely manner, while mixed methods can combine the best of both worlds. Across all these methods of diagnosis, researchers have found similar broad categories of misconceptions that are described in the following section.

**Types of Mathematical Misconceptions**

In addition to the studies specifically designed to diagnose misconceptions in the previous section, there is a large body of research into the general types of algebraic errors students make and their corresponding underlying misconceptions. Much of this research is over twenty years old, evidence that the problem has been around for decades, though there is very little research from that era about remediation of these errors. The root problem is that students (and perhaps teachers) focus on memorizing and using rules without reason (instrumental understanding) rather than understanding the structure of the math itself and knowing both what to do and why (relational understanding) (Skemp, 2007). Math teachers must help students understand the underlying structure and pattern rather than simply offering rules to memorize. Further, because students overgeneralize and respond incorrectly to what is visually salient or what looks correct, “algebra curricula need to give explicit attention to parsing and to structural
notions in ways that will make structure a strong competitor for perceptual salience among the many impulses competing for student attention” (E. Merlin, 2008, p. 82).

**Variable use.** Variables are the fundamental building blocks of algebra, but students often misunderstand their meaning and use. A variable represents an unknown, varying quantity in an equation. For instance, the variable \( A \) might stand for “the number of apples in a bowl”, though students may instead misunderstand \( A \) to mean “an apple”. The most common, well documented variable misconceptions include viewing a variable as a label for an object, assigning a particular value to a letter, and assigning values based on their position in the alphabet (Booth, 1988; Lucariello et al., 2014; MacGregor & Stacey, 1997; Stephens, 2005). Students incorrectly based interpretations of letters on intuition, analogies, or “a false foundation created by misleading teaching materials” (MacGregor & Stacey, 1997, p.15) After working with fifth grade students for 18 months, Lannin (2008) found that use of formal variable notation must be developed very gradually if students are to avoid merely manipulating symbols without connecting them to their underlying meaning. Variable misconceptions are at the root of the problem that opened Chapter 1. Students make this error because they forget the rule about crossing out letters. However, if \( a \) and \( b \) are understood to be variables representing numbers, then no memorized rule is needed at all.

In the classic student-professor problem, students commonly reverse variables when translating an expression like “there are six times as many students as professors” into math, coding this incorrectly as \( 6S = P \) rather than correctly as \( S = 6P \) (Booth, 1987). The misconception is that \( 6S \) stands for \textit{6 students} rather than \textit{6 times the number of students}. When studying variations of this problem, only 15% of high school students could correctly answer a similar problem dealing with the value and number of dimes and pennies (Philipp, 1993).
**Equals sign.** Equality, often represented with an equal sign, is a fundamental and seemingly simple algebraic notion. However, math teachers often train students to view the equal sign as a signal to do something (operational view) instead of an indication that the two sides of the equation are equivalent to each other (relational view). The relational view is fundamental to success in algebra; students who view the equal sign as relational outperform their peers who view it strictly as an operator (Knuth, Alibali, Hattikudur, McNeil, & Stephens, 2008). In a similar way, students also frequently confuse an expression ($2 + x$) with an equation ($2 + x = 6$), demonstrating an inability to treat an expression as an object rather than an operation to be carried out. For example, students often incorrectly treat an expression as an equation by cross-multiplying two fractions being added together.

**Order of operations.** Unlike reading text, math is not strictly read from left to right. Instead, there is a hierarchy known as the order of operations. In the U.S., students typically memorize the acronym PEMDAS, which stands for Parentheses, Exponents, Multiplication, Division, Addition, and Subtraction. This acronym, while intended to help students remember the order, can foster misconceptions in students who do not understand the underlying structure. For instance, addition comes before subtraction in the acronym, but these two operations have the same priority and are computed left-to-right. Hence, for the problem $6 - 4 + 3$, students mistakenly calculate $6 - 7 = -1$ instead of $2 + 3 = 5$. This very common misconception is directly attributable to blindly applying the acronym without understanding the meaning (Hewitt, 2012; Schwartzman, 1996).

**Algebraic notation.** Conventional notation is notoriously difficult for students learning algebra. In order to participate in the language of mathematics, students need to develop notational fluency. More authors now realize that “symbolization activities are intrinsically
associated with the emergence of meaning and formation of new concepts” (Vlassis, 2008, p. 558). In other words, using symbols (notation) properly is a clear indication of understanding. Students must be savvy enough to understand that \( x^{-1} = \frac{1}{x} \), but \( f^{-1}(x) \neq \frac{1}{f(x)} \). When expressing the equation of a line, students often believe that the standard-form convention \( y = mx + b \) is superior to all other forms used to characterize a line. As a result, they inflexibly ignore other more useful forms (Fisher, Borchert, & Bassok, 2011). Student use of parentheses is also problematic, as they can be used both to indicate order of operations, as in \((3+4) – (4-5)\), as well as multiplication, as in \((3+4)(6)\) (Welder, 2012). In another example, students are uncomfortable leaving \((2 + x)\) as is, so they concatenate, or position the symbols next to each other, to give \((2x)\) (Hewitt, 2012). Concatenation errors also stem from confusing notational inconsistencies: for instance, \(8 \frac{1}{3}\) means \(8 \text{ plus } \frac{1}{3}\), while \(8a\) means \(8 \times a\).

**Use of the negative sign.** The negative sign is particularly problematic for students, since it has different meanings depending on how it is used in an expression:

\[-2x - (6x^{-3} + 2)\]

In this single expression, the minus sign represents a negative quantity, subtraction, and a reciprocal. Students often have an inflexible understanding of the minus sign and hence are unable to distinguish between the various functions it plays, falsely believing that subtraction is its only role (Vlassis, 2008). Another common misconception is the sticky sign, in which students incorrectly view \(-9^2\) and \((-9)^2\) as equivalent (Cangelosi et al., 2013). Perhaps due to lack of clear examples in textbooks, students are particularly bad at handling signed numeral forms correctly such as \(-4.7^2\) or \(-8 \frac{1}{3}\) (M. A. Lee, 2000). Overall, “a weak understanding of negative numbers seems to be the fundamental source” of student errors (Titus, 2011).
**Rules of exponents.** Classic errors involving operations with exponentiated variables include addition/subtraction: $x^3 + x^4 \neq x^7$, multiplying: $x^3 \times x^7 \neq x^{21}$, dividing: $x^6 \div x^2 \neq x^3$, power to a power: $(a^3)^2 \neq a^5$, or the zero exponent: $a^0 \neq 0$. The concept of exponentiation is “difficult for two reasons. First, one has to consider the relationship between symbols, meanings, and properties of exponents… Second, the conceptual understanding of exponents involves a deep understanding of the number system with its strict hierarchy” (Pitta-Pantazi et al., 2007). Once again, students merely memorize rules but then do not know when a particular rule applies.

While it is well documented that students have problems with exponents, there are very few practical strategies for ways to teach the underlying structure. One option used by the researcher is to expand the expression to better understand what it represents. For instance, $x^2 \times x^3 = (x \times x) \times (x \times x \times x) = x^5$.

**Negative exponents.** While students have difficulties with rules of exponents in general (Bush & Karp, 2013), these troubles intensify when there is a negative in the exponent. In study of 202 high school students, 25% were unable to extend their knowledge of positive exponents to negative exponents, while another 50% could do so with only limited success (Pitta-Pantazi et al., 2007). In another study, students had trouble with exponential expressions containing a negative sign due to fundamental misconceptions such as the meaning of inverse (Cangelosi et al., 2013). Many students incorrectly think a negative exponent makes a term negative.

**The distributive law.** There are many common situations where students apply the distributive law incorrectly. One ubiquitous error is failing to distribute the negative sign inside parentheses: $-(a - b) \neq -a - b$. Another error is so common it was nicknamed the “Freshman’s Dream” due to the struggles of college freshman taking algebra: $(a + b)^2 \neq a^2 + b^2$ (Hungerford, 1974). The error is incorrectly distributing an exponent across addition or
subtraction, which is also commonly seen as $\sqrt{x^2 - y^2} \neq x - y$. Other errors include incorrectly distributing multiplication across multiplication: $3(xy) \neq 3x \cdot 3y$ or failing to distribute the exponent across multiplication: $(ab)^2 \neq ab^2$ or division $\left( \frac{a}{b} \right)^2 \neq \frac{a^2}{b}$. All these errors “indicate an inability to recognize when one operation is distributive over another” (Schwartzman, 1977, p. 594). Although some students successfully memorize the rules of exponents, many do not understand the structural properties that underpin the rules, leading to frequent errors (Schwartzman, 1996). It is difficult to find teaching materials targeting these misconceptions, though some websites discuss a geometric or adjacent levels approach (“Combining operations (distributive laws),” n.d.; Schwartzman, 1977).

**Simplifying algebraic fractions.** In this error, students randomly cancel the same letter or number from a fraction without regard to the underlying structure: $\frac{3x+6}{3y+4} \neq \frac{x+6}{y+4}$. This common error has plagued students and teachers for decades. In 1924, Grossman lamented that every teacher of experience knows that a great many of his algebra pupils all the way from the first year in high school up to college continue with almost comical regularity to make strange mistakes in the subject of "cancellation" in fractions - mistakes that show clearly that the essence of the matter has escaped them (Grossman, 1924, p. 104).

The root of this problem is that students do not structurally understand the difference between a factor (a part of a product) and a term (a single number, variable, or product of them) (Grossman, 1924; Martinez, 1988). Despite the prevalence and persistence of this kind of error, there is little literature on the subject except to classify these kinds of simplification errors through interviews (Hong, 2013; Makonye & Khanyile, 2015). Very few have imposed their own structure, such as “glue and trees” (Merlin, 2014), making a strong analogy to non-algebraic fractions (Grossman, 1924), or “slashing” (Merlin, 2008).
Despite the extensive literature around diagnosing and cataloguing misconceptions, practical solutions are rarely found anywhere in the literature. It is often necessary to go to the internet to find such work posted by classroom teachers rather than finding published journal articles (Schechter, 2009; Schofield, 2003). Given the broad agreement on the existence and stickiness of mathematical misconceptions, there is a clear need for effective ways of changing students’ underlying beliefs that can be implemented in real classrooms.

**Conceptual Change Models**

Given the identification of student’s misconceptions, how can they be eliminated? Since these issues have persisted over decades, it seems the standard methods tried in the past just do not work (Marquis, 1988; Smith, 1946; Webber, 1929). When misconceptions are present, changing those beliefs requires conceptual change, which is “much more challenging than other kinds of learning…. (it is difficult) for even the best algebra teachers to correct misconceptions that students have internalized through years of prior mathematics instruction” (Welder, 2012). Math teachers need to learn from researchers in conceptual change, which occurs when a student reorganizes prior knowledge to correctly understand an idea.

Conceptual change theory is well researched in science settings, but much less so in mathematics. Given that there are similarities in “naïve understandings” in math and science, that both are based on the understanding of certain core principles, and that there are common difficulties in learning them, it is reasonable to conclude that conceptual change models can be successfully applied to learning mathematics (Tirosh & Tsamir, 2004; Vosniadou & Verschaffel, 2004b). According to Vamvakoussi and Vosniadou (2006), cases where understanding requires conceptual change can and do occur in mathematics learning. This may suggest that mathematics education can profit from the educational
guidelines provided by the conceptual change approach literature. Increasing mathematics educators’ sensitivity about conceptual change problems is a first step and may help them to explain why students keep “forgetting” some things (p. 465).

The core tenet of most conceptual change models is that effective instructional interventions are designed to create cognitive conflict in students (Guzzetti et al., 1993). The idea is that by forcing students to confront their misconceptions through counterexamples and contradictions, they are more likely to change their conceptions. In these interventions, after students first identify their preconceptions, the teacher presents some kind of conflicting information that forces the student to reevaluate the prior conception, which now appears unacceptable.

The evidence on the effectiveness of cognitive conflict to produce conceptual change is mixed. Contemporary research indicates that it is not enough merely to introduce conflicting evidence; in addition, the intervention must create meaningful conflict by addressing student motivational factors such as goals and self-efficacy, students’ prior knowledge, values and attitudes, cognitive engagement and reappraisal, social factors, and affective factors (Lee et al., 2003; Limón, 2001; Yok-Cheng Sam, Choo-Yee Ting, & Chee-Onn Wong, 2012). There are many instructional formats designed to induce this conceptual conflict. This section will briefly discuss a range of techniques, focusing on empirical studies using varied methods of conceptual change, including the techniques that were used in the intervention.

**Conceptual change discussions.** One technique is to use conceptual change discussions to induce cognitive conflict. In one study, students learning about force and motion were involved in weekly 20-minute conceptual change discussions in which the teacher led with an exposing event to allow students to reveal their misconceptions, then followed up with a
discrepant event, which created conflict between the exposed misconceptions and some observed phenomenon. The discussions ended with examples, summary, and a feeling of progress and growth for students. Teachers believed the method would have been even more effective if they had more training on facilitating the discussions (Eryilmaz, 2003).

**Anomalous data.** Another technique is to use anomalous data, or data that does not fit with a current conception, to change students’ beliefs about controversial topics. In a science study, a pre-test showed the majority of students believed that the extinction of the dinosaurs was due to a meteor hitting the Earth. She then gave students a text that blamed a volcano’s eruption for the extinction. Finally, she gathered data to see if the students had changed their beliefs and why. While some students did change their conceptions of dinosaur extinction, there were limitations of the study. These included the lack of a “right” answer to compel the students, lack of conversation and discussion, and lack of strong stimulation to recognize a conceptual conflict (Mason, 2002). In another study, using anomalous data, Chow and Treagust (2015) showed a positive effect on facilitating students’ conceptual understanding of algebra. This method appears to be most effective when applied to a misconception that has a clear truth and a distinct correct answer, such as in mathematics.

**Bridging analogies.** Another method is bridging analogies, in which students use a series of analogies to bridge from their current misconceptions to target correct conceptions. Students are first tested to identify misconceptions about a target concept. The teacher then presents other, simpler examples related to the target concept. If students correctly understand them, they are known as anchoring analogies. The students then explicitly compare the anchor and target cases to see if they are analogous. The teacher may introduce intermediate analogies if necessary. The intent is to link a known concept with to an unknown, problematic example
Science researchers have found this technique effective in reducing student misconceptions (Savinainen, Scott, & Viiri, 2005; Yilmaz, Eryilmaz, & Geban, 2006).

**Diverse instruction.** Yet another method designed to generate conceptual conflict is diverse instruction, in which more teachers employ more than one way of changing a conception at the same time. Hayes et al. (2003) found that when instruction challenged diverse aspects of children’s misconceptions, it was more likely to produce conceptual change. Young children’s belief that the earth is flat was somewhat changed when they were presented with videos about gravity or earth’s size, but the effect was multiplied when they were presented with information on gravity and earth’s size. It makes sense that multiple avenues of attacking a misconception would be more successful than one way alone.

**Augmented activation.** Before presenting new, correct information, researchers can activate students’ prior beliefs about a topic to prime them for the new concept. However, merely activating a prior misconception may only serve to reinforce it, making the task of conceptual change more difficult (C. Hynd et al., 1997). Augmented activation is a variation that activates “students’ prior knowledge of a concept, supplemented by information to cause dissatisfaction with a current belief or correct a misconception” (Guzzetti, Snyder, & Glass, 1992, p. 646), hence creating cognitive conflict. Typically, augmented activation includes combining an activity, example, or text with a warning to look out for forthcoming information that might be different than current beliefs (Alvermann & Hague, 1989; Alvermann & Hynd, 1989). Presenting an example or demonstration that creates surprising results can also cause students to become dissatisfied with their current thinking (Guzzetti, 1990). A meta-analysis of studies using augmented activation with post-secondary students showed an “average effect size of .80 (almost an entire standard deviation unit) when compared to control conditions” (Guzzetti
et al., 1992). Despite many studies that support the positive effect of augmented activation, there are exceptions. For instance, Pena and Alessi compared augmented activation to expository instruction and showed no advantage for augmented activation (1999). However, the description of the instructional strategy more closely matched regular activation, as it did not include an explicit warning to attend to contradictory beliefs. Overall, augmented activation has been shown to be one of the most effective strategies to create conceptual change.

**Refutational text.** In contrast to expository text that does not reference students’ prior conceptions, refutational text directly describes a misconception then refutes it by providing evidence for the correct conception (Mason, Gava, & Boldrin, 2008). Optionally, a third element called a refutational cue can be added to alert the reader that the refutational of the misconception will follow (Lem et al., 2015). An example of a refutational text that contains all three elements reads:

*Warning!!* When you look at a boxplot you might think that the larger part of the box represents more results than the smaller part of the box. This is incorrect! In each of the four parts of a boxplot (approximately) the same number of observations is represented (Lem et al., 2015, p. 916).

A meta-analysis concluded that “two decades of research indicate that reading refutational text rather than traditional expository text is more likely to result in conceptual change” (Tippett, 2010, p. 951). Among all the cognitive change strategies, refutational text has shown to be particularly effective at changing misconceptions (Guzzetti, 2001).

Many interventions, mainly in science, have been conducted using refutational text. Many have found that refutational texts have improved post-test results (Broughton, Sinatra, & Reynolds, 2010; Lem et al., 2015). More specifically, refutational text worked better for
students with weaker understanding (Södervik, Virtanen, & Mikkilä - Erdmann, 2015) or when combined with worksheets (Korur, Enil, & Göçer, 2015), demonstrations (C. Hynd et al., 1997), videos (Caleon & Subramaniam, 2014), or multiple sources (Kendeou, Braasch, & Bråten, 2016). Students also preferred the refutational text format over an expository or narrative format (Mason et al., 2008; Sinatra & Broughton, 2012). In a study on changing preservice teachers’ beliefs about how children best learn math, the researchers used both augmented activation and refutational text to create conceptual conflict, which promoted a larger change in teacher beliefs than the expository text (Gill et al., 2004).

This overview of conceptual change models presents different ways researchers can induce a state of cognitive conflict that sets the stage for conceptual change to occur. The commonality among these studies is that “each strategy found effective as a single intervention in some way created in students’ minds a degree of discomfort with their prior beliefs” (Guzzetti et al., 1992). While most of these studies produced significant results compared to control groups that used traditional teaching methods, future work needs to include the contemporary criticisms of cognitive conflict so that it addresses motivation and other student characteristics. Finally, there is no reason to use these techniques in isolation. Researchers could combine the techniques in ways that induce an even more powerful state of cognitive conflict.

Now that common algebraic misconceptions and techniques to create cognitive change have been discussed, this review next examines interventions that have been conducted in math classrooms, focusing on the use and effectiveness of conceptual change models.

**Mathematics Classroom Interventions**

While there are have been many studies aimed at understanding and diagnosing students’ misconceptions (Almog & Ilany, 2012; Cangelosi et al., 2013; Knuth et al., 2008; Lucariello et
al., 2014; MacGregor & Stacey, 1997; Pitta-Pantazi et al., 2007; Vlassis, 2008), there have been very few studies that take the concept to its next logical level: how does one actually intervene and change student conceptions? There are even fewer studies that are set in a math classroom, and fewer still that directly reference conceptual change models. The following studies employ such models in a math context. The interventions include teacher-made formative assessments, computer-based interventions, calculator experiments, and discourse based instruction. Though more studies may exist, this section includes every study found that combined mathematics interventions and conceptual change. At the highest level, these studies demonstrate that even when armed with cognitive change theory, it can still be difficult to get students to change their misconceptions appropriately.

Adams (1998) conducted one of the few math studies that explicitly employed a conceptual change method. She created a conceptual change assignment to change students’ beliefs about functions and their domain, range, and graphing. Disappointingly, Adams found that students who used the conceptual change assignment actually performed worse on a post-test than control group students. She attributed this to either a lack of applicability of the model, or more likely, the weakness of the assignment. As discussed above, modern research indicates than cognitive conflict is more complicated than previously thought, and perhaps Adams did not take all the student variables into account.

A common student misconception is that all relationships are linear and proportional. For example, if 3 cups of water fill 1 jar, then 12 cups of water fill 4 jars. There is a proportional relationship between cups and jars. However, if the side of a square is tripled, the area is increased by a factor of 9, not 3; this is a quadratic relationship. Researchers used a cognitive change approach to change students’ misconception that every relationship is linear (Van
Dooren, De Bock, Hessels, Janssens, & Verschaffel, 2006). In the pretest, students mistakenly treated every relationship as linear, but by the time they took the post-test, they were thoroughly confused and mistakenly treated very few relationships as linear. Like in Adams’ (1998) study, it seems that Van Dooren et al. succeeded in getting students to question their misconception that everything was linear without providing enough instruction to help students choose correctly between linear vs. non-linear relationships.

Kramarsky (2004) researched how cognitive change accomplished through metacognitive instruction might help students address misconceptions around graphing and graph sense. He found that students who experienced cooperative learning with metacognitive instruction had more positive outcomes and fewer misconceptions that cooperative learning alone. His metacognitive activities included asking questions related to comprehension, reflection, connection, integration, and generation. He found that “mathematical discourse raised cognitive conflicts, which in turn encouraged students to discuss the conflicts and suggest ways for resolving them” (p. 611). He emphasizes that students must be made explicitly aware of their misconceptions or they will not realize there is a disconnect between their thinking and mathematical truth.

Moschkovitch (1998) looked at the robustly entrenched misconception that the x-intercept is directly shown in the equation of a line in standard form $y = mx + b$. The author prefers to call this a transitional conception, since she wanted teachers to view the confusion as an opportunity to bridge student thinking toward conceptions that are more accurate. Though she does not use the term cognitive conflict directly, she details how students worked collaboratively and participated in structured discussions designed to focus student attention on
the misconception. After the intervention, the author found that while there was improvement, many of the students continued to use the x-intercept incorrectly.

In another study on student misconceptions, Moschkovitch (1999) used linear functions, specifically the role of the x-intercept, to show how students working with graphing software and a peer were able to successfully change their misconceptions. They used four types of available resources: specific cases, connections to other conceptions, procedures, and descriptive language. She recommended that teachers explicitly identify and use these resources for their students to bring about cognitive change. For instance, teachers can use directed case studies that ask students to examine preconceptions, try to connect them to other conceptions, and negotiate shared understanding.

To address misconceptions about fractions, researchers used a two-stage performance test, a worksheet, and an conceptual change text (Koparan, Yıldız, Köğce, & Güven, 2010). They found that students in the experimental group using the instructional materials had increased success over the control group. Unfortunately, the researchers did not include any examples of their instructional materials for evaluation, so we only have the statistical results.

Digital interventions are growing in popularity and show promise for improving student learning (Bokhove & Drijvers, 2012; Craig et al., 2013; Russell et al., 2009). While it is relatively straightforward for a computer-based system to teach help students simply learn new techniques, the system must employ conceptual change methods if the goal is to remediate misconceptions. Bokhove and Drijvers (2012) developed a computer based intervention designed to increase algebraic expertise. They intentionally created conceptual crises, bringing in elements of conceptual change theory such as bridging analogies and anomalous data. In this method, students solved a pre-crisis item in a familiar way. They were then given a crisis item
that could not be solved with the usual method. After working through this item, they received a post-crisis item to practice the new skill. The students showed significant improvement in their fluency. The authors contend that causing intentional crisis is an effective approach for spurring learning and conceptual change. However, given van Dooren’s (2006) experience with changing conceptions of linearity, perhaps the post-crisis items should contain both pre- and post-crisis type items so the students must learn to choose the appropriate technique.

Russell, O’Dwyer, and Miranda (2009) wrote a computer program called the Diagnostic Algebra Assessment System (DAAS) to diagnose students that held misconceptions related to variables, equality, and graphing. In their study, misconceptioners were then subject to some combination of classroom intervention or control. Unfortunately, there was very little information in the article about the design of the lesson plans or conceptual change model employed. They found that full intervention (diagnostic, reports, and lesson plans) had a net positive, though not statistically significant, effect. Given the article’s focus on the diagnostic end, perhaps their lack of attention to the classroom intervention meant the lesson plans did not sufficiently create the cognitive conflict necessary to change misconceptions.

As mentioned above, many researchers have designed interventions using refutational text in science classrooms to study conceptual change. However, only two of the published studies on refutational text take place in math classrooms (Gill et al., 2004; Lem et al., 2015). Both found advantages to using refutational text as part of a classroom intervention.

Given that these articles represent an exhaustive search for empirical studies involving conceptual change theory in math, there is a clear need for more research to understand how to use this theory to remediate student misconceptions in math classrooms. Only one study (Gill et al., 2004) takes into account contemporary understandings of warm cognitive conflict and the
many variables that can affect the strength of the conflict, such as motivation, meaning, and affect. Another theme that emerged from the articles, especially from Adams (1998), Van Dooren (2006), and Moschkovitch (1998), is that it can be harmful to students to begin the process of conceptual change and not conclude it satisfactorily. If the teacher succeeds in dislodging a misconception without adequately implanting a correct new one, students may end up more confused than when they started. Finally, not all of these studies successfully changed students’ misconceptions in measurable ways, confirming the difficulty of the task. Having looked at the literature around ways to change misconceptions, the review now turns to ways to effectively plan the timing of lessons to ensure student learning is retained over time.

**Spaced and Interleaved Learning**

Spaced or distributed practice spreads out learning a concept over time, while interleaved learning mixes different kinds of problems together during a single study session (Rohrer, 2012). The two are interconnected, since interleaving inherently spaces learning out over a longer period of time. For instance, if a unit contains the three concepts A, B, and C, a blocked schedule would run AAABBBCC, while an interleaved and spaced schedule would be ABCABCABC.

The spacing effect has been studied extensively over time, with large effect sizes giving evidence of its robust effectiveness in increasing long term retention (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Kapler, Weston, & Wiseheart, 2015; Mettler, Massey, & Kellman, 2016; Rohrer, Dedrick, & Stershic, 2015). In these studies, it was more effective to space out practice on a topic than to mass practice all at once. For instance, in a study conducted in a physics classroom, students were randomly taught concepts on Hooke’s Law and String Laws with either a massed or a spaced model. Students using the spaced model earned significantly higher scores on a series of follow up tests (Grote, 1995). Distributed practice “works across
students of different ages, with a wide variety of materials, on the majority of standard laboratory measures, and over long delays” (Dunlosky et al., 2013). Further, spaced learning can be easy and inexpensive to implement, since current textbooks can be repurposed by thoughtfully spacing out problems that are printed in a massed fashion in the text.

Interleaved practice benefits students by requiring them to decide which procedure is the correct one for the problem at hand. In contrast, in block learning the student already knows the practice problems are all related to a single procedure (Rohrer & Taylor, 2007). This discrimination between problem types is both beneficial both for test taking and for real life, since problem types often arise in a mixed fashion. For instance, when a family doctor sees patients, she must discriminate between many possible diagnoses. In a study of college students learning about permutations, Rohrer and Taylor conducted two experiments that manipulated the timing and interleaving of practice sessions. Both spacing and interleaving produced large gains in test scores over a blocked schedule (2007). Interleaved practice provided “near immunity against forgetting” in a study of middle school students learning to graph lines. The test performance of the interleaved group decreased by only 6% over 30 days, as opposed to the blocked group, which decreased by 22% (Rohrer et al., 2015, p. 906).

There are some perceived downsides to interleaving practice. For instance, a student’s practice performance can be impaired by interleaving, though retention over time is often improved (Taylor & Rohrer, 2010). Unfortunately, students often don’t like interleaved learning, though it is good for them in the long run. Because working a group of interleaved problems is harder and more time consuming than working a blocked group, students may balk at the extra effort (Rohrer, 2012). Further, this extra effort might lead students and teachers to “falsely yet
reasonably conclude that interleaving is less effective than blocking... Among subjects who did benefit from interleaving, only 25% believed that interleaving was more helpful” (p. 365).

Despite the “relatively dramatic effect on students’ learning and retention of mathematical skills,” literature is still sparse with “enough null effects to raise concerns” (Dunlosky et al., 2013, p. 44). For instance, in a study in which students were learning to pronounce French words, all of the experiments favored blocked learning over interleaving (Carpenter & Mueller, 2013). The best solution may be a mixed practice schedule in which topics are first learned in a blocked format and then practiced in an interleaved format, though research is lacking on what combination of methods might lead to optimal learning (Carpenter & Mueller, 2013; Dunlosky et al., 2013; Rohrer, 2012; Rohrer et al., 2015). In addition to spaced and interleaved learning, there are other techniques that can increase learning and retention. One such technique is the use of mnemonics, which will be explored in the next section.

**Process Mnemonics**

Even when a student correctly learns a concept, it can be a challenge to retain that learning over time. Mnemonics, defined as a word, sentence, or pictorial device or technique for strengthening memory, can help (Lombardi & Butera, 1999). There are two basic types: fact mnemonics for remembering a specific piece of information, and process mnemonics used for remembering rules and procedures (Manalo, 2002). An example of a fact mnemonic is HOMES, in which each letter stands for one of the Great Lakes. A well-known spelling process mnemonic is “i before e, except after c, or when it sounds a, as in neighbor and weigh.” There is a large body of research supporting the use of mnemonics to increase student performance and retention (Irish, 2002; Lombardi & Butera, 1999; Maccini, Mulcahy, & Wilson, 2007; Manalo, 2002). This review focuses on process mnemonics due to its relevance to the current study.
While there is much research on mnemonics in general, there is little in the literature on using process mnemonics in math classes. One study analyzed the effectiveness of a technique for learning fractions nicknamed with the process mnemonic “LAP,” or “Look, Ask, Pick” (Test & Ellis, 2005). Six students, of which three were learning disabled (LD), were taught the LAP process to do computations on fractions with unlike denominators. Five of the six students achieved mastery on both strategy and performance steps and maintained scores of at least 80% six weeks after the intervention ended. In a study of 29 LD eighth-grade students learning computations on decimals, students in the process mnemonic group not only showed significantly higher initial learning, but also demonstrated the greatest retention eight weeks after the intervention’s end ($d = 2.22$) (Manalo, Bunnell, & Stillman, 2000). It is interesting that both these studies were conducted with learning disabled students. It is likely that techniques could benefit a much broader range of students by enhancing their thinking skills and long-term retention.

Some have criticized mnemonics as shielding students from true comprehension by encouraging them to simply memorize a rule (see the PEMDAS discussion above). While this can happen, the fault appears to lie in the teaching that occurs concurrently with the learning of the mnemonic rather than the mnemonic itself. Indeed, mnemonics can “make information more understandable by not only simplifying it but also focusing on salient and most important aspects… all these features point to mnemonics actually being facilitative of comprehension” (Manalo, 2002, p. 76). As a part of a comprehensive teaching strategy, mnemonics can be a valuable tool. As a further benefit, the teacher need only remind students of a previously taught mnemonic to activate retrieval and use of the underlying process (Lombardi & Butera, 1999). In the context of the CRKM (Dole & Sinatra, 1998), a process mnemonic could serve as a
peripheral cue to catch students’ attention and draw them into high engagement with the message.

Conclusion

Many students struggle in high school math classes because they have internalized many misconceptions in prior math classes. They never properly learned the full structure and meaning of the math they are studying. Instead of developing understanding, students memorize rules, but then do not know how or when to apply them. The literature documents students’ deep misunderstanding of the structure of algebra, including the meaning and use of variables, equality, exponents, negative numbers, and parentheses. The literature has shown that while misconceptions are tenacious and resistant to business as usual, there are ways to attack them using conceptual change theory, spaced and interleaved practice, and process mnemonics to enhance learning and long-term retention.

Implications for this research

This research was designed to address gaps and uncertainties in the literature. First, there is no suitable diagnostic instrument to target the specific misconceptions discussed above. Although there is a rich body of existing research that identifies, categorizes, and diagnoses students’ algebraic misconceptions, there are very few interventions designed to remediate such misconceptions set in math classrooms and even fewer that employ conceptual change models. Looking even more narrowly, there was only one study that used a warm conceptual change model to create cognitive conflict in mathematics students in a classroom setting. Finally, while the literature tends to favor interleaved learning over blocked learning, there is still debate over which is the better method and whether or not a mixed schedule of blocked followed by interleaved learning might be best.
The math education community shares a common frustration with the persistence of algebraic misconceptions, yet teachers continue down the same unsuccessful path decade after decade. Educators must do something different to break the cycle so students can achieve their full potential in math and STEM-related fields. One possibility is to design interventions that harness the power of conceptual change theory to remediate the algebraic misconceptions that hold students back. This study explores the effectiveness of such an intervention. If math educators do not find a way to solve the problem, the children of today’s students will continue to make the same mathematical mistakes in high school as their parents and grandparents made.
Chapter 3: Research Design

The purpose of this study was to evaluate the effectiveness of a warm conceptual change theory (WCCT) based intervention on remediating common algebraic misconceptions held by students at a private secondary school. The No-to-Yes (NTY) intervention focused on eight common misconception categories (MCs): negative function inputs, rules of exponents, distribution of the negative sign, distribution errors, simplification of algebraic fractions, rules of fractions, rules of negative exponents, and a mixed review. Because depth of engagement is a critical component of WCCT, student engagement in the intervention was also examined.

Research Questions and Hypotheses

As discussed in Chapter 1, the presence of students’ algebraic misconceptions was measured by the Misconception Error Score (MES), which is the total number of errors attributable to a misconception. Student engagement was measured by statistics obtained from records of activity in the intervention, such as number of correct practice quizzes, number of video views, and number of practice slips correctly completed. Engagement also varied at the class level since the two intervention group teachers engaged at different levels of intensity.

Four research questions were investigated. Research Questions 1 and 2 examined the differences between the intervention groups and the control group by overall result and by MC. Focusing on the intervention groups, Research Question 3 explored the relationship between engagement and gain in test scores. Finally, Research Question 4 explored how the participating students experienced the intervention. For all RQs, the subjects are students in Algebra 2 who attend a private secondary school in New England.
**Research question 1 (RQ1).** To what extent were there significant differences over time in the MES on a test of algebraic misconceptions among the control group and two intervention groups?

**Research question 2 (RQ2).** For each MC, to what extent were there significant differences in the change in MES among students in the control group and successful students in the intervention group, after excluding students who earned a “Yes!” on the pre-test for that MC?

**Research question 3 (RQ3).** To what extent did engagement predict the depth of conceptual change, as measured by the change in MES in the two intervention groups?

**Research question 4 (RQ4).** How did participating students experience the intervention?

**Research Design**

This study employed a quasi-experimental, pre-test post-test, control group design as well as student focus groups. Students in both the intervention and control groups took a pre-test to measure their baseline level of algebraic skills and demonstration of misconceptions. After the 8-week intervention, students in both groups took an identical post-test and the results were analyzed using two-way mixed ANOVA, Chi-Squared testing, and Multiple Linear Regression.

Quasi-experimental designs are used to establish causality when the ability to randomly assign subjects is limited. Experimental methods, including quasi-experimental, are the best method for establishing cause and effect relationships among variables (Creswell, 2011). Since this study sought to establish that the independent variable (the intervention) caused a change in the dependent variable (the post-test score), an experimental design was appropriate. However, since it was not possible to assign students randomly to groups, a quasi-experimental design was required. Quasi-experimental designs are intended to approximate the benefits of a pure experimental design when it is not possible to randomly assign subjects to groups.
On the positive side, quasi-experimental designs enjoy enhanced ecological validity (Muijs, 2011). Ecological validity refers to “whether or not one can generalize from observed behavior in the laboratory to the natural behavior in the world” (Schmuckler, 2001). Since this study was carried out in a natural classroom setting rather than a laboratory environment, there is increased confidence that the results have external validity, which implies they can be generalized to other classroom settings. This means other teachers and schools could implement the same No-to-Yes intervention with a reasonable expectation of seeing similar results.

**Threats to internal validity.** Unfortunately, quasi-experimental designs are not as effective at establishing causality as pure experimental designs due to the inability to control a large component of variability (Muijs, 2011). Since students cannot be assigned randomly to treatment groups, quasi-experimental designs are more vulnerable to threats to internal validity than true experimental designs (Creswell, 2011). This design weakness requires the researcher to pay strict attention to standardizing conditions, enhancing the fidelity of implementation, and closely monitoring how the intervention is being deployed (Muijs, 2011). Further, each of the two teachers participating in the intervention was considered a separate intervention group, explicitly acknowledging that the teacher was a source of potentially large internal variability.

There are many other threats that can jeopardize an experiment's internal validity if not controlled, such as the following (Fraenkel, Wallen, & Hyun, 2014):

**Subject characteristics threat.** Since it was not possible to randomly assign students, intact classes were assigned to control and intervention groups based on their teacher’s willingness to participate.

**Mortality threat.** Only on rare occasions does a student leave school during the winter term, perhaps due to illness or disciplinary action, and in those cases the loss is unavoidable and
unforeseeable. Of the 73 students who took the pre-test, six did not take the post-test due to illness or disciplinary action and hence were excluded from the analysis.

**Location threat.** The intervention was deployed in fundamentally identical math classrooms in the same building on campus, so this was not a concern.

**Instrumentation threat.** The validity of the researcher’s instrument is discussed below. Neither the instrument nor the scoring procedure was changed during the intervention, and a detailed rubric helped the researcher classify errors properly.

**Data collector characteristics.** The researcher did all grading and updating during the intervention, which minimized the possibility of different graders having inconsistent results. Further, objective data gathered from the online component was not subject to interpretation from the researcher.

**Data collector bias.** This threat was minimized by standardizing procedures and creating rubrics. Though code names were used to minimize bias, by the end of the intervention the researcher was familiar with the work from certain students, which may have introduced bias. The researcher also knew from which section the practice slips originated. The largely impersonal nature of the online component and practice slips helped minimize this threat. Further, each slip question was short and easy to grade, with little room for grader bias.

**Testing threat.** Given that the pre- and post- tests were identical, it is possible that students did better on the post-test merely by having taken it before. However, since students were not allowed to keep their graded pre-test, they could not refer back to it or study it. The post-test was given seven to eight weeks later, which should have been enough time for students to forget the precise pre-test questions. Due to the nature of the intervention, the students knew exactly what kinds of questions would be on the post-test, and hence gained familiarity during
the intervention. This could have led to improved scores. Instead of being a threat, this was the goal of the intervention.

**History threat.** Teachers in all groups were instructed to avoid discussing these seven MCs during class more than usual during the intervention, except when individual students asked specific questions. No unplanned or unanticipated events arose during the study.

**Maturation threat.** Some improvement in test results should be expected due to the passing of time and exposure to more algebra as part of regular class. However, since the misconceptions in this study have proven to be quite sticky over time (Bush & Karp, 2013; Cangelosi et al., 2013; Pitta-Pantazi et al., 2007), it was unlikely students would grow out of them over the time of the intervention. Any maturation effect should have been duplicated in both the control and experimental groups and been taken into account in the statistical analysis.

**Attitude of subjects threat (Hawthorne Effect).** Students in the intervention group may have performed better merely because something special was happening in class. However, it is unlikely that simply being part of a research study gave students the ability to overcome their misconceptions. Also, in general, students in the control group may have been resentful that they were not receiving special treatment and hence their performance may have suffered. However, in this case, the control group was more likely grateful they were not part of the intervention (hence doing more work) rather than resentful.

**Regression threat.** Students scoring at very high or very low levels on a pre-test are more likely to have post-test scores nearer the mean. It is plausible that students scoring poorly on the pre-test might have improved much more than students scoring very highly on the pre-test, who did not have much room to improve. With a range of misconception errors from 3 to 23 on the 28-question pre-test and the middle 50% of students making between 12 and 18
misconception errors, there were not many students in danger of hitting the testing ceiling. On the post-test, only one student made no errors at all.

Overall, these threats were minimized by careful planning, standardization of procedures and grading, and careful monitoring.

**Population and Sampling**

**Research site.** The research site for this study was the private secondary school in New England where the researcher is a mathematics teacher. This selective school enrolls approximately 500 boarding students. The site was chosen not only for convenience, but also because it allows the researcher to continue to improve her own practice. The school granted approval to the researcher to conduct the study in Algebra 2 classes.

**Research participants.** The research participants were two teachers of non-Honors Algebra 2 during the winter term of the 2016-2017 school year. Honors Algebra 2 classes were excluded from this research because pilot studies have shown their pre-test scores to be significantly higher, indicating they have fewer misconceptions than the non-Honors students. None of the students were in the researcher’s own classes. Two teachers (Mr. G and Mr. M) volunteered to deploy the intervention in their classes while two other teachers served as the control group. Mr. G was a first-year high school math teacher who began teaching in 2016 after a career in engineering, marketing, and corporate training. Mr. M has been teaching at this school and teaching this Algebra 2 course for three years. He participated in the pilot program in past years. Previously, he taught science at the university level.

Of the 73 students who took the pre-test, six did not take the post-test due to absence, illness, or other reasons, resulting in 67 students who took both the pre- and post-test. The control group initially consisted of 30 students in four separate classes taught by two teachers.
Four of these students were absent when the post-test was administered, so they were excluded from the study, leaving 26 students in the control group. The two control group teachers administered the pre- and post-test approximately eight weeks apart with no other change to their normal classroom routine.

The intervention group initially consisted of 43 students in five separate classes (three classes taught by Mr. M and two by Mr. G). Two of these students were absent when the post-test was administered, so they were excluded from the study, leaving 41 students in the intervention group. During the intervention, Mr. G and his students were much more engaged in the intervention than Mr. M, which confirmed the need to analyze data from the two teachers separately. Mr. G’s intervention group, named the “Full” intervention group, consisted of 20 students in two classes. Mr. M’s intervention group, named the “Partial” intervention group, had 21 students in three classes.

Of the 67 students in the study, 31% \((n = 21)\) were freshmen, 55% \((n = 37)\) were sophomores, and 14% \((n = 9)\) were juniors. The gender split was 48% \((n = 32)\) male and 52% \((n = 35)\) female. There were 57% \((n = 38)\) students new to the school this year, while 43% \((n = 29)\) had been at the school for at least one prior school year. No other demographic data was collected.

It was important to know if the groups were approximately equivalent in size, demographics, and incoming algebra skills before the study began. Demographics were collected through the schools’ learning management system while students’ baseline algebra skills were compared by examining performance on two common assessments: an algebra diagnostic test taken in September 2016 and the fall term test taken in November 2016.
The three groups were approximately equal in size, with 20 students in the Full Group, 21 in the Partial Group, and 26 in the Control Group. As shown in Figure 2, the demographic split among the three groups was relatively uniform with two exceptions. First, the Full Group had no juniors, who were instead found in the Partial Group (44%, n = 4) and the Control Group (56%, n = 5). The Control Group had the smallest proportion of freshmen (19%, n = 4) compared to the Full Group (43%, n = 9) and Partial Group (38%, n = 8).

![Figure 2: Distribution of grade levels across the treatment groups.](image)

Since juniors in Algebra 2 historically tend to be weaker math students while freshman tend to be stronger math students, the distribution of age appears to favors the Full Group. However, results from the diagnostic exam show that while juniors (M = 14.25) did score lower than freshmen (M = 18.65) and sophomores (M = 17.19), the difference was not substantial. This may be due to the higher variability in juniors’ scores, as seen in Figure 3.
The second demographic difference was that the 20 students in the Full Group were predominantly females (70% female, 30% male), while the 21 students in the Partial Group were predominantly males (62% males, 38% females). The 26 students in the Control Group were evenly split by gender. Given that males earned a mean score of 16.73 on the diagnostic compared to the girls mean score of 17.98, there does not appear to be a gender advantage.

Looking at the three groups overall, results were almost identical on the diagnostic test, with median score of 18 for all groups (see Figure 4). The mean scores (out of 40) were 17.95 in the Full Group, 17.42 in the Partial Group, and 16.85 in the Control Group.
The fall term exam scores differed by five points among the groups, with mean scores (out of 100%) of 82% for the Full Group, 87% for the Partial Group, and 82% for the Control Group (see Figure 5).

![Box plot showing fall term test scores across treatment groups.]

**Figure 5:** Distribution of fall term test scores across treatment groups.

Overall, since the three groups were relatively similar across demographics and prior assessment results, differences in the post-test results among groups could be plausibly attributed to the intervention.

**Description of the No-to-Yes Intervention**

The No-to-Yes (NTY) intervention targeted a set of pervasive misconceptions through testing, online work, and class practice. Overall, the goal was for students to decrease the number of misconception errors and earn a “Yes!” in eight misconception categories (MCs). For each MC, students earned a “Yes!” in one of two ways: by doing well on the pre-test or by completing the corresponding module. To complete a module, the student first went online to watch a video and earn 100% on a practice quiz and then had to complete a series of practice slips in class. This section describes each of these components of the intervention in detail, and complete student instructions are given in Appendix D.
**Misconception categories.** The MCs included in the intervention originated from both the review of the literature and from common problems identified by math teachers at the researcher’s school. The MCs were:

1. Negative function inputs
2. Rules of exponents
3. Distribution of the negative sign
4. Distribution errors
5. Random canceling (simplification of algebraic fractions)
6. Rules of fractions
7. Rules of negative exponents
8. Mixed Review (paper slips only)

As will be discussed in Chapter 4, students found some of these categories easier (such as MC1), while some proved far more difficult (such as MC5 and MC8).

**No-to-Yes intervention components.** The NTY intervention consisted of four main components and a class wall chart that displayed each student’s status. The first component was a diagnostic pre-test designed to identify which misconceptions a student demonstrated. The second was an online course with a video and practice quiz for each MC. The third component was a set of in-class paper practice slips for each MC designed to help students practice newly changed conceptions. Finally, the fourth component was a post-test (identical to the pre-test) to show any changes that occurred over the eight-week intervention. Grade incentives helped motivate students to engage in the intervention.

Spaced and interleaved practice was incorporated in the online course and practice slips to leverage the many benefits discussed in the literature review. Practice was interleaved in the
online quizzes, each paper practice slip, and more broadly in the 8th misconception category (mixed review). On each quiz and slip, students needed to discriminate between types of problems to determine which technique to use before executing that technique. Further, students experienced the benefits of distributed learning as they completed practice slips over time, with a minimum of three class days required to earn a “Yes!” in any MC.

**The No-to-Yes test.** Since open-ended questions give insight into student thinking, the NTY pre- and post-test in Appendix E consisted of 28 short, free response questions across seven MCs, with four problems in each category similar to tests created by Poon & Leung (2010). The problems were presented to the student in mixed order. All questions were carefully targeted to elicit a misconception response with minimal opportunity to make other careless errors such as addition. Each question was graded as correct, misconception error, or other error. Since students were given ample time to complete the pre-test, blanks were counted as misconception errors since they indicated the student did not know what to do. While it was not always possible to distinguish between the types of errors, there were common errors that typified particular misconceptions. Table 1 shows the question, correct answer, misconception answer, and possible other answer for two of the diagnostic questions. Appendix F contains instructions for administering the test and Appendix G contains a sample of the grading rubric.
Table 1

**Sample Misconception Questions and Answers**

<table>
<thead>
<tr>
<th>Question</th>
<th>Question 1</th>
<th>Question 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simplify, if possible: $(3 + x)^2$</td>
<td>Simplify, if possible: $\frac{2x+6}{2x-10}$</td>
<td></td>
</tr>
<tr>
<td>4. Distribution Errors</td>
<td>5. Random Canceling</td>
<td></td>
</tr>
<tr>
<td>Correct Answer</td>
<td>$9 + 6x + x^2$</td>
<td>$\frac{x + 3}{x - 5}$</td>
</tr>
<tr>
<td>Misconception Error</td>
<td>$9 + x^2$</td>
<td>$-\frac{6}{10} - \frac{3}{5} \frac{x + 6}{x - 10}$</td>
</tr>
<tr>
<td>Other Error</td>
<td>$9 + 3x + x^2$</td>
<td>$\frac{x + 2}{x - 5}$</td>
</tr>
</tbody>
</table>

Students received two grades on the pre-test: a Misconception Error Score (MES) that gave the total number of misconception errors, and a Yes Earned Score (YES) that gave the number of “Yes!” indicators earned. To earn a “Yes!” in an MC, the student must have correctly answered all four questions related to that MC. If a student earned a “Yes!” for an MC on the pre-test, no further work was required in that category. Otherwise, the student needed to do more work in that MC until a “Yes!” was earned, as described below. The number of correct answers was not analyzed, since the focus of the research was on reducing errors due to misconceptions and not due to other random errors like adding improperly.

**Online course.** Each student in an intervention group was enrolled in an online NTY Project course created in Canvas. Students currently use the Canvas learning management system (“Learning Management System | LMS | Canvas by Instructure,” n.d.) for each of their classes, so they were already familiar with the platform. For each of the first seven MCs, the researcher created a video and a practice quiz organized in modules on Canvas.
**Online videos.** The researcher created animated videos using PowToon (“PowToon | Create Animated Videos for Work or Play,” n.d.) to engage students and Edpuzzle (“EDpuzzle,” n.d.) to embed pop-up questions. The 3-5 minute videos incorporated augmented activation and refutational text in order to create the cognitive conflict that leads to conceptual change (C. R. Hynd, 2001). To do this, each video started with a triggering problem for the student to solve which was intended to activate their (mis)conception. After giving evidence for the correct conception, the video clearly warned students to pay close attention to the ideas that follow. The video then explicitly addressed what many students think and explained why that thinking is incorrect. Each video then continued by giving additional examples, teaching correct conceptions, and summarizing the ideas learned (see Appendix H for video screen shots).

To increase engagement and discourage students from walking away while a video was playing, each video had 3-5 embedded pop-up questions like the one shown in Figure 6. Edpuzzle prohibited students from skipping ahead through the video and tracked how many times a student watched each segment of a video, how many minutes were spent watching, and the answers given to the embedded questions. Unfortunately, once the link was clicked, Canvas did not prevent students from skipping past the video to the practice quiz, so it was possible for students to avoid the conceptual change based lesson.

![Figure 6: Example of an embedded Edpuzzle pop-up question.](image-url)
Additionally, as discussed in the literature review on process mnemonics, each video incorporated one or more “brands” or logos designed to trigger the recall of previously learned procedures. Appendix I lists all the brands developed for the videos. Some of these brands, such as “Plug In!,” were reused in multiple videos. Once a student learned these brands, a teacher could use them at any time in class to trigger recall of the ideas taught in the videos.

**Online quizzes.** Each MC also had a four-problem multiple-choice practice quiz that drew from a test bank of similar items. For each question, there were four variations in the test bank, each consisting of a correct answer and 1-3 misconception answers to serve as distractors (Russell et al., 2009). After completing a quiz, the student received immediate feedback that either reinforced the correct choice or offered tutoring on an incorrect choice (see Appendix J). Canvas tracked the grade for each quiz attempt and how long a student spent taking each quiz.

Students were encouraged to watch the videos and take the practice quizzes as many times as they wanted. However, in order to move on to the paper practice slips in class, a student had to “unlock” an MC by watching its video and scoring 100% on the related practice quiz.

**Practice slips.** For each MC, there were four different versions of a paper practice slip consisting of variations of the problems practiced online (see Appendix K). The slips were organized in each classroom in boxes sorted by MC and version number (see Figure 7). Every day in class, a student could attempt one practice slip per unlocked MC, completing as many as class time permitted. The students left their completed practice slips in a designated box for the researcher to grade. If the student correctly answered all the problems on the slip, the researcher marked the slip with a cheerful “Yes!” (see Figure 8). If there were errors, the researcher wrote brief feedback to the student on the slip. The researcher then copied the graded slips and returned them to their designated box for the students to pick up the next day. Once a student
completed three different versions of the slip on three different days, she earned a “Yes!” for the MC. Once a student earned a “Yes!” on the first seven MCs, the 8th MC was unlocked and students could work on the related practice slips. The intervention ended when the student earned a “Yes!” in all eight MCs or the eight-week period ended.

Figure 7: Practice slips organized into boxes.

Figure 8: Sample of a graded practice slip.
Wall chart. A wall chart in each classroom displayed, by student code name, the status of each MC for each student (see Appendix L). The wall chart was updated daily to show the exact status of each student. For each MC the student earned on the pre-test, the chart was initially marked with a green “Yes!” The rest of the cells were covered with orange stickers that signified the student needed to do work online. Once a student unlocked an MC online, the orange sticker was removed. Next, as the student completed practice slips in class, successful attempts were noted on the class wall chart with date and slip version, and a “Yes!” indicator was added after three successful attempts. After a student earned the first seven MCs, the orange slip from the 8th MC was removed. Finally, if a student then earned a “Yes!” on the 8th MC, thus finishing the intervention, her name was highlighted in pink on the chart. The publicly displayed chart added a social dimension that might have increased motivation and competition among students and hence enhanced conceptual change (Dole & Sinatra, 1998), but there was also a risk that it discouraged those that were further behind.

Post-test. At the end of the intervention, all students in the control and intervention groups were given a post-test identical to the pre-test. The researcher graded all the post-tests for consistency.

Student incentives. To motivate students as well as offer a reward for the amount of work involved, students could earn extra credit points for making progress on the intervention. Students received one bonus point on a quiz for each MC they unlocked and one point for each “Yes!” they earned, for up to 16 points on a previously graded quiz. In addition, if the student completed all eight MCs, they scored 100% on a new bonus quiz. The estimated impact of completing the entire intervention was approximately 0.5% of their final grade for the term.
Sampling plan. While the true population of interest was the entire cohort of non-Honors Algebra 2 students across the nation, given the inability to access these students, this study did not sample from the population but instead used a convenience sample from the researcher’s school. Results may not generalize well to the broader population, since students at this school may be different than the general population of students in other schools. However, in reviewing the literature and from hearing anecdotal reports from other math teachers, it appears the students at this school commit the same fundamental math errors as students in many other schools.

The sample size of 67 students in the control and intervention groups should have been sufficient if the effect size was at least medium size. A preliminary power analysis suggested that for a repeated measures ANOVA with 3 levels of the dependent variable, a medium effect size, alpha of 0.05, a power of 0.80, the desired total sample size was 28 (Intellectus Statistics, 2013).

Since it was not possible to randomly assign students to classes and the intervention was to be given to all of a teacher’s classes, teachers (and hence their students) were assigned to treatment groups. As part of the recruitment process, the researcher asked each Algebra 2 teacher about their willingness to participate in the study, implement the protocols with a high degree of fidelity, and provide classroom time for a focus group at the end of the intervention. The two teachers who adopted the intervention as classroom practice had all their sections of Algebra 2 participate. While this was not ideal from a statistical perspective, since the effect of teacher on the results was confounded with the intervention, it ensured that a potentially beneficial treatment was not withheld from any students of the intervention teachers. Two other teachers agreed to be in the control group and to administer the pre- and post-tests.
**Information collected.** A variety of data was collected from the participants. Demographic information, diagnostic assessment results, fall term test scores, and pre- and post-test scores were collected from all students. From the intervention groups, data was also collected from the online and class portions of the intervention and from focus groups.

**Research relationship.** The researcher was an occasional visitor to each classroom who introduced the intervention and audited implementation fidelity. The researcher graded all pre- and post-tests as well as the practice slips, using student code names to avoid bias and maintain confidentiality. Outside of mathematics, the researcher was the adviser of one student, Head of House for four students, and squash coach for one student. These varied relationships did not provide any other sources of information for the study. None of the students were in the researcher’s own classes. Mr. M is a colleague and friend of the researcher, as well as the third reader of this dissertation. Mr. G is the researcher’s husband and colleague.

**Data Collection**

Data was collected to measure the presence of algebraic misconceptions before and after the intervention, to validate the pre-test instrument, and to assess the level of student engagement. The following data (and data type) was collected for each student:

In the pilot group: Pre-test score (ratio) plus informal qualitative feedback

In both the control and intervention groups:

- Pre- and post–test results, recording whether each question was correct (C), a misconception error (M) or other error (O) (Categorical)
- Demographics, including Gender, New/Returning, Teacher, Section (Nominal); Grade Level (Ordinal)
- Fall Term Exam Grade (0 to 100) (Ratio)
• Fall Diagnostic Assessment Grade (0 to 40) (Ratio)

In the experimental group only:

• Intervention statistics by MC, including the number of practice slips completed correctly, number of video segments watched, and number of practice quizzes completed correctly (ratio)

• Audio recordings of focus groups with students at the end of the intervention

**Instrument.** The instrument in this study was the researcher designed NTY test that was used as both a pre- and post-test for all students. While much research exists on algebraic misconceptions in general, there are very few diagnostic instruments, and none that directly target ubiquitous problems like distribution of the negative sign and random canceling. This instrument was designed to fill this gap by tightly targeting selected misconceptions.

**NTY pre- and post-testing.** In order to answer research questions RQ1 and RQ2, the NTY pre- and post-test instrument was used to assess the presence of algebraic misconceptions. A comparison of performance on the identical pre- and post-test indicated if the intervention had been successful in reducing the number of misconception errors and increasing students’ mastery of each MC.

**NTY test validity.** Validity, which asks if the study is measuring what it is intended to measure, is “probably the single most important aspect of the design of any measurement instrument in educational research” (Muijs, 2011, p. 57). To assess the validity of an instrument, content validity, criterion validity, and construct validity must be addressed. All three types must be present to demonstrate overall validity.

**Content validity.** An instrument with content validity measures what it is intended to measure (Fraenkel et al., 2014). The NTY test has been piloted over the past three years and
continually revised, with extensive review from the other Algebra 2 teachers to ensure the content validity of the questions. For this study, a group of math teachers from the researcher’s school reviewed the test and assessed whether or not each question tested for the presence of the misconception it was intended to test. This panel also assessed the clarity of formatting and instructions. Revisions were incorporated into the test and the group again reviewed it. The test was then piloted with an excluded group of the researcher’s students. They were asked to comment on clarity of the test’s design and instructions as well as the face validity, or whether the test looked valid to them (Muijs, 2011). Based on their feedback, some items were reworded, some made easier to allow the test to be taken in a reasonable period of time, and some modified to more accurately target a particular misconception.

*Criterion validity.* For a test to have criterion validity, it should be related to other measures of the same construct (Muijs, 2011). For this test of algebraic misconceptions, it was reasonable to expect that other measures of students’ algebraic skills might show similar results. The closest related measure to show concurrent validity was the student’s grade on a diagnostic exam that was given in September 2016 to all students entering Algebra 2. General success on the diagnostic exam was plausibly related to how many misconceptions the student had, though the content on the diagnostic covered a broader range of topics than the NTY pre-test.

To evaluate criterion validity, a Pearson’s product-moment correlation was run to assess the relationship between the diagnostic exam and pre-test score for the 62 students who took both exams. Preliminary analysis showed the relationship to be linear with one outlier that was kept in the analysis. Both variables were normally distributed, as assessed by Shapiro-Wilk’s test ($p > .05$). There was a moderate negative correlation between the diagnostic exam and pre-test score, $r = -0.585$, $p < .0001$, which means that students who scored higher on the diagnostic
tended to have fewer misconception errors (see Table 2). This supports the criterion validity of the pre-test.

Table 2

Correlation Between Diagnostic Score and Misconception Errors on the Pre-test

<table>
<thead>
<tr>
<th># Misconception Errors Pre-Test</th>
<th>Pearson Correlation</th>
<th>Significance (2-tailed)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-.585**</td>
<td>.000</td>
<td>62</td>
</tr>
</tbody>
</table>

**Note:** Correlation is significant at the 0.001 level (2-tailed).

**Construct validity.** Construct validity is the broadest measure of validity, drawing together multiple sources of evidence to support overall validity (Fraenkel et al., 2014). Construct validity “refers to the degree to which inferences can legitimately be made from the operationalizations in your study to the theoretical constructs on which those operationalizations were based” (“Construct Validity,” n.d.). In other words, construct validity means the NTY test measures what it claims to measure, namely the presence of algebraic misconceptions. Given the content validity affirmed by a panel of teachers and a pilot group of students, high face validity, and the strong correlation with a diagnostic test, it is reasonable to assume an acceptable level of construct validity.

**MC test reliability.** Reliability refers to the consistency of test scores obtained, both for one individual on two different versions of the assessment and at two different times. Inter-rater reliability was not relevant to this study, as no repeated observations were taken and only one person completing all the grading. Ideally, both repeated measures and internal consistency should be evaluated for an instrument to assess reliability (Muijs, 2011). However, given the design of the study, it was difficult to assess these types of reliability rigorously. Because
revisions were made to the pre-test after the first pilot group took it, it was not possible to re-test with the same pilot group. Further, while the intervention students took the test twice (pre- and post-test), the goal was that their grade would improve. Hence this repeated testing could not be used to assess reliability.

Internal consistency. The only measure of internal consistency that could have been obtained was a student’s propensity to answer all questions in the same category correctly or incorrectly. However, since there were both easier and harder questions in each MC, it was not expected that a student who answered one question correctly would answer all four questions in a category correctly. This was not a fault in internal consistency, but instead was related to how deeply the student knew the material.

Procedures

This section will discuss the timeline of events leading up to and through the intervention. In addition, the detailed procedures that ensured both implementation fidelity among teacher participants and reproducibility for future researchers are included. Supporting documentation is included in the appendices.

Timeline. During the fall term (September – November 2016), the researcher finalized the proposed test, intervention, tutorials, protocols, and procedures. The proposal was defended and approved by the IRB. The intervention took place in the winter term (November 2016 – March 2017) as detailed in Table 3.
Table 3

*NTY Intervention Timeline*

<table>
<thead>
<tr>
<th>When</th>
<th>What/Who</th>
</tr>
</thead>
</table>
| Fall term (September – November 2016) | - A panel of teachers reviewed the intervention materials.  
- The pre-test was piloted with a group of excluded students and results used to edit the test and finalize the grading rubric.  
- Algebra 2 teachers were recruited and assigned to groups.  
- Teachers in the intervention groups were trained on intervention protocols. |
| Week 2 of winter term (week of December 5) | - All non-Honors Algebra 2 students took the pre-test.  
- Demographic, diagnostic, and fall term exam grades were collected. |
| Week 4 of winter term (week of Jan 8) | - Intervention students enrolled in the online course.  
- The intervention was kicked off in each classroom. |
| Weeks 4-10 of winter term | - Daily procedures were in progress, including checking for “unlocked” MCs, grading slips, and updating wall charts. |
| Week 11 of winter term (week of Feb 26) | - The post-test was given to all groups.  
- Focus groups were held in intervention classes.  
- Intervention teachers were interviewed.  
- Students received their incentive grade. |
| Spring Term 2017 | - Though not part of this study, one control group teacher implemented the intervention and one intervention teacher did the project again, using the post-test as a new pre-test. |

**Control group procedures.** The control group’s participation in the intervention was limited to taking the pre- and post-test. The majority of class time for both groups of students was spent learning the normal Algebra 2 curriculum, with only 5-10 minutes of the intervention group’s daily class time differing from the control group.

**Detailed procedures.** It was vital that implementation fidelity be maintained across all teachers and sections in order to have valid results. This section lists detailed instructions for intervention teachers and the researcher.
**Administering the pre-test.** The researcher distributed pre-tests to each teacher. Each teacher was instructed to distribute the pre-tests and read the printed instructions aloud. Each teacher collected the completed pre-tests and gave to the researcher for grading.

**Grading the pre-test.** The researcher graded each pre-test according to the established rubric. Any ambiguous cases were used to update the rubric. Once graded, the pre-test data was entered in Excel and SPSS for analysis and the tests were stored.

**Preparing for the intervention.** The researcher prepared for the intervention as follows:

1. Assigned code names to students in the intervention groups. Once code names were assigned and added to the Excel sheet, the column listing actual names was hidden. The researcher and students used only code names during the study.

2. Prepared and hung status wall charts for each class that listed student code name and status of each MC. Appendix L contains a sample wall chart as it might have looked in the middle of the intervention.

3. Copied, cut up, and distributed practice slips to all experimental classrooms using a similar organizational system in each class.

4. Enrolled intervention group students in the NTY Canvas course.

**Rolling out the intervention.** The researcher attended each intervention group class to introduce the intervention, working from a prepared script that included information about the procedures and the bonus grade they would receive for completing the intervention. Students were allowed to briefly review their graded pre-test along with their score, after which the pre-tests were collected. The researcher gave students their code names.

The researcher explained the wall chart, demonstrated the online course by projecting it live, and demonstrated the practice slips and how they work. The students then used their own
computers or a school computer to enter the Canvas course, watch a video, and locate the practice quizzes. They were instructed that they must first watch the related video and successfully complete the associated online quiz before unlocking the practice slips for an MC and that they could watch the video and complete online quizzes as many times as they want. They were assigned homework to “unlock” an MC by watching the video and earning 100% on the associated quiz. Students were instructed to ask questions about the intervention only to their teacher or the researcher. This introduction took about 30 minutes.

**Daily procedures during the intervention.** The researcher checked the Canvas course each morning to see if students had unlocked new MCs and updated the wall charts accordingly. Nearly every class day, except for assessment days, students worked online and/or retrieved graded practice slips and completed new ones. Mr. G directed his students to work on slips during the first 5-10 minutes of class, while Mr. M offered students time to work on slips at the end of class. After school each day, the researcher collected all slips, graded and copied them, updated class wall charts with the results, and returned the slips to the appropriate classroom.

**Monitoring student progress.** Students who were not making reasonable progress, defined as completing at least one MC each week, received a weekly reminder email. Teachers also encouraged students to work on slips in class, especially those who were not making sufficient progress.

**Terminating the intervention.** At the end of eight-week intervention, the researcher distributed post-tests to control group and intervention group teachers to give in class. Each teacher collected completed post-tests and gave them to the researcher to grade. Teachers did not inform the students of the post-test date so that students could not attempt to study for the test. The intervention groups also participated in a focus group run by the researcher that was
recorded and transcribed by Rev.com (see Appendix C for the protocol). Finally, the researcher conducted recorded interviews with the two intervention teachers.

**Threats to validity.** In addition to the threat from lack of implementation fidelity, or not following the protocols exactly as written, the following may have been threats to the validity of the study associated with data collection procedures.

*Online Canvas data.* Students were not monitored while using Canvas away from class. While Canvas recorded how long a student had a quiz open, it is possible a student might have walked away and returned later. Similarly, while the EdPuzzle application recorded how many times students watched each segment of a video, there is also the possibility that students walked away from a video while it was playing. To reduce this risk, pop-up questions were built into the video that had to be answered while viewing. Finally, given that Canvas was accessed as homework and not in class, it was not possible to know if the student actually answered his or her own quiz questions. Students were also able to click past the videos in Canvas and go directly to the quizzes. Overall, the Canvas data can only be used cautiously to assess engagement.

*Delayed testing variability.* Students completed MCs and the overall intervention at varying times. For instance, Student A might have completed MC1 in week 3 and the entire intervention in week 6, while Student B completed MC1 in week 6 and the entire intervention in week 7. They both took the post-test in week 8. It is possible that Student B might score higher on MC1 than Student A on the post-test because he finished it more recently.

**Data Analysis**

In this section, the plan for analyzing the data will be discussed, including rationale for each choice of statistical test and assumptions regarding those tests. For all statistical tests, an alpha level of 0.05 was used to determine statistical significance and all p-values were reported.
**Preparation of data file.** All quantitative data was entered in an Excel spreadsheet for future analysis in IBM’s SPSS Version 24. For all variables, frequencies were generated to look for data entry errors and missing data that was then fixed or re-coded properly. Outliers and transformations of the data required to meet the requirements of a particular statistical test are discussed in Chapter 4.

**Transformation of raw data.** Totals and composite variables were calculated as follows:

For pre- and post-tests, totals were created for each student overall and in each MC. Specifically, for each student and each MC, the number of misconception errors was totaled as well as whether or not each student earned a “Yes!” for that MC. Further, overall scores were created for each student by combining the individual MC numbers into the total Misconception Error Score (MES).

To create intervention engagement statistics, for each student in each MC, the following totals were calculated: the total number of practice slips completed correctly (TotalYesSlips), total amount of video usage (TotalVideo), and total number of online quizzes completed correctly (Total4Quiz). Additionally, a new variable (TotalEngage) combined these individual measures to serve as a single measure of engagement.

**Data analysis for RQ1.** RQ1, which investigated the overall effectiveness of the intervention by looking at pre- and post-test results, was analyzed using a repeated measures ANOVA. The purpose of this test is to “test the differences between two or more independent groups while subjects are repeatedly measured on some dependent variable in each level of the within-subjects factor” (Verma, 2015, p. 126). Not only can this type of ANOVA measure the main effects of the independent variables, but it can also measure the interaction between them.
This test is appropriate for a continuous dependent variable and two categorical independent variables or factors (Laerd Statistics, 2015d). The main advantage of this test is that it allows for investigation of the interaction between the within-subjects variable and the between-subjects variable.

In this study, the dependent variable was the MES. The between-subjects independent variable was group membership (Full, Partial, Control), while the within-subjects independent variable was time (pre-test, post-test). This test showed if MES changed over time for each group, which was likely since students enrolled in an Algebra 2 class should have been improving their skills and decreasing their MES even without an intervention. It also determined if MES changed with group membership (at each time), which highlighted the difference among the groups’ scores at pre-test or at post-test. Finally, the interaction term between time and group showed if the intervention caused the MES to change differently for the three groups over time. The last interaction was of crucial importance in this study.

Before running the model, a number of assumptions were checked as described in Chapter 4, which included checking for the presence of outliers, normality for every combination of time and group, and homogeneity of variances. Once the assumptions were checked, the results were interpreted.

**Data analysis for RQ2.** RQ2 examined the change in test scores for each MC individually. At the MC level, the range of possible MES was 0-4 on both the pre-test and post-test, resulting in a change score that could range from -8 to 8, though typical values ranged more from -3 to 1. As before, a negative change score was a positive result indicating that fewer errors were made on the post-test than on the pre-test. Because the dependent variable (MESchangeMCx) was ordinal, a parametric test such as ANOVA or t-test could not be used, so
the Mann-Whitney U test was chosen. This is a rank-based, nonparametric test used to determine if there are differences between two distributions (both location and shape) when the dependent variable is ordinal (Laerd Statistics, 2015a). It is often used as an alternative to parametric tests such as a t-test when data is not normally distributed or the dependent variable is not continuous and tests whether “one variable tends to have higher values than the other” (Hart, 2001, p. 391). A series of seven Mann-Whitney U tests were run, one for each MC.

Since the purpose of RQ2 was to determine the effect of the intervention among students who needed to change their misconceptions, students who had already demonstrated their mastery of the MC by earning a “Yes!” on the pre-test were excluded. Next, all remaining control group students were included, along with all intervention group students who successfully earned a “Yes!” during the intervention by completing the online component and in-class slips. Given these criteria, the number of students included in the test was different for each MC.

The only requirement to use a Mann-Whitney U test is that the observations must be independent. Since a student in the control group could not also be in an intervention group, this assumption was met. To be conservative, for all MCs, the assumption was also made that the distributions were not similar, hence mean ranks were compared instead of medians.

Data analysis for RQ3. RQ3 asked about the relationship between engagement and the change in MES, which is the difference between misconception error pre- and post-test score. The hypothesis that greater overall engagement would lead to greater conceptual change was first examined using a measure of total engagement that combined the three different ways students could engage in the intervention: videos, online quizzes, and in-class practice slips.
Next, the hypothesis that the three individual ways of engaging could have different relative impacts on the change in MES was tested.

**Video engagement.** The variable TotalVideo was created to quantify how much each student engaged in the online videos. The variable consisted of two parts. First, the number of video segments watched (TotalVidViews) was totaled per student. Each of the seven videos was divided into ten segments, and students could watch each segment more than once. For example, if a student watched an entire video once and then re-watched two segments, her TotalVidView was 1.2. Second, the number of embedded video questions correctly answered (TotalVidQs) was totaled. The number of correct answers was a better measure of engagement than questions attempted, since students might have clicked randomly without engaging just to get through the video. TotalVidView and TotalVidQ were highly correlated ($r = .844, p < .001$). Therefore, since it was determined they were measuring roughly the same thing, a new variable (TotalVideo) was created to combine them so that each variable was weighted equally (twice TotalVidView + TotalVidQ). This new variable was used as an independent variable.

**Online quiz engagement.** The variable Total4Quiz was created to quantify how much each student engaged in the online quizzes. There were seven online quizzes, one for each MC, which could be taken as many times as needed to earn 100% (or more frequently if desired as practice). The number of quizzes each student got completely correct (a score of 4 out of 4) was totaled. Correct quizzes were used instead of total quizzes attempted, since a student who attempted many quizzes may have been active in the intervention but engaging very superficially and perhaps guessing randomly. For example, one student in the Partial Group (Gauss) took the online quiz for MC5 eleven times before successfully earning a grade of four. It appears he was superficially engaged and simply hoping to get lucky on one of his multiple attempts.
**Practice slip engagement.** The variable TotalYesSlips was created to quantify how much each student engaged with the in-class practice slips by totaling the number of slips each student got completely correct (a score of 4 out of 4). As with the videos, correct slips were used instead of total slips attempted, since a student who attempted many slips may have been active in the intervention but engaging very superficially. For example, one student in the Full Group (Einstein) attempted ten slips for MC2 but only completed one correctly. Based on teacher observation, he appeared to be simply doing the same thing repeatedly without taking the time to engage and remediate his misconception. Completing a slip correctly was a more direct sign of successful deep engagement with the work.

**Total engagement.** Finally, a single measure of engagement for each student (TotalEngage) was computed by assigning weights to each of the three previously discussed engagement variables so that they would have equal weight in the combined measure. The formula was TotalEngage = TotalVideo + 4.7(Total4Quiz) + 2(TotalYesSlips).

**Regression modeling.** Given that warm conceptual change theory posits that engagement causes conceptual change, the most appropriate test for continuous variables was linear regression. To answer the research question, two models were run, the first using TotalEngage as the only independent variable and a second using the three separate measures of engagement as independent variables. Change in MES was the dependent variable in both models.

The simple linear regression and multiple linear regression tests were run after checking the assumptions for each test. Regression requires checking four assumptions (Laerd Statistics, 2015b) as detailed in Appendix N. These assumptions were independence of residuals, homoscedasticity of residuals, normal distribution of errors, and linearity between the independent and dependent variables. The presence of outliers or influential points was also
examined. The overall $p$-value and the $p$-values of the independent variables were assessed to see whether or not the model was significant as a whole and which independent variables were most causal, respectively. See Chapter 4 for more details.

**Data analysis for RQ4.** RQ4 sought to qualitatively understand the experiences of students and teachers in the intervention. The focus groups conducted with each class in the Full and Partial Groups were recorded and transcribed by Rev.com. Data analysis then consisted of reading through the transcripts several times and coding the data using MAXQDA 12. The initial round of coding was largely informed by focus group questions such as what the student liked or didn’t like about the intervention. However, once the initial coding was complete, themes such as time pressure began to emerge. The data was then re-coded to reflect these emerging themes. Overall, this process was inductive and cyclical in nature, starting from detailed data and ending with generalized trends or themes (Creswell, 2011).

**Validity, Reliability and Generalizability**

Taken together, strong validity, reliability, and generalizability ensure a study consistently measures what it intends to measure in a way that is useful to a broad audience. As with any experiment, this research faced threats in all three dimensions, which were identified and mitigated when possible. Full disclosure in this written report will allow the reader to judge the study on its merits and decide how to interpret the results.

Validity refers to “the appropriateness, meaningfulness, correctness, and usefulness of the inferences a researcher makes” based on the data collected from a particular instrument (Fraenkel et al., 2014, p. 148). The content, construct, and criterion validity of the researcher written NTY test has been discussed and the threats to internal validity of the study explored in earlier sections. One of the biggest threats was the lack of randomization inherent in quasi-
experimental design, which was addressed through thoughtful assignment of groups and analysis of the relative strength of each group at the outset. Implementation fidelity was another strong threat, which was addressed through standardization in protocols, rubrics, and training.

Reliability refers to the consistency of the scores or answers from one administration of an instrument or intervention to another. High reliability ensures the results are tightly clustered around a (hopefully valid) target (Fraenkel et al., 2014). Though reliability of the NTY test was difficult to explore quantitatively since neither the test-retest method nor split halves method was possible to implement, group of teachers deemed the test a reliable indicator of the presence of misconceptions. Reliability of the overall study was enhanced by strict adherence to protocols and rubrics.

Generalizability, or external validity, is the ability to use the results of this study to say something about its applicability to other settings, other subject groups, or other periods of time (Creswell, 2011). For this study, the nature of a convenience sample limits the generalizability to all Algebra 2 students and all high schools, since this sample was drawn from a relatively unusual cohort of boarding school students. However, in the researcher’s experience, these regular Algebra 2 students are not more advanced mathematically than the general high school population and may be quite representative of the broader population of algebra students. Further, the misconceptions addressed in this study have been plaguing algebra students for decades (Marquis, 1988; Smith, 1946; Webber, 1929), further supporting the idea that the misconceptions in these students are representative of a more universal problem. Before generalizing across specific conditions, researchers must show similar outcomes through replicated experiments across those conditions (Fraenkel et al., 2014). Finally, external validity
was enhanced by assigning intact groups to treatments, thus keeping students in their “natural setting” (Dimitrov & Rumrill, 2003, p. 163).

Protection of Human Subjects

Northeastern’s Institutional Review Board (IRB) approved all procedures, protocols, and instrumentation to ensure the protection of human subjects. The appendices contain the Teacher Recruitment Email (Appendix A), Teacher Signed Consent Form (Appendix B), and the Focus Group Protocol (Appendix C).

Protection from harm. There was no identifiable risk of harm to the teacher participants or their students in this study. Although the students in Algebra 2 were minors aged 13-17 who are normally considered a vulnerable population that require special protections, it was reasonable to assume these students were sufficiently mature to make the assessment of risk to themselves (Fraenkel, 2014). Data collected was quantitative, except for student focus groups, for which parental permission was obtained. Like most education research, this study “involves activities that are within the customary, usual procedures of schools… and as such involve little or no risk” (Fraenkel, 2014, p. 63).

Students might have become anxious about completing the intervention because there it was graded. However, the grade assigned was for completion only, not for level of performance, and all students who completed the intervention received a grade of 100% on a bonus quiz. To minimize anxiety among students about their grade, or possibly their grade being affected by participation in the study, the researcher made clear to all students the scope and justification for the study.

It is possible to do harm by withholding a beneficial treatment. If the intervention was found to significantly help students reduce their algebraic misconceptions, then withholding it
from the control group could be considered harmful. For this reason, teachers who agreed
to participate in the intervention group deployed the intervention in all their sections. Further, the
researcher encouraged the teachers of the control group to deploy the intervention to their classes
in the spring term. One of the two control group teachers did so.

Another way to do harm is to deprive students of instruction they would have received
during the time when the intervention was taking place. Since students in the intervention group
only used minimal class time each day (5-10 minutes) to work on practice slips, they did not
miss out on the regular activities that were happening in the control class groups, so there was no
“opportunity cost” of missing other lessons to participate in the study.

**Informed consent.** As part of the recruitment process, teachers were asked to sign an
informed consent form that fully explained the research, risks, confidentiality, and their right to
withdraw from the study at any time without penalty.

It is important to note that, if a teacher decided to deploy the intervention, his or her
students’ participation in the intervention itself was not optional. This intervention has been
piloted for three years and continually revised in the researcher’s and others’ classrooms. As
such, it was part of the normal classroom practice for the researcher and some of the other
teachers in this math department.

For the focus group portion of the study, the research fell into the IRB “Full Review”
category due to the students’ status as minors. However, current Health and Human Services
guidelines currently allow exemptions for most categories of educational research that involve
“research on the effectiveness of or comparison among instructional techniques” and “use of
educational tests”, though this might not apply to research involving minors (Frankel, 2014, p.
With this in mind, parents provided written consent for focus group participation via email and students were allowed to opt-out of participating in the focus group by remaining silent.

**Confidentiality.** All students were assigned code names, and all post-tests and practice slips were signed using code names only. In addition, all published data will only refer to code names and aggregate data, with no possibility of linking the data to individual students. The cross-reference between code names and actual names was kept in a hidden column in an Excel spreadsheet, and all pre- and post-tests and any copies of all practice slips were kept in the researcher’s locked home. Once the research has been completed and published, the key linking code names and real names will be destroyed, though the tests and slips may be retained for further research related to the original purpose.

**Deception.** No deception was employed by the researcher or the intervention. The participants knew the reason for the intervention and its goals.

**Conclusion**

This quasi-experimental study with student focus groups was designed to test the effectiveness of the No-to-Yes intervention at changing persistent algebraic misconceptions. Since success in mathematics helps students attain greater success in school and in life, the topic is important and timely. The literature supported the use of augmented activation and refutational text in the context of a warm conceptual change based intervention. The design of the intervention focused on implementation fidelity and the use of multiple methods such as interleaving to help students improve their skills. Finally, the inclusion of student focus groups allowed a deeper exploration of how students reacted to the intervention. Given the success the intervention has shown, the researcher has already begun to share it with the math education community to help more students change their conceptions and experience math success.
Chapter 4: Research Findings

The purpose of this study was to assess the effectiveness of a warm conceptual change based intervention on remediating algebraic misconceptions held by students at a private secondary school. Overall, the study’s research findings showed that the intervention was successful at helping students overcome their persistent algebraic misconceptions.

As discussed in Chapter 3, this research study employed a quasi-experimental, pre-test post-test, control group design as well as student focus groups. Students in both the intervention and control groups took a pre-test to measure their baseline level of algebraic skills and demonstration of misconceptions. After the eight-week intervention, students in both groups took an identical post-test and the results were analyzed using repeated measures ANOVA, Mann-Whitney U tests, and linear regression.

The No-to-Yes (NTY) intervention targeted a set of pervasive misconceptions through testing, online work, and class practice. The eight misconception categories (MCs) were negative function inputs, rules of exponents, distribution of the negative sign, distribution errors, simplification of algebraic fractions, rules of fractions, rules of negative exponents, and a mixed review. Overall, the goal was for each student to decrease the number of misconception errors made. During the intervention, students could earn a “Yes!” in a MC in one of two ways: by doing well on the pre-test or by completing the corresponding module. To complete a module, the student first went online to watch a video and earn 100% on a practice quiz. After the online portion, the student needed to correctly complete a series of three practice slips in class. Daily progress was tracked on a wall chart. At the end of the intervention, students took an identical post-test and the differences were analyzed.

The results of the analysis of the four research questions are reported in this chapter.
RQ1 investigated the overall effectiveness of the intervention by looking at pre- and post-test results, while RQ2 examined the change in test scores for each MC individually. RQ3 explored how engagement predicted change in test scores. Finally, RQ4 qualitatively sought to understand the experiences of the students and teachers in the intervention. The statistical analysis was supported by Tabachnick and Fidell (2013), Verma (2015) and Laerd Statistics (2015a, 2015b, 2015d). Supporting details for all analyses are in the appendices.

**Research Question 1 (RQ1)**

To what extent were there significant differences over time in the MES on a test of algebraic misconceptions among the control group and two intervention groups?

**Descriptive exploration of MES.** The MES could range from 0 to 28 errors. Though the Full Group had on average 1.5 fewer misconception errors per student on the pre-test than the other two groups, this difference was relatively small (Full Group: $M = 13.8$, $SD = 4.2$; Partial Group: $M = 15.1$, $SD = 5.2$; Control Group: $M = 15.4$, $SD = 5.1$) (see Figure 9). The researcher re-graded the outliers in the Full Group and confirmed they were accurate.

![Figure 9](image.png)

*Figure 9: Boxplots of pre-test misconceptions error scores by group.*
On the post-test, there were more obvious differences between the groups, with the mean MES of the Full Group dropping to 4.6 ($SD = 2.7$), the Partial Group to 8.1 ($SD = 4.7$), and the Control Group to 11.4 ($SD = 5.7$) (see Figure 10). The Full Group improved on average by 9.2 questions (67%), the Partial Group by 7.0 (46%), and the Control Group by 3.9 (26%) (see Figure 11). This descriptive analysis supported the effectiveness of the intervention at reducing the number of misconceptions errors.

Figure 10: Boxplots of post-test misconception error scores by group.

Figure 11: Mean pre-test and post-test misconception errors scores by group.
Null hypothesis 1. There was no interaction between MES over time among the control group and two intervention groups.

As discussed in Chapter 3, a repeated measures ANOVA was used to test Hypothesis 1 (full details can be found in Appendix M). The continuous dependent variable was MES, which was measured at two levels of the independent variable time (MESpre and MESpost). The second categorical independent variable was group membership (ThreeGroups). The assumptions of the ANOVA model were checked and two minor violations found as detailed in Appendix M. As shown in Figure 12, the greatest decrease in MES from Time 1 (pre-test) to Time 2 (post-test) was in the Full Group, followed by the Partial Group, and finally the Control Group. The unequal slopes suggest that the interaction term might be significant. In fact, there was a statistically significant interaction between group membership and time on MES, $F(2, 64) = 12.83, p < .001$, partial $\eta^2 = .286$ (see Table 4).

![Profile plot showing estimated marginal mean number of errors by group](image-url)
Repeated Measures ANOVA for MES by Group

Tests of Within-Subjects Contrasts

<table>
<thead>
<tr>
<th>Source</th>
<th>Time</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
<th>Noncent. Parameter</th>
<th>Observed Powera</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Linear</td>
<td>1479.127</td>
<td>1</td>
<td>1479.127</td>
<td>243.21</td>
<td>.000</td>
<td>.792</td>
<td>243.207</td>
<td>1.000</td>
</tr>
<tr>
<td>Time * ThreeGroups</td>
<td>Linear</td>
<td>156.052</td>
<td>2</td>
<td>78.026</td>
<td>12.830</td>
<td>.000</td>
<td>.286</td>
<td>25.659</td>
<td>.996</td>
</tr>
<tr>
<td>Error(Time)</td>
<td>Linear</td>
<td>389.232</td>
<td>64</td>
<td>6.082</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Computed using alpha = .05

Since the assumption of homogeneity of variances was not met, the Games-Howell test was used to look at overall pairwise differences. As shown in Table 5, the MES was statistically significantly different between the Full Group and the Control Group over time (Mean Difference = -4.19, SE = 1.213, p = .004). The Full Group and Partial Group were not significantly different (Mean Difference = -2.44, SE = 1.214, p = .124) nor were the Partial Group and the Control Group (Mean Difference = -1.75, SE = 1.426, p = .445).

Table 5

Pairwise Differences for MES in the Repeated Measures ANOVA

<table>
<thead>
<tr>
<th>(I) ThreeGroups</th>
<th>(J) ThreeGroups</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Games-Howell</td>
<td>Control</td>
<td>4.19</td>
<td>1.213</td>
<td>.004</td>
<td>1.24 - 7.14</td>
</tr>
<tr>
<td></td>
<td>Full</td>
<td>1.75</td>
<td>1.426</td>
<td>.445</td>
<td>-1.71 - 5.20</td>
</tr>
<tr>
<td></td>
<td>Partial</td>
<td>-4.19</td>
<td>1.213</td>
<td>.004</td>
<td>-7.14 - -1.24</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>-2.44</td>
<td>1.214</td>
<td>.124</td>
<td>-5.42 - .53</td>
</tr>
<tr>
<td></td>
<td>Full</td>
<td>-1.75</td>
<td>1.426</td>
<td>.445</td>
<td>-5.20 - 1.71</td>
</tr>
<tr>
<td></td>
<td>Partial</td>
<td>2.44</td>
<td>1.214</td>
<td>.124</td>
<td>-.53 - 5.42</td>
</tr>
</tbody>
</table>

* Based on observed means.

The error term is Mean Square(Error) = 19.701.

* The mean difference is significant at the .05 level.

Looking at the simple main effects of group by time, there was not a statistically significant difference in MES between groups at pre-test (p = .52), which indicates the three
groups started out with approximately equivalent scores. However, there was a significant difference in MES at post-test, $F(2, 64) = 12.116, p < .001$, partial $\eta^2 = .275$ (see Table M2 in Appendix M). At the time of the post-test, the Full Group had a significantly lower MES than the Partial Group ($p = .045$) and the Control Group ($p < .001$), while the difference between the Partial and Control Groups was on the borderline of significance ($p = .052$).

Looking at the simple main effects of time by group, there was a statistically significant difference in MES between times for each of the three groups ($p < .001$) (see Table M3 in Appendix M). This means that all three groups improved over time. However, the effect size was greatest for the Full Group (partial $\eta^2 = .867$), followed by the Partial Group (partial $\eta^2 = .815$) and the Control Group (partial $\eta^2 = .584$).

Overall, the evidence from this test led to rejecting the null hypothesis that there was no significant difference over time in the post-test Misconception Error Scores among the groups. While there was not a significant simple main effect of group at pre-test, there was a significant main effect at post-test, with the Full Group making significantly fewer errors than both the Partial and Control Groups. Finally, while all three groups made fewer errors over time, the Full Group had the largest effect size, followed by the Partial and Control Groups. All evidence from this test indicated that the Full Group had significantly lower MES across time and as compared to the other groups.

**Research Question 2 (RQ2)**

For each MC, to what extent were there significant differences in the change in MES among students in the control group and successful students in the intervention group, after excluding students who earned a “Yes!” on the pre-test for that MC?
As discussed in Chapter 2, it was plausible that some misconceptions were “stickier” than others and would show a small or no effect from pre- to post-test. To answer RQ2, the data was first explored descriptively and then a series of seven Mann-Whitney U tests were run to look at each MC individually. For each MC, the dependent variable (MESchangeMCx) was computed by subtracting the pre-test score from the post-test score for each student, so that a negative number was a successful outcome indicating that the student made fewer error on the post-test.

**Descriptive exploration of MES by MC.** On the pre-test, there were a total of 991 misconception errors made by the 67 students. As shown in Figure 13, students made the most errors in MC5 and the least in MC1. For each MC, students in each group made roughly similar numbers of errors.

![Figure 13: Distribution of pre-test misconception errors by MC and group.](image)

On the post-test, the same 67 students made 506 misconception errors. As shown in Figure 14, students still made the most errors in MC5 and the least in MC1, though total numbers
decreased. A visual inspection confirmed that students in the Control Group made more post-test errors than students in either the Partial or Full Groups.

![Bar chart showing the distribution of post-test misconception errors by MC and group with total percent decrease from pre-test below each bar.](image)

*Figure 14:* Distribution of post-test misconception errors by MC and group with total percent decrease from pre-test below each bar.

Drilling deeper, it was important to know the distribution of the decrease in errors across the three groups so that the decrease could be attributed to the intervention or to regular Algebra 2 coursework. Figure 15 shows the percent decrease in each MC by group. For example, on MC1, the Full Group made 82% fewer misconception errors on the post-test than on the pre-test, while the Partial Group made 42% fewer errors and the Control Group 5% fewer errors. Across all MCs, the Full Group improved the most, followed by the Partial Group, and then the Control Group. Based on differences in percent improvement, MC1 appeared the most responsive to the intervention while MC4 was the least responsive.
Figure 15: Percent decrease in misconception errors from pre-test to post-test by MC and group.

The guide published by Laerd Statistics (2015a) was used extensively for the series of Whitney-Mann U tests that were run to answer RQ2. Instead of using data from all students in the study, for each MC, all students who earned a “Yes!” at pre-test were excluded, thus only retaining students who needed remediation in that MC. Further, for the Full and Partial Groups, only students who earned a “Yes!” during the intervention were included. Thus, each test compared only students who needed to improve in an MC, with one group successfully completing the intervention for that MC and the other group serving as a control. A significant result meant that, for students who needed to work on this MC, successful intervention group students decreased the number of misconception errors they made in that MC more than the control group students did. Detailed discussion for each MC follows the summary statistics given in Table 6. Due to SPSS requirements, the successful intervention group students were coded as Group 1, and the control group as Group 2.
Table 6

<table>
<thead>
<tr>
<th>MC</th>
<th>Group</th>
<th>Sample Size</th>
<th>Mean MES pre-test (max of 4)</th>
<th>Mean MES post-test (max of 4)</th>
<th>Mean Decrease in MES</th>
<th>Percent decrease in MES</th>
<th>p-value of U-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC1</td>
<td>Intervention</td>
<td>15</td>
<td>1.40</td>
<td>0.07</td>
<td>-1.33</td>
<td>95%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>12</td>
<td>1.58</td>
<td>1.00</td>
<td>-0.58</td>
<td>37%</td>
<td>0.02*</td>
</tr>
<tr>
<td>MC2</td>
<td>Intervention</td>
<td>20</td>
<td>2.15</td>
<td>0.95</td>
<td>-1.2</td>
<td>56%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>26</td>
<td>2.44</td>
<td>2.12</td>
<td>-0.32</td>
<td>13%</td>
<td>0.012*</td>
</tr>
<tr>
<td>MC3</td>
<td>Intervention</td>
<td>17</td>
<td>1.71</td>
<td>0.35</td>
<td>-1.35</td>
<td>79%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>21</td>
<td>2.19</td>
<td>1.48</td>
<td>-0.71</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td>MC4</td>
<td>Intervention</td>
<td>17</td>
<td>2.35</td>
<td>0.65</td>
<td>-1.71</td>
<td>73%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>23</td>
<td>2.35</td>
<td>1.35</td>
<td>-1.00</td>
<td>43%</td>
<td>0.135</td>
</tr>
<tr>
<td>MC5</td>
<td>Intervention</td>
<td>19</td>
<td>3.37</td>
<td>0.58</td>
<td>-2.78</td>
<td>83%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>24</td>
<td>3.33</td>
<td>2.42</td>
<td>-0.92</td>
<td>28%</td>
<td>0.001*</td>
</tr>
<tr>
<td>MC6</td>
<td>Intervention</td>
<td>11</td>
<td>2.00</td>
<td>0.18</td>
<td>-1.82</td>
<td>91%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>24</td>
<td>3.25</td>
<td>2.13</td>
<td>-1.13</td>
<td>35%</td>
<td>0.060</td>
</tr>
<tr>
<td>MC7</td>
<td>Intervention</td>
<td>10</td>
<td>1.9</td>
<td>0.4</td>
<td>-1.50</td>
<td>79%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>25</td>
<td>2.44</td>
<td>1.92</td>
<td>-0.52</td>
<td>21%</td>
<td>0.002*</td>
</tr>
</tbody>
</table>

* Significant at alpha = 0.05

**Null hypothesis 2A.** For MC1, there was no difference in the distribution of the change in MES between students in the control group and students in the intervention group who earned a “Yes!” during the intervention, after excluding all students who earned a “Yes!” on the pre-test in MC1.

A Mann-Whitney U test was run to determine if there were differences in the change in MES scores between 15 successful students in the intervention group and 12 students in the control group (excluding all students who earned a “Yes!” on the pre-test in MC1). The dependent variable was the change in MES for MC1 (MES\text{changeMC1}) and the categorical independent variable was group membership (YesDuring1). Distributions of the MES for students in the control group and successful students in the intervention groups were not similar,
as assessed by visual inspection (see Figure 16). MES for the successful intervention group students (mean rank = 11.13) and the Control group (mean rank = 17.58) were statistically significantly different, \( U = 47.0, z = -2.322 \), asymptotic \( p = .020 \) (see Table 7). The asymptotic \( p \) was used because there were ties in the data. Therefore, for MC1, we reject the null hypothesis and conclude that successful intervention group students made fewer misconception errors on MC1 than the Control group.

*Figure 16:* Side-by-side histograms for the change in MES by group for MC1.
Table 7

*Mann-Whitney U-Test for MC1*

<table>
<thead>
<tr>
<th>YesDuringNum</th>
<th>N</th>
<th>Mean Rank</th>
<th>Sum of Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>MESchangeMC1</td>
<td>1</td>
<td>15</td>
<td>11.13</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>12</td>
<td>17.58</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Test Statistics*<sup>a</sup>

<table>
<thead>
<tr>
<th></th>
<th>MESchangeMC1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mann-Whitney U</td>
<td>47.000</td>
</tr>
<tr>
<td>Wilcoxon W</td>
<td>167.000</td>
</tr>
<tr>
<td>Z</td>
<td>-2.322</td>
</tr>
<tr>
<td>Asymp. Sig. (2-tailed)</td>
<td>.020</td>
</tr>
<tr>
<td>Exact Sig. [2*(1-tailed Sig.)]</td>
<td>.037&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup> Grouping Variable: YesDuringNum

<sup>b</sup> Not corrected for ties.

**Null hypothesis 2B.** For MC2, there was no difference in the distribution of the change in MES between students in the control group and students in the intervention group who earned a “Yes!” during the intervention, after excluding all students who earned a “Yes!” on the pre-test in MC2.

A Mann-Whitney U test was run to determine if there were differences in the change in MES scores between 20 successful students in the intervention group and 26 students in the control group (excluding all students who earned a “Yes!” on the pre-test in MC2). The dependent variable was the change in MES for MC2 (MESchangeMC2) and the categorical independent variable was group membership (YesDuring2). Distributions of the MES for students in the control group and successful students in the intervention groups were not similar, as assessed by visual inspection (see Figure 17). MES for the successful intervention group students (mean rank = 18.03) and the control group (mean rank = 27.71) were statistically significantly different, $U = 150.5$, $z = -2.499$, asymptotic $p = .012$ (see Table 8). Therefore, for
MC2, we reject the null hypothesis and conclude that successful intervention group students made fewer misconception errors on MC2 than the control group.

![Graph showing side-by-side histograms for the change in MES by group for MC2.]

Figure 17: Side-by-side histograms for the change in MES by group for MC2.

Table 8

**Mann-Whitney U-Test for MC2**

<table>
<thead>
<tr>
<th>YesDuring2Num</th>
<th>N</th>
<th>Mean Rank</th>
<th>Sum of Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>MESchangeMC2</td>
<td>1</td>
<td>20</td>
<td>18.03</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>26</td>
<td>27.71</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>46</td>
<td></td>
</tr>
</tbody>
</table>

**Test Statistics**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MESchangeMC2</td>
<td></td>
</tr>
<tr>
<td>Mann-Whitney U</td>
<td>150.500</td>
</tr>
<tr>
<td>Wilcoxon W</td>
<td>360.500</td>
</tr>
<tr>
<td>Z</td>
<td>-2.499</td>
</tr>
<tr>
<td>Asymp. Sig. (2-tailed)</td>
<td>.012</td>
</tr>
</tbody>
</table>

a. Grouping Variable: YesDuring2Num
Null hypothesis 2C. For MC3, there was no difference in the distribution of the change in MES between students in the control group and students in the intervention group who earned a “Yes!” during the intervention, after excluding all students who earned a “Yes!” on the pre-test in MC3.

A Mann-Whitney U test was run to determine if there were differences in the change in MES scores between 17 successful students in the intervention group and 21 students in the control group (excluding all students who earned a “Yes!” on the pre-test in MC3). The dependent variable was the change in MES for MC3 (MESchangeMC3) and the categorical independent variable was group membership (YesDuring3). Distributions of the MES for students in the control group and successful students in the intervention group were not similar, as assessed by visual inspection (see Figure 18). MES for the successful intervention group students (mean rank = 16.21) and the control group (mean rank = 22.17) were not statistically significantly different, $U = 122.5$, $z = -1.715$, asymptotic $p = .086$ (see Table 9). Therefore, for MC3, despite the lower mean ranking that indicates intervention group students made fewer misconception errors on MC3, we cannot reject the null hypothesis that there was no difference between the groups.
Figure 18: Side-by-side histograms for the change in MES by group for MC3.

Table 9

Mann-Whitney U-Test for MC3

<table>
<thead>
<tr>
<th>MESchangeMC3</th>
<th>YesDuring3Num</th>
<th>N</th>
<th>Mean Rank</th>
<th>Sum of Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>17</td>
<td>16.21</td>
<td>275.50</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>21</td>
<td>22.17</td>
<td>465.50</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>38</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Test Statistics\(^a\)

<table>
<thead>
<tr>
<th>MESchangeMC3</th>
<th>Mann-Whitney U</th>
<th>Wilcoxon W</th>
<th>Z</th>
<th>Asymp. Sig. (2-tailed)</th>
<th>Exact Sig. [2*(1-tailed Sig.)]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>122.500</td>
<td>275.500</td>
<td>-1.715</td>
<td>.086</td>
<td>.101(^b)</td>
</tr>
</tbody>
</table>

\(^a\) Grouping Variable: YesDuring3Num
\(^b\) Not corrected for ties.
Null hypothesis 2D. For MC4, there was no difference in the distribution of the change in MES between students in the control group and students in the intervention group who earned a “Yes!” during the intervention, after excluding all students who earned a “Yes!” on the pre-test in MC4. A Mann-Whitney U test was run to determine if there were differences in the change in MES scores between 17 successful students in the intervention group and 23 students in the control group (excluding all students who earned a “Yes!” on the pre-test in MC4). The dependent variable was the change in MES for MC4 (MES\(_{\text{changeMC4}}\)) and the categorical independent variable was group membership (Yes\_During4). Distributions of the MES for students in the control group and successful students in the intervention group were not similar, as assessed by visual inspection (see Figure 19). MES for the successful intervention group students (mean rank = 17.41) and the control group (mean rank = 22.78) were not statistically significantly different, \(U = 143.0, z = -1.493, \text{asymptotic } p = .135\) (see Table 10). The asymptotic \(p\) was used because there were ties in the data. Therefore, for MC4, despite the lower mean ranking that indicates intervention group students made fewer misconception errors on MC4, we cannot reject the null hypothesis that there is no difference between the groups.

Figure 19: Side-by-side histograms for the change in MES by group for MC4.
Table 10

**Mann-Whitney U-Test for MC4**

<table>
<thead>
<tr>
<th>YesDuring4Num</th>
<th>N</th>
<th>Mean Rank</th>
<th>Sum of Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>MESchangeMC4</td>
<td>1</td>
<td>17</td>
<td>17.41</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>23</td>
<td>22.78</td>
</tr>
<tr>
<td>Total</td>
<td>40</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Test Statistics**

<table>
<thead>
<tr>
<th></th>
<th>MESchangeMC4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mann-Whitney U</td>
<td>143.000</td>
</tr>
<tr>
<td>Wilcoxon W</td>
<td>296.000</td>
</tr>
<tr>
<td>Z</td>
<td>-1.493</td>
</tr>
<tr>
<td>Asymp. Sig. (2-tailed)</td>
<td>.135</td>
</tr>
<tr>
<td>Exact Sig. [2*(1-tailed Sig.)]</td>
<td>.156(^b)</td>
</tr>
</tbody>
</table>

a. Grouping Variable: YesDuring4Num
b. Not corrected for ties.

**Null hypothesis 2E.** For MC5, there was no difference in the distribution of the change in MES between students in the control group and students in the intervention group who earned a “Yes!” during the intervention, after excluding all students who earned a “Yes!” on the pre-test in MC5.

A Mann-Whitney U test was run to determine if there were differences in the change in MES scores between 19 successful students in the intervention group and 24 students in the control group (excluding all students who earned a “Yes!” on the pre-test in MC5). The dependent variable was the change in MES for MC5 (MESchangeMC5) and the categorical independent variable was group membership (YesDuring5). Distributions of the MES for students in the control group and successful students in the intervention groups were not similar, as assessed by visual inspection (see Figure 20). MES for the successful intervention group students (mean rank = 15.08) and the control group (mean rank = 27.48) were statistically significantly different, \( U = 96.5 \), \( z = -3.301 \), asymptotic \( p = .001 \) (see Table 11). Therefore, for
MC5, we reject the null hypothesis and conclude that successful intervention group students made fewer misconception errors on MC5 than the control group.

**Figure 20**: Side-by-side histograms for the change in MES by group for MC5.

**Table 11**

*Mann-Whitney U-Test for MC5*

<table>
<thead>
<tr>
<th>MESchangeMC5</th>
<th>YesDur5Num</th>
<th>N</th>
<th>Mean Rank</th>
<th>Sum of Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>MESchangeMC5</td>
<td>1</td>
<td>19</td>
<td>15.08</td>
<td>286.50</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>24</td>
<td>27.48</td>
<td>659.50</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>43</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Test Statistics*

<table>
<thead>
<tr>
<th></th>
<th>MESchangeMC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mann-Whitney U</td>
<td>96.500</td>
</tr>
<tr>
<td>Wilcoxon W</td>
<td>286.500</td>
</tr>
<tr>
<td>Z</td>
<td>-3.301</td>
</tr>
<tr>
<td>Asymp. Sig. (2-tailed)</td>
<td>.001</td>
</tr>
</tbody>
</table>

a. Grouping Variable: YesDur5Num
**Null hypothesis 2F.** For MC6, there was no difference in the distribution of the change in MES between students in the control group and students in the intervention group who earned a “Yes!” during the intervention, after excluding all students who earned a “Yes!” on the pre-test in MC6.

A Mann-Whitney U test was run to determine if there were differences in the change in MES scores between 11 successful students in the intervention group and 24 students in the control group (excluding all students who earned a “Yes!” on the pre-test in MC6). The dependent variable was the change in MES for MC6 (MESchangeMC6) and the categorical independent variable was group membership (YesDuring6). Distributions of the MES for students in the control group and successful students in the intervention groups were not similar, as assessed by visual inspection (see Figure 21). The MES for the successful intervention group students (mean rank = 13.36) and the control group (mean rank = 20.13) were not statistically significantly different, $U = 81.0, z = -1.881$, asymptotic $p = .060$ (see Table 12). For MC6, despite the lower ranking that indicates intervention group students made fewer misconception errors on MC6, we cannot reject the null hypothesis of no difference between the groups.

![Figure 21: Side-by-side histograms for the change in MES by group for MC6.](image-url)
Table 12

*Mann-Whitney U-Test for MC6*

<table>
<thead>
<tr>
<th>YesDuring6Num</th>
<th>N</th>
<th>Mean Rank</th>
<th>Sum of Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>MESchangeMC6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>11</td>
<td>13.36</td>
<td>147.00</td>
</tr>
<tr>
<td>2</td>
<td>24</td>
<td>20.13</td>
<td>483.00</td>
</tr>
<tr>
<td>Total</td>
<td>35</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Test Statistics*<sup>a</sup>

<table>
<thead>
<tr>
<th></th>
<th>MESchangeMC6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mann-Whitney U</td>
<td>81.000</td>
</tr>
<tr>
<td>Wilcoxon W</td>
<td>147.000</td>
</tr>
<tr>
<td>Z</td>
<td>-1.881</td>
</tr>
<tr>
<td>Asymp. Sig. (2-tailed)</td>
<td>.060</td>
</tr>
<tr>
<td>Exact Sig. [2*(1-tailed Sig.)]</td>
<td>.072&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup> Grouping Variable: YesDuring6Num

<sup>b</sup> Not corrected for ties.

**Null hypothesis 2G.** For MC7, there was no difference in the distribution of the change in MES between students in the control group and students in the intervention group who earned a “Yes!” during the intervention, after excluding all students who earned a “Yes!” on the pre-test in MC7.

A Mann-Whitney U test was run to determine if there were differences in the change in MES scores between 10 successful students in the intervention group and 25 students in the control group (excluding all students who earned a “Yes!” on the pre-test in MC7). The dependent variable was the change in MES for MC7 (MESchangeMC7) and the categorical independent variable was group membership (YesDuring7). Distributions of the MES for students in the control group and successful students in the intervention groups were not similar, as assessed by visual inspection (see Figure 22). MES for the successful intervention group students (mean rank = 10) and the control group (mean rank = 21.2) were statistically significantly different, $U = 45.0$, $z = -3.09$, asymptotic $p = .002$ (see Table 13). The asymptotic $p$
was used because there were ties in the data. Therefore, for MC7, we reject the null hypothesis and conclude that successful intervention group students made fewer misconception errors on MC7 than the control group.

*Figure 22:* Side-by-side histograms for the change in MES by group for MC7.

Table 13

*Mann-Whitney U-Test for MC7*

<table>
<thead>
<tr>
<th>MESchangeMC7</th>
<th>YesDuring7Num</th>
<th>N</th>
<th>Mean Rank</th>
<th>Sum of Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>10</td>
<td>10.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>25</td>
<td>21.20</td>
<td>530.00</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>35</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Test Statistics*\(^a\)

<table>
<thead>
<tr>
<th></th>
<th>MESchangeMC7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mann-Whitney U</td>
<td>45.000</td>
</tr>
<tr>
<td>Wilcoxon W</td>
<td>100.000</td>
</tr>
<tr>
<td>Z</td>
<td>-3.090</td>
</tr>
<tr>
<td>Asymp. Sig. (2-tailed)</td>
<td>.002</td>
</tr>
<tr>
<td>Exact Sig. [2*(1-tailed Sig.)]</td>
<td>.003(^b)</td>
</tr>
</tbody>
</table>

\(^a\) Grouping Variable: YesDuring7Num

\(^b\) Not corrected for ties.
Overall, there were significant differences between the included successful intervention group Students and control group students for MCs 1, 2, 5, and 7. MCs 3, 4, and 6 did not have significant results, though in all cases the successful intervention group students improved 30%-56% more than the control group students. MCs 5 and 7 had the largest difference in mean MES change, with successful students in the intervention groups making 1.82 fewer errors than the control group for MC5 and .98 fewer errors in MC7.

**Research Question 3 (RQ3)**

To what extent did engagement predict the depth of conceptual change, as measured by the change in MES in the two intervention groups? RQ3 investigated the warm conceptual change theory tenet that the interaction between the student and the message (the intervention) leads to depth of engagement, which in turn leads to conceptual change (Dole & Sinatra, 1998). Two regression models were run to investigate if any of the methods of engagement caused conceptual change as measured by the change in MES.

RQ3 only included students in the Full and Partial Groups. For each student, the change in MES (MESchange) was computed by subtracting the post-test MES from the pre-test MES so that a negative number was the desired result showing that the student made fewer errors on the post-test than the pre-test. There were three measures of individual engagement: a combination of videos watched and video questions correctly answered (TotalVideo), completely correct online quizzes (Total4Quiz), and completely correct in-class practice slips (TotalYesSlips). These three measures were combined into one overall measure (TotalEngage). Table 14 gives descriptive statistics for the engagement variables.
Table 14

Descriptive Statistics for Variables in RQ3

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>TotalEngage</td>
<td>41</td>
<td>0.2</td>
<td>115.5</td>
<td>56.45</td>
<td>32.91</td>
</tr>
<tr>
<td>TotalVideo</td>
<td>41</td>
<td>0</td>
<td>38.6</td>
<td>18.77</td>
<td>11.9</td>
</tr>
<tr>
<td>Total4Quiz</td>
<td>41</td>
<td>0</td>
<td>7</td>
<td>4.00</td>
<td>2.52</td>
</tr>
<tr>
<td>TotalYesSlips</td>
<td>41</td>
<td>0</td>
<td>22</td>
<td>9.44</td>
<td>7.46</td>
</tr>
<tr>
<td>MESChange</td>
<td>41</td>
<td>-16</td>
<td>-1</td>
<td>-8.02</td>
<td>3.66</td>
</tr>
</tbody>
</table>

**Null hypothesis 3A.** There was no relationship between an overall measure of total engagement (TotalEngage) and the depth of conceptual change as measured by change in MES (MESchange) among students in the two intervention groups.

A simple linear regression was run to understand the effect of engagement (TotalEngage) on the change in MES (MESchange). The Laerd Statistics (2015c) online guide was used to guide the analysis. Before interpreting the results of the test, the assumptions were checked. First, a scatterplot of total engagement against change in MES with the regression line (see Figure 23) showed a linear relationship between the variables. There was independence of residuals, as assessed by a Durbin-Watson statistic of 1.464. There was homoscedasticity (as shown by a visual inspection of a scatterplot of standardized predicted values vs. standardized residuals) and normality of the residuals (as shown by a visual inspection of a Normal P-P plot) (see Figure 24). Finally, there were no outliers with a standardized residual greater than three standard deviations from the mean.
Figure 23: Scatterplot of total engagement vs. change in MES

Figure 24: Scatterplot of predicted vs. residual values of MES and Normal P-P plot

Once the assumptions were checked, the results were interpreted (see Table 15). Total engagement significantly predicted MES change, $F(1, 39) = 8.494, p = .006$, accounting for
17.9% of the variation in MES change (adjusted $R^2 = 15.8\%$). There was a medium-large effect size (Cohen’s $f^2 = .218$) (Selya, Rose, Dierker, Hedeker, & Mermelstein, 2012). Therefore, the null hypothesis was rejected. On average, students included in this analysis made eight fewer errors on the post-test than on the pre-test. The significant intercept value of $\beta = -5.366$ ($p < .001$) can be interpreted to mean that students who did not engage in the intervention at all experienced a drop of about five points from pre-test to post-test. The remaining three point decrease was accounted for by total engagement in the intervention.

Table 15

**Simple Regression Results for RQ3**

<table>
<thead>
<tr>
<th>Model Summary $^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), TotalEngage  
b. Dependent Variable: MESChange

<table>
<thead>
<tr>
<th>ANOVA $^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

a. Dependent Variable: MESChange  
b. Predictors: (Constant), TotalEngage

<table>
<thead>
<tr>
<th>Coefficients $^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

a. Dependent Variable: MESChange
**Null hypothesis 3B.** There was no relationship between video engagement (TotalVideo), number of correct online quizzes (Total4Quiz), and number of correct in-class practice slips (TotalYesSlips) and the depth of conceptual change as measured by change in MES (MESchange) among students in the two intervention groups.

To test this hypothesis that broke down engagement into three components, a multiple linear regression model was run following the online guide published by Laerd Statistics (2015b). Assumptions were checked and met, with details in Appendix N. The multiple linear regression analysis indicated a statistically significant overall model, $F(3, 37) = 3.219, p = .034$, adjusted $R^2 = 0.143$, with a medium to large effect size (Cohen’s $f^2 = .261$) (see Table 16).

Therefore, the null hypothesis was rejected. However, none of the three independent variables had a statistically significant p-value; TotalVideo had the highest significance ($p = .122$), followed by Total4Quiz ($p = .222$) and TotalYesSlips ($p = .879$). As predicted, the coefficients for TotalVideo and Total4Quiz were negative, meaning they were associated with a decrease in misconception errors. The coefficient for TotalYesSlips was slightly positive (0.016). Regression coefficients and standard errors can be found in Table 17. Only the intercept of -5.016 was significant ($p < .001$), which again can be interpreted to mean students’ scores dropped by about 5 points without engaging at all in the intervention.

Overall, the analysis for RQ3 supported the CRKM theory, since a higher level of engagement in the intervention led to a decrease in the number of errors on the post-test, presumably via a student’s conceptual change.
Table 16

Results for the Multiple Linear Regression Model Predicting Change in MES from Pre-Test to Post-Test

Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
<th>Durbin-Watson</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.455&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.207</td>
<td>.143</td>
<td>3.392</td>
<td>1.552</td>
</tr>
</tbody>
</table>

- a. Predictors: (Constant), Total4Quiz, TotalVideo, TotalYesSlips
- b. Dependent Variable: MChange

ANOVA

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Regression</td>
<td>3</td>
<td>37.050</td>
<td>3.219</td>
<td>.034&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>37</td>
<td>11.509</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>40</td>
<td>536.976</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- a. Dependent Variable: MChange
- b. Predictors: (Constant), Total4Quiz, TotalVideo, TotalYesSlips

Table 17

Coefficients for the Multiple Linear Regression Model Predicting Change in MES from Pre-Test to Post-Test

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unstandardized Coefficients</td>
<td>Standardized Coefficients</td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>B</td>
<td>Std. Error</td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>-5.016</td>
</tr>
<tr>
<td></td>
<td>TotalYesSlips</td>
<td>.016</td>
</tr>
<tr>
<td></td>
<td>TotalVideo</td>
<td>-.098</td>
</tr>
<tr>
<td></td>
<td>Total4Quiz</td>
<td>-.328</td>
</tr>
</tbody>
</table>

- a. Dependent Variable: MChange
Research Question 4 (RQ4)

How did participating students experience the intervention? This qualitative question sought to understand the experience of the students in the intervention groups. The question was explored through post-intervention focus groups with each of the intervention classes and augmented with interviews with intervention group teachers. Questions included what students liked and did not like, how they used the online component, and recommendations for improvement (see Appendix C). As discussed in Chapter 3, the transcripts were analyzed through an inductive, cyclical process of coding and re-coding as themes emerged. MAXQDA was used to streamline the process of identifying patterns and commonalities in the data. As discussed below, the six themes that emerged were time pressure, motivation, spaced and interleaved learning, video design (including augmented activation), recommendations for improvement, and success stories.

Theme 1: Time pressure. Time pressure arose in a number of contexts in all the focus groups and interviews. The largest concern among students was the daily pressure on them to find time for the intervention on top of their regular work. In addition, there was pressure to finish the intervention in the number of weeks allotted. This pressure, along with student recommendation to alleviate it, are discussed below.

Daily time pressure. Across all the intervention classes, students expressed their frustration with the added daily time requirements of this intervention. Students at this school have very busy schedules, usually including athletics and other extra-curriculars on top of their academic workload. As a result, stress levels are high and students resent further demands on their time. One student commented that “I had more free time to work on it, I feel like I would have done more, but I just had a lot of more pressing obligations.” Another added “it was hard
to find time to do it with all our other work here.” Many students stated something similar to “in the beginning, I was definitely stressed because I have so much homework in my other classes and then I had to watch videos for this.” However, not everyone agreed, including one student from the Partial Group (Kepler) who disagreed with her classmates by stating, “I think if you have to find time to do it, you can.”

To relieve daily time pressure, both teachers offered students class time for the intervention. Mr. M typically offered 5-10 minutes at the end of class, while Mr. G offered 5-10 minutes at the beginning of class. Mr. M’s students tended to use the time at the end of class to start their homework instead of working on the intervention, so Mr. M’s students made much less progress overall. Mr. M reflected that:

the students that most need the No-to-Yes structure and reinforcement and knowledge to help them are the students who are the least likely to do it…I would give time at the end of class to do the slips, but I would never say, ‘Now is the time to do the No-to-Yes.’ I would say, ‘Now is the time you can work on the No-to-Yes or homework,’ and the stronger students would go do the No-to-Yes, and the weaker students were like, ‘I'm getting my homework done.’

A student in that class concurred, stating, “During class, I never really had time, because I would always use the extra time to get started on homework…even though it's like five or ten minutes, I just couldn't do the project.”

In contrast, in Mr. G’s classes the students universally took advantage of the time offered to work on the intervention when it was at the beginning of class. The amount of time given was “about right.” They effectively used the down time when “everyone's kind of filing into the classroom,” since “everyone gets here at different times and you can just go straight to it and
work on it.” They even acknowledged “we would waste a few minutes at the beginning of class if we didn't have the project.” Mr. G was “adamant” that any teacher implementing this intervention make time at the beginning of class, because at the end “you're either trying to wrap up a topic or they're trying to do homework, and they're unmotivated to stop doing homework to work on the slips.” Also, in the beginning of class, they “all gather around the box to look at their slips and they're excited and energized, and then they rush to get back and get their new slips done. It's motivating because they know they only have five to seven minutes.”

In addition to having time during class, students had other suggestions on ways to relieve some time pressure. They suggested that the teacher might “combine it with the amount of homework we usually get,” rather than just adding to it. One student suggested “it would maybe be easier if maybe they (the slips) were online or on Canvas so that you can really take them whenever you had the time.”

**Intervention completion pressure.** Even when students were working diligently on the intervention, most did not have enough time to completely finish it before the term was over. Only two students (in the Full Group) out of 41 intervention students completely finished the intervention before the end of the term, with another nine students working on their last MC. The students felt pressure to work faster and get more right answers in order to get it all done. “Maybe we should have…just a bit more time to actually finish it all,” commented a student in the Partial Group, since “it's simply just not possible to finish because you run out of time.” One student in the Full Group (Cauchy) added,

> I remember, over the last two weeks, you started saying, ‘All right, we're ending soon. This is just a reminder. Everybody keep doing your slips.’ And that's when I thought, ‘Oh no, I have a lot to do. A lot.’
This frustration was shared by many of the students who were deeply engaged in the work.

Some students ran out of time to complete the intervention because they started off slowly in the beginning. When the intervention kicked off, students were instructed that they needed to unlock one category for homework that night. Not all of them complied and there was no penalty, which led to some students delaying entry into the intervention. “I don't know how, but somehow finding a way to help kids get into it sooner on,” suggested one student. To do this, Mr. G dedicated one full class period early in the intervention, which he noted “made a huge difference. After that everybody finally got some momentum.” Another student suggestion was to send “an email every two weeks with, ‘Hey, work on your slips.’ Or a reminder in the beginning of class, we only have this much time left.” Since earlier engagement led to more completion in the intervention, Mr. G. recommended finding the time so every student can “unlock one category on the first day.”

Despite the daily time pressure, students in Mr. G’s classes found the time to complete 71% of the intervention, while those in Mr. M’s completed only 13%. This strong difference in results reflected the difference in motivation between the two groups, which is the second theme that emerged from the focus groups.

**Theme 2: Motivation.** It can be difficult to convince students that extra work in the short term is worthwhile in the long run. Because student buy-in was crucial to create engagement, the intervention was designed to foster motivation in three ways: intrinsically, extrinsically, and socially. Unless a student experienced some or all of these interrelated forms of motivation, they were unlikely to engage deeply with this optional intervention. As discussed in Chapter 3, a variety of methods were employed to motivate students. The success of these
methods was highly varied across students and especially between the Full and Partial Groups. In this theme, students’ motivation or lack thereof and their recommendations are discussed.

Intrinsic motivation. To foster intrinsic motivation, students needed to buy in to the long-term benefits to be gained from deeply engaging in the intervention. To convince them of the need to engage, the researcher first handed back their graded pre-tests on kick-off day. Most students were shocked and dismayed when they realized how poorly they had done. The researcher then explained some of the most common misconception errors on the quiz and how students in calculus classes continued to make the same errors. Further, during the intervention, each video was designed to activate their misconceptions to again startle them into buying in to the need for change. Throughout the intervention, the teachers pointed out when these common errors arose in regular classwork and assessments.

Students in the Full Group were convinced that there was long-term intrinsic benefit in fixing their misconceptions and hence got excited about the intervention as they did the work. It was worth doing “because also it helps you with tests and exams and stuff in general.” One student thought that “we benefited a lot” from engaging in all parts of the intervention, and that figuring out what you were doing wrong was “most effective.” Mr. G noted that Bernoulli’s primary motivation during the intervention was to “master these fundamentals so he wouldn’t miss problems on assessments because of basic mechanics. More than once, he mentioned that he hadn’t really learned how to do these concepts correctly or that he had forgotten the rules.” Bernoulli was “obviously pleased” that he was “mastering the concepts he knew were necessary for his math success in the foreseeable future.”
Once students started experiencing success in the program, their intrinsic motivation and enthusiasm grew. Aristotle (Full Group) felt the intervention became “kind of fun and exciting, how you would get a “Yes!” and it would be so exciting!” Aristotle went on to explain that

As you got on with the eight boxes (MCs), once you got close to the finish line, we all got into it so much because every “Yes!” counted. And you could see that we didn't get the third slip right and we'd be, "What!?" And we'd go back and check it and then it was just really exciting.

Mr. G was “surprised” at how much the students “liked the whole idea of paper slips… (it was) sort of like getting a little present every day when they come back and see if they got all four right on their slips. They were really energized by that.” Once students in his classes “started to see progress in different categories. That was a motivator.”

With intrinsic motivation came the persistence to keep trying and the desire to learn. Chebyshev, who scored 100% on the post-test, felt her intrinsic motivation grew with every green “Yes!” on a slip or the wall chart. She explained that she “liked having little successes in my day because at (school) you lose sometimes. (It’s like) getting a little wind.” She completed eight separate slips in the eighth category but only got one completely correct. Despite that, “I kept going and I kept getting them wrong and so I feel if I had more practice for another term I might be able to apply the things I learned and then be able to get that eighth section.’’

In contrast, no talk of excitement arose in the Partial Group, which indicated many students were never convinced of the intrinsic value of engaging deeply in the intervention. Mr. M “had one class (in which) basically no one did it in, and the other classes were sporadic. The two strongest students who did the most were my two stronger students, were like, ‘This is easy to do.’” However, “most of the students who are struggling in class are the ones who hate math,
the ones who don’t have that intrinsic motivation or excitement to do this work.” Instead, they focused on extrinsic motivation.

*Extrinsic motivation.* The intervention teachers gave bonus points on quizzes for progress in the intervention as extrinsic motivation to engage with the optional project. The students could earn two points for earning a “Yes!” on an individual MC and also 100% on a bonus quiz if they completed the intervention. There was a clear divide between the Full and Partial Groups on the role of this external motivation and whether, in the future, the intervention should be mandatory or optional.

The Full Group liked the extra credit and wanted the intervention to remain optional. Since their intrinsic motivation was high, the extrinsic incentives were just an added bonus. As students said, “I would have done it anyway for extra credit,” and “being optional took a lot of the pressure off... It was, oh, I can do this to get extra credit. I think the idea of not being forced made it look a little bit more appealing.” One student spoke for many of her classmates when she said, “I think we’ll all do it! I mean, it’s a (free) quiz.” Mr. G reflected that by giving extra credit points on quizzes, the students “seemed to be very enthusiastic and felt like it was a lot...it was interesting that the most weight was a full 100-point quiz, but (to them) it seemed to be less incentive than two points at a time.”

The Partial Group, who largely felt there was no intrinsic benefit, wanted the intervention to be “mandatory because people otherwise won’t do it.” Some wanted it to be “assigned, like maybe as homework. That would have been more efficient and more incentive to do it.” Or another student suggested it should be “mandatory, but not for like every student. Have some sort of cutoff below (a certain grade).” Others were not sure, such as a student in the Partial Group who said “I'm not sure if making it mandatory is good … obviously it's a good way to get
kids to do it but… not everyone has that kind of time and it takes different amounts of time for different people.” Another Partial Group student who completed only 2.5 out of eight MCs, added “I also liked how the project…helped with your class grade. I think that's a way to motivate kids to do it.” Finally, another Partial Group student theorized, “if you want to fix people's mistakes, I feel like it has to be more forced rather than an option.”

Jacobi (Partial Group) typified the student who felt no intrinsic motivation and hence needed more extrinsic motivation to participate. “If you spent an hour doing this, you might complete it, but if you spent an hour studying for your next quiz, it might be more productive… time on this is not as effective as studying for your next test.” This student did not see the link between spending time fixing misconceptions and future success in mathematics. He admitted to skipping through the videos and “gaming the system” to attempt to get the work done quickly, though he only unlocked one MC during the intervention. While he did not believe the intervention should be mandatory, he believed that incentives are a “very important” part of this project and that the current incentive system was “not good enough.” If the goal was to get 100% participation, we would need to “use higher values. Let's say everything you complete, you get ten points worth” instead of one point.

Lovelace (Partial Group) argued for the intervention to be optional, because the motivation ought to be intrinsic:

Us getting extra points shouldn't be the reason (to do the project), it should be our own motivation to go and do it, so that's why I think it should be optional, not mandatory…

(At) school you should be more independent. Know your boundaries of when you need to do something and when you don't. (Otherwise) the school is babysitting you basically.
You need to learn when to do it on your own, and when not to. I think it being optional is a better idea, people will get more out of it than it being mandatory.”

Unfortunately, Lovelace did not himself feel that intrinsic motivation. Despite earning no “Yes!” indicators on the pre-test, he was only motivated enough to complete one MC during the intervention. Mr. M, reflecting on the small amount of the intervention completed by his classes, stated:

I do think it needs to be mandatory, and I think it needs to have like flexible benchmarking based on where you start…If they start from nothing, by a month into this term they should have any two of the categories done. There needs to be a built-in, you-know-this-is important structure for the student that can't get motivated… At what point are you trying to create intellectual maturity in the students, and at what point are you just admitting that without a decent enough carrot you can't get them to go anywhere?

**Social motivation.** As proposed by the CRKM (Dole & Sinatra, 1998), social context is an important part of an environment that can motivate a learner to engage with a message (the intervention) in order to create conceptual change. This social context in this intervention was created by both the classroom culture and the wall chart that publicly displayed progress or lack thereof.

The social context created in the two teachers’ classrooms was quite different. In the Full Group, the culture around the intervention was full of excitement and fun, as discussed above. Mr. G served as a cheerleader who encouraged his students enthusiastically, creating motivated students who were excited about the intervention and the success they were experiencing. In turn, the students’ enthusiasm was contagious among their peers. It became clear that Mr. G’s classes felt the intervention was worthwhile when students in one of his classes fist-bumped each
other to acknowledge the “awesome job” they did on the intervention. They enjoyed watching the wall chart change as they made progress. One student said it was “really rewarding to get the orange slips off. That was actually weirdly helpful.” Even better, earning a “Yes!” caused even Einstein, who made the least total progress in the Full Group, to declare that he “loved the color change when you got it. It was green!” Galileo agreed that earning a green cell “motivates you to get more.”

Social and external motivation were often closely linked. For example, a few weeks from the end of the intervention, Mr. G added a fun, surprisingly effective donut-based incentive to motivate all students to unlock any remaining categories by using both external and social motivation. He reported,

We were three weeks out from the end of the term and a lot of people had two or three categories left... I said, ‘Would it give everybody an incentive if I brought in donuts for Saturday morning if the entire class, everybody, gets all the way through seven categories and gets them unlocked so you can work on the slips?’ It was unanimous. There was not one person in class who was like, ‘Well, I don't want to have to do that.’ Everybody was like, ‘Oh yeah, we'll make it happen.’ They ended up earning the donuts. I said, ‘Look, there was one person here that didn't make it but I'm going to give that person a donut on loan because I know that he will unlock it eventually.’

In contrast, the culture in the Partial Group was fairly indifferent towards the intervention. Since very few were making significant progress, there was no peer pressure in class to increase the amount of work done. Mr. M was positive about the intervention, but left it up to the students to decide how involved they would get.
**Theme 3: Spaced and interleaved learning.** As discussed in Chapter 2, spaced and interleaved learning is effective but often students don’t like it. The two main aspects of spaced and interleaved learning in the intervention were that students discussed in the focus groups were daily slip limitations and the variety of activities that required spacing to complete.

Many students in both groups did not like the limitation of doing only one slip per MC per day, which ensured the student could correctly solve the same type of problems over at least one week. Multiple students from both the Full and Partial Groups did not like “that we could only do one slip per section per day,” because “then it took so much time.” However, many students also acknowledged that the spacing was effective. A student in the Full Group thought spacing “helped in the long run because I would have forgotten if we could do all of them (the slips) and then move on, but we couldn't.” A student in the Partial Group “liked that you had to do the slips three times, because it really helped you to practice the math.” Mr. G noted “they would always go back to understanding why, but there was a bit of frustration because while they were doing the slips well they wanted to get them all out of the way and move on.” Mr. M questioned whether or not the intervention was spaced enough. He suggested that instead of taking skips on three different days, students should take them over “three different weeks, to create long-term absorption of those skills.”

Students also saw value in mixing in a variety of activities over time. Maxwell (Partial Group) liked “how by doing the slips after watching the videos and taking online quizzes, it kind of forced the ideas to stay in your head a little bit more.” Cauchy (Full Group) agreed:

So I think the fact that you had to go through a lot of different trials and do a lot of things to get a “Yes!” in the box helped a lot. It wasn't just a one-time thing that you would get
right away, because that doesn't seem like it would be as thorough as it was. I think we benefited a lot from that.”

Galileo (Full Group) thought the spacing and variety was worth it, or otherwise you would “forget it the next day because you crammed it all and then you don't remember.”

Hypatia (Partial Group) spoke for many when he addressed both sides of the argument. “It doesn’t matter how many slips we do to get that green box, but it shouldn't be on three different days. Then it just takes up a lot more time, which is annoying...but I understand that the whole point of this is to make sure we know it down by heart and everything. I didn't like it, but I do understand the point of having it on three different days.”

**Theme 4: Video design.** The videos created for this intervention were intended to initiate the process of conceptual change using augmented activation and refutational text, along with creating process mnemonics to help students remember what they learned. Usage of the videos was mixed as were the opinions on their usefulness and general design.

*Augmented activation and refutational text.* Each video began with an activating problem that students sharing a misconception were expected to get wrong. The cartoon narrator asked, “Did you answer (misconception answer)?” The next screen would give a large “Noooooo!” and then warn students that if they got the wrong answer, they should pay close attention to the ideas that followed. Student opinions were mixed on whether or not starting each video this way was effective. Only students in the Full Group discussed this topic during their focus group, perhaps because the Partial Group watched many fewer videos and hence had less to say. Chebyshev liked how the videos gave the wrong answer first “because for me it'd be like, ‘did you think like this?’ I'd be like, ‘oh, yeah I did think like that.’ Then it'd be like, ‘oh, because you were wrong.’ But then it explained how to do it right.”
DeMoivre agreed that it was effective to start out “if you said this, this would be wrong and then you understood our thought process as to how we got that answer and what we can do next time.” However, she also thought that “sometimes seeing the wrong way to do it first was confusing because I'd mix up which way it was right and which way it was wrong.” Euclid didn’t like that the wrong answer was presented first. “The videos obviously helped, but I didn't like that it would go through what you did wrong first. I would have rather seen the correct way to do it so then I could see why the other ways were incorrect.” Einstein appreciated that “even if you got it right they demonstrated how you could get it wrong, what could have been done.”

**Process mnemonics.** A few students brought up the process mnemonics used in the videos. Liebniz and Fourier mentioned “E-X-P-A-N-D,” Pascal mentioned the “Fancy One,” one class discussed “ENDOSA,” and three mentioned “Married vs. Neighbors,” including Euclid who said it “helped me so much.” All but one of the comments were from the Full Group, again most likely because they watched many more segments of video than the Partial Group.

**Other video issues.** Students had many other things to say about the videos. Some liked that they were “easy to follow,” readily accessible on Canvas, and concise. Mobius (Partial Group) really liked “how the videos were short, and very to the point, vs. other videos that can be extremely boring and long, and lose your focus. I like how it kept you in the picture, the whole video.” However, he also though it “was annoying” that you couldn’t skip ahead through the video the first time you watched it. Hippocrates simply thought the videos “were weird.” Brahmagupta appreciated that “within the actual process of getting your boxes checked off you were able to go back to the videos in case you forgot anything, or you wanted to review a topic.”

Hypatia and other students in both groups skipped the videos by exploiting a loophole in Canvas that only required you to access the page rather than watch the whole video by “starting a
video and clicking next.” Jacobi agreed that he and others “gamed the system” to avoid spending time on watching the videos.

Students in the Full Group mentioned they liked the embedded pop-up problems in each video. Aristotle watched the video, and “after the example problem it would be like, ‘Try this on your own,’ and I'd do it. And then it would be like, ‘Did you do this?’ And I'm like, ‘Yes!’ And then it would be like, ‘Nooooo!’” Galileo liked that “in the videos when they had you stop and do problems in the middle because it's why would I watch and then try it. It's keeping actively engaged. You remember things as you go because you're applying them.”

**Theme 5: Recommendations for improvement.** Mr. M. felt strongly that “each of these topics needs a paper packet of things to work on” that includes some self-guided material that explains the concept, has example problems including things that can go wrong, a worksheet with answers, and a quiz. “I think that some of the kids really need a resource to be able to sit down, read, highlight, look at some examples… where they can keep working through it and reinforcing it.” Mr. G wanted to add a “printed piece of paper to hand out with each of the major mnemonics and a quick definition” that they could keep in the front of their notebooks as reference.

Mr. G wanted to add information on slips attempted to the wall chart in addition to correctly completed slips to get a better indication of where students were spending their time, perhaps by adding a dot to represent each attempt. That way, “if they're really good at that category, you would see three dots because they did three and they got three right and they're done, vs. you see ten dots in there.” If you saw many dots but few successful attempts, you could “stop and have a conversation and tell them to watch the video.”
Mr. G was surprised that none of his students Googled their mathematical code name. “They loved their names and they would call them out and they'd take pride in having their name, yet they wouldn't even know when the person lived or what they did.” He recommended adding a small quiz in which they research and write a paragraph about their own code name and two others on the list.

Finally, Mr. G recommended one additional category focused on PEMDAS, the process mnemonic for parentheses, exponents, multiplication, division, addition, and subtraction. As discussed in Chapter 2, this is another well-known area for persistent misconceptions.

**Theme 6: Success stories.** There were many instances of students and teachers reporting successful outcomes from the intervention. In the Partial Group, Mr. M reported some students began to point out things they learned from watching the videos, and that “two of the students who were extremely weak who kept telling the other students how useful the (project was) because it was helping them with the things they kept doing wrong.” He “still sees some random canceling, but it's less. Or now when someone suggests random canceling, another other person says, ‘Oh no, you can't do that!”’

For Mr. G, “at the top level I'm sold on the program and I would like to try it on other math class groups before the Algebra 2 level, like Algebra 1 and Geometry… I'm excited about trying again in the spring.” Many students in the Full Group summed up their successful experience with the intervention during the focus group. DeMoivre commented, “I felt myself getting better on the seven different things as we went throughout the winter term. It was a good way to feel like I was constantly learning.” Mr. G lauded Cauchy’s efforts, since after the intervention “she was not getting tripped up on misconceptions, grasped new concepts faster, and scored higher on assessments. I feel that this program will be instrumental in her confidence and
ability to be successful in precalculus.” Before the intervention, Mr. G explained that Aristotle was struggling with pretty much all of the algebraic misconceptions…she felt totally defeated by the class each time we had an assessment.” However, he added:

in terms of confidence growth for the entire year, I would rank her as #1 out of both of my classes, due primarily to the No-to-Yes program. She was diligent in doing the slips and was excited with each ‘Yes!’ she earned. For the first time in the class, she was seeing a direct correlation between effort and positive gains (before that, she had only experienced difficult effort with poor results). She also saw how these misconceptions were playing a role in completing problems for new topics in the class. By the middle of the spring term No-to-Yes program, she was saying aloud “I got this!” – and she did. She started to show the attitude that “I can do math.”

Aristotle herself similarly summarized:

I really liked my progress. I remember the first test that we took on all of this. I was stumped. I thought it was going to easy because it was math that we'd learned before, but I was stumped on a lot of the questions. And then the one that we just took, it felt a lot easier, and throughout the quizzes and stuff, it was a good refresher and I just felt like I was getting better at a lot of them.”

Finally, Brahmagupta summed up her experience in the intervention by comparing her experience taking the post-test to the pre-test experience:

We just took the second diagnostic test, (I was) just flying through those problems and thinking it was much, much easier than the first time. Like Aristotle said, how before it really stumped me too, and now I could just easily complete most of the problems on that. It felt good.”
Conclusion

Results from the four RQs supported the effectiveness of the NTY intervention both qualitatively and quantitatively. Students in the intervention made fewer misconception errors after the eight weeks relative to the control group, with more engaged students improving more than less engaged students. Looking at the MCs individually, students made significantly fewer errors in four MCs and non-significantly fewer errors in the other three MCs. Students in the focus groups related their experiences relative to time pressure, motivation, and what they liked and didn’t like about the intervention. Success stories from the Full Group gave witness to the power of the intervention to not only decrease misconception errors, but also to increase self-confidence and inspire a more positive attitude towards math.
Chapter 5: Discussion of the Research Findings

This final chapter revisits the problem of practice and the research methodology, then provides a discussion of key research findings from the No-to-Yes intervention related to changes in students’ Misconception Error Scores (MES) and their lived experiences. The findings will be related back to the theoretical framework as well as the literature. Next, the implications for practice, limitations, and significance of the study will be discussed. Finally, the chapter concludes with suggestions for future work related to the intervention.

Revisiting the Problem of Practice

For students to reach their full potential academically and professionally, a strong grounding in mathematics is essential. As shown by the publication of numerous articles on common errors in algebra over the decades (L. Lee & Wheeler, 1989; Marquis, 1988; Smith, 1946; Webber, 1929), many students enter high school with persistent algebraic misconceptions that limit their success in mathematics and, by extension, limit educational attainment and future earnings (Rose & Betts, 2001). To remediate these deeply held misconceptions, a warm conceptual change based intervention was created based on the Cognitive Reconstruction of Knowledge Model (CRKM) (Dole & Sinatra, 1998). The CRKM argues that conceptual change occurs because of student engagement created by the interaction between learner and message. By analyzing the change in scores from pre-test vs. post-test and engagement, this study sought to determine if the intervention created conceptual change in students and if depth of engagement played a role in the depth of change. Overall, the purpose of this study was to assess the effectiveness of a warm conceptual change based intervention on remediating algebraic misconceptions held by students at a private secondary school.
On the first day of the intervention, the pre-test results confirmed that this intervention was desperately needed. Students in Algebra 2 made an average of 15 misconception errors on the 28-question pre-test covering Algebra 1 topics, correctly answering only an average of 46% of the questions correctly. These students had not firmly established the algebraic foundation necessary to succeed in Algebra 2 and beyond. Mr. M expressed his deep belief in the need for this intervention as a result of his experiences teaching high school and college.

I tell students ‘these are fundamental misconceptions that keep plaguing you, and are going to block you in every math class, whether it's stats or calculus, you're only going to get in your own way.’ …These misconceptions have always been there. But I think now they're worse, and at a much deeper level because we have exposed people to symbolic representations at an age before they're ready to do it and appreciate it. And then we just keep pushing them through that path. So we give them this crappy foundation and they don't really understand what they're doing. This year I had students who have just given up… A third of our student body should be in Algebra 1 in the fall (instead of only a handful). I think calling what we offer Algebra 2, but really having it match the curriculum of Algebra 1 plus a couple of units that appear in most Algebra 2 classes, is a disservice to everyone…(until that changes) we need to have that parallel track of working on these foundational skills that are weak. They're going to keep appearing but we need to have something that focuses on that as a primary learning goal… I remember when I was teaching at the college level and the calculus teachers said students were great at concepts, but when it came to just the algebraic manipulation of the answer that they got, all hell broke loose… their algebra skills were terrible. (This is important because) all of the studies show that in both physics and in chemistry in college, your
first college chemistry grade or your first college physics grade are most strongly correlated to your most recent math course grade. And they are not correlated at all to your high school chemistry grade. Nope, zero, nope.

**Review of the Methodology**

This study employed a quasi-experimental, pre-test post-test, control group design as well as student focus groups and teacher interviews. There were three student groups: a control group that only took the pre-and post-test but did not otherwise participate in the intervention, a partial intervention group in which the teacher did not create a culture of participation and hence many of the students engaged only at a low level, and a full intervention group in which the teacher and students were fully engaged in the intervention. Students in all three groups took a pre-test to measure their baseline level of algebraic skills and demonstration of misconceptions. Overall, the goal of the intervention was for students to make fewer misconception errors and demonstrate mastery by earning a “Yes!” in all eight misconception categories (MCs). The eight misconception categories (MCs) were negative function inputs, rules of exponents, distribution of the negative sign, distribution errors, simplification of algebraic fractions, rules of fractions, rules of negative exponents, and a mixed review. For each MC, students earned a “Yes!” in one of two ways: by doing well on the pre-test or by completing the corresponding module. To complete an MC’s module, the student first went online to watch a video and earn 100% on a practice quiz and then had to correctly complete a series of three practice slips in class. Daily progress was tracked on a wall chart. After the eight-week intervention, students in all groups took an identical post-test and the results were analyzed using repeated measures ANOVA, Mann-Whitney U testing, and regression. Focus groups with students were also conducted at the end of the intervention.
Four research questions were investigated, of which the first three were quantitative and the fourth qualitative. RQ1 and RQ2 examined the differences between the intervention groups and the control group in pre- and post-test misconception error scores (MES) overall and by individual MC. RQ3 explored the relationship between engagement and change in MES for intervention group students. Finally, RQ4 used data from focus groups and teacher interviews to explore how the participating students experienced the intervention. For all RQs, the subjects were students in Algebra 2 classes at a private secondary school in New England. The four research questions were:

RQ1: To what extent were there significant differences over time in the MES on a test of algebraic misconceptions among the control group and two intervention groups?

RQ2: For each MC, to what extent were there significant differences in the change in MES among students in the control group and successful students in the intervention group, after excluding students who earned a “Yes!” on the pre-test for that MC?

RQ3: To what extent did engagement predict the depth of conceptual change, as measured by the change in MES among students in the two intervention groups?

RQ4: How did participating students experience the intervention?

**Discussion of Key Findings**

The quantitative and qualitative analyses described in Chapter 4 led to the key findings listed in Table 18. Their relationship to the theoretical framework and the literature are discussed in this chapter. Findings 1-3 address RQs 1-3 respectively while Findings 4–8 address RQ4.
<table>
<thead>
<tr>
<th>Finding #</th>
<th>Finding Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Full Group students made significantly fewer misconception errors after the intervention than either the Partial or control group students. The Full Group’s MES decreased on average approximately five errors more than the Control Group and two more than the Partial Group on the 28-question test. As discussed in Chapter 4, there was a significant interaction between group membership and time, which confirmed there was a measurably different impact among the three groups from pre-test to post-test. Further, while the three groups scored statistically the same at pre-test, there was a significant simple main effect of group at post-test, with the Full Group having a significantly lower MES than both the Partial Group and the Control Group. Finally, while all three groups made fewer errors over time, the effect size was greatest for the Full Group, followed by the Partial Group, and finally the Control Group. These</td>
</tr>
<tr>
<td>2</td>
<td>The intervention’s effectiveness at reducing misconception errors for each MC varied from highly significant to non-significant.</td>
</tr>
<tr>
<td>3</td>
<td>The more deeply a student engaged in the intervention, the fewer misconception errors he or she made.</td>
</tr>
<tr>
<td>4</td>
<td>Students felt the intervention created a great deal of time pressure, both to make progress daily and to finish the intervention in eight weeks.</td>
</tr>
<tr>
<td>5</td>
<td>Three types of motivation were crucial to creating student engagement in the intervention: extrinsic, intrinsic, and social.</td>
</tr>
<tr>
<td>6</td>
<td>A class culture of excitement and engagement in the intervention was vital to sustained progress and conceptual change.</td>
</tr>
<tr>
<td>7</td>
<td>Although students recognized the benefits of spaced and interleaved learning, they disliked it in practice.</td>
</tr>
<tr>
<td>8</td>
<td>As students experienced success in the intervention, they gained self-confidence and developed more positive attitudes towards math.</td>
</tr>
</tbody>
</table>

**Finding 1.** Full Group students made significantly fewer misconception errors after the intervention than either the Partial or Control Group students. The Full Group’s MES decreased on average approximately five errors more than the Control Group and two more than the Partial Group on the 28-question test. As discussed in Chapter 4, there was a significant interaction between group membership and time, which confirmed there was a measurably different impact among the three groups from pre-test to post-test. Further, while the three groups scored statistically the same at pre-test, there was a significant simple main effect of group at post-test, with the Full Group having a significantly lower MES than both the Partial Group and the Control Group. Finally, while all three groups made fewer errors over time, the effect size was greatest for the Full Group, followed by the Partial Group, and finally the Control Group. These
results can be seen in Figures 9-11 in Chapter 4, which show the pre-and post-test MES for each group. Both the descriptive and inferential statistics gave strong support for the effectiveness of the intervention at reducing the number of misconception errors students made. This finding confirms that the intervention overall was worthwhile and benefited students more than taking an Algebra 2 class with no intervention.

**Finding 2.** The intervention’s effectiveness at reducing misconception errors for each MC varied from highly significant to non-significant. Given that some misconceptions are quite “sticky” (Bush & Karp, 2013; Cangelosi et al., 2013; Hewitt, 2012; Knuth et al., 2008; Lucariello et al., 2014; Pitta-Pantazi et al., 2007; Schwartzman, 1977), it was plausible that some MCs would have a greater response than others. Mann-Whitney U tests showed there was a statistically significant decrease in mean rank MES from the intervention group to the control group for MCs 1, 2, 5, and 7, which meant the intervention was effective at reducing the frequency of these errors. For these four MCs, the successful intervention group students decreased their number of errors 43% - 58% more than the control group decreased their number of errors. While the decreases in mean rank MES for MCs 3, 4, and 6 were not large enough for the Mann-Whitney U tests to be significant, in these three MCs the successful intervention group students decreased their number of errors made 30%-56% more than the control group decreased. Overall, in every MC, students in the intervention group improved more than the control group.

Overall, across all MCs, successful intervention group students made fewer errors than control group students. Figures 13 and 14 in Chapter 4 show the frequency of misconception errors made by each group at pre-test and at post-test, and Figure 15 shows the decrease as a
percentage. Table 6 in Chapter 4 shows the relative decrease in MES and significance of the U test for each MC. Highlights of the analysis are discussed below.

Random canceling (MC5) is a notoriously common problem that algebra students make with “comic regularity” (Grossman, 1924, p. 104), so it was not surprising that it was the most difficult category on the pre-test. Students made a total of 216 misconception errors on MC5, representing an average of 3.2 errors per student out of 4 questions. Given the ubiquity of this error, it was gratifying to see that MC5 showed the greatest improvement across all MCs, with the successful intervention group students improving on average 1.86 questions more than the control group ($p = .001$). In the video, students were instructed to find the “Fancy 1” before canceling, thus eliminating the random aspect. Students that used this technique on the post-test usually solved the question correctly. The evidence indicates that the intervention was especially effective at remediating this persistent error.

MC7 (negative exponents) showed the second largest improvement ($p = .002$), with successful intervention group students improving on average almost one full question more than the control group. The video instructed students to distinguish “married” factors from those that were merely “neighbors” to determine how the negative exponent affected them. The video also gave students “cross the line, change the exponent’s sign” as a process mnemonic to help them remember the proper procedure.

MC1 (negative inputs) had a significant $p$-value (.02) and a small mean difference in MES (about a 0.75 question advantage to the Full Group out of four questions). MC1 was the easiest category on the pre-test, with many students earning a “Yes!” and a total of only 62 misconception errors committed, or 0.9 errors per student on average, so there was very little room for improvement. This error was remediated thoroughly in the Full Group, where the
number of errors decreased 82% from 17 to 3, while only decreasing by 5% in the control group. The videos’ instruction to “Reveal the Hidden Parentheses!” had a strong impact on the Full Group.

MC4 (Distribution Rules) had the largest $p$-value (.135) resulting from the relatively similar decrease in errors among all three groups. Control group students improved the most in this category, perhaps due to emphasis from control group teachers as part of their regular coursework. Since regular coursework was effective at decreasing errors in this MC, it could be replaced by a new category in future deployments of the intervention such as a focus on misconceptions related to PEMDAS.

There were smaller sample sizes for some of the $U$ tests, either because many students earned a “Yes!” on the pre-test (MC1) or because few students earned a “Yes!” during the intervention (MC6 and MC7). It was unfortunate that so many students ran out of time before completing MC6 and MC7, thereby weakening the overall analysis. MC7, for instance, had the smallest decrease in MES from pre- to post-test (53%) of all the MCs and also had the smallest number of students who completed it. However, it is not clear if the relatively small decrease in errors was due to its inherent “stickiness,” the way the video taught the lesson, or because students simply ran out of time.

Finding 3. The more deeply a student engaged in the intervention, the fewer misconception errors he or she made. This was shown by the regression model run to test Hypothesis 3A, in which total engagement was significantly predictive of change in MES ($p = .006$). With a significant intercept of -5.366 ($p < .001$), the model predicted a decrease of about five errors from pre-test to post-test for students who had no engagement in the intervention at all. Since control group students experienced a similar decrease of approximately
four errors, the intercept’s five error decrease appears to be a benefit of simply being in an Algebra 2 class over the eight-week intervention and practicing algebra skills daily. Since students in the intervention group made about eight fewer errors overall from pre-test to post-test, the remaining three error decrease was attributed to total engagement in the intervention.

Given that total engagement significantly predicted change in MES, the next step was to examine the relative impact of the individual measures of engagement. The multiple linear regression showed that a model with three predictor variables (videos watched, online quizzes correctly completed, and completely correct in-class practice slips) significantly predicted the decrease in MES ($p = .034$). However, only the intercept had a significant $p$-value ($p < .001$); none of the individual predictor variables’ $p$-values were significant. This result can occur when there is multicollinearity, which was not an issue in this case (see Appendix N). Small sample size is another possible explanation, since most rules of thumb would call for larger than a 41 student sample size with three predictors. For instance, with one common rule of thumb calls for a sample size of least $50 + 3$ times the number of predictors, which in this case would be 59. (Tabachnick & Fidell, 2013). A small sample size would lead to low power, which is the test’s ability to correctly determine a false null hypothesis. Another possibility is that since students learn differently, for some the videos may have been more effective, while for others the slips may have had more impact, thus making it impossible to generalize across all students. It’s also possible the total in this case actually is more than the sum of its parts, so that all the pieces working together created a rich environment that fostered conceptual change.

**Finding 4.** Students felt the intervention created a great deal of time pressure, both to make progress daily and to finish the intervention in eight weeks. Students at this school are very busy and resentful of additional time demands, so teachers were aware they needed to limit
the extra time required for the intervention to alleviate both kinds of time pressure. Both teachers allotted 5-7 minutes of class time daily, with time at the beginning far more effective than at the end. Mr. G kept emphasizing “even if the teacher is engaged, I think you can be fighting an upstream battle if you do it towards the end of the class.”

Creating a strong start was one of the keys to the Full Group’s progress, since an effective kickoff gave students time to progress further overall. To create such a strong start, Mr. G asked students to unlock one category on the first night of the intervention. While not all students followed though, it did help most students to get started. Mr. G also dedicated a full class period early in the intervention to give more momentum to student progress. While it was difficult for Mr. G to find the time to suspend the regular curriculum for a class, he believed it was worthwhile, and the results of his Full Group students supported that belief. Another dedicated class a few weeks later in the intervention would have helped more students finish the intervention on time.

Unfortunately, the students that most needed to do this remedial work were the ones who were most crunched for time, since their algebra skills were weaker and they took longer to do their regular assignments, leaving them with even less time for the intervention. This is a difficult problem to solve but is crucial to the future success of the student, since building on a shaky foundation guarantees the students will never truly master new material. To ensure such students dedicate time to this work in future deployments, a dedicated study hall hour could be created specifically to work on the intervention.

There were more complaints about time pressure from the Partial Group than the Full Group, perhaps because the Partial Group was not fundamentally convinced that learning these skills well would save them time and improve their grades in the long run. Some felt they were
better off studying for a regular test or quiz than doing the intervention. As discussed in the next finding, the Partial Group made much less progress than the Full Group since they lacked motivation.

**Finding 5.** Three types of motivation were crucial to creating student engagement in the intervention: extrinsic, intrinsic, and social. The difference between the Full and Partial Groups in motivation was apparent in both the statistical results and the focus groups. The Full Group was intrinsically motivated, with the external incentive of extra credit points providing an additional carrot. In contrast, the Partial Group was not intrinsically motivated; aside from a few who made good progress, many engaged at the minimum level possible to earn bonus points without really engaging their minds, and there were some who did not engage at all. Students felt the external motivator needed to be stronger if they were to engage in the intervention.

The three types of motivation were highly interrelated in the way they led to greater engagement. In the Full Group, as students walked in and checked their graded slips, they received a boost to all three kinds of motivation: first, they received a bit of extrinsic motivation in the form of bonus quiz points. Next, they experienced social motivation as they together saw what had changed on the wall chart and felt the peer pressure to turn more squares green. Finally, they got a boost to intrinsic motivation as they saw their own success building. In contrast, the Partial Group generated no excitement around doing slips at the end of class, since they saw homework completion as more important. There was no social pressure to make progress, and peer pressure might even have actually discouraged students from doing the work.

To nurture an environment where true learning can occur, it’s vital that the external motivation not eclipse the intrinsic. One student wanted the incentive to be at ten times the current level. While that sounded like a good idea to him and may have spurred an increase in
engagement in the short term, only intrinsic motivation can inspire students to engage deeply in any work to create lasting change.

Finding 6. A class culture of excitement and engagement in the intervention was vital to sustained progress and conceptual change. In both tone and content, the focus groups highlighted the starkly different culture between the two groups. In Mr. G’s Full Group classes, a social culture of engagement, fun, and enthusiasm set the stage for student achievement. Students in the Full Group talked about excitement and fun in the focus groups; they did more work but also enjoyed it more as they experienced success. Mr. G created that culture through his own personal enthusiasm, creation of time for the work, and daily reminders to complete slips. In contrast, the Partial Group students never mentioned fun or excitement. Instead, they focused on making the intervention mandatory and/or raising incentives as the only possible way to get students to do the work. The culture of excitement and engagement in the Full Group enabled a virtuous circle of learning.

Finding 7. Although students recognized the benefits of spaced and interleaved learning, they disliked it in practice. As discussed in Chapter 4, even the students that recognized the need for spacing and interleaving did not like it, since it meant it took longer to complete an MC and they risked forgetting what they had temporarily learned. It was striking that both the successful and unsuccessful students disliked spaced and interleaved practice.

When students attempted MC8, which interleaved all the previous categories together, their confusion made it clear they did not recognize when to apply the techniques they had learned studying each individual MC. Chebyshev, for example, moved quickly through the first seven MCs, completing on average 3-5 slips to get three totally correct and earn a “Yes!” indicator. However, in MC8, she had a great deal of trouble correctly solving problems that
combined three random MCs. She turned in eight attempts but only got one completely correct. Once the categories were mixed, she had a hard time identifying which techniques to use. Intentionally focusing student attention on discerning which MCs are present in each problem may help students in future deployments complete MC8 as well as give them an effective study technique for related work outside the intervention.

**Finding 8.** As students experienced success in the intervention, they gained self-confidence and developed more positive attitudes towards math. For Mr. G, this finding was “one of the most exciting results for me as a teacher.” Mr. G shared:

> there were three cases that stood out in particular – one of my top students, one mid-range student, and one struggling student. In each case, they put the work into the program and saw the connection between work and results. They gained a new confidence that they could master these math topics and math in general through a strategic way of practicing problems.

This finding is particularly important for struggling math students, who often developed a negative fixed mindset about math in middle school and often feel as if they are and always will be “bad at math” no matter how hard they try. These students often stop taking math classes as soon as allowed by graduation requirements. Through full participation in this intervention, students will not only improve their skills but also improve their self-confidence and attitude towards math, sparking a mindset change that will open paths involving future study in STEM fields.

**Discussion of Findings in Relation to the Theoretical Framework**

Dole and Sinatra’s (1998) Cognitive Reconstruction of Knowledge Model (CRKM) posits that a learner’s interaction with a message leads to engagement, which in turn causes
conceptual change. The greater the level of engagement, the greater the degree of conceptual change that can occur. Key characteristics of the learner and the message (the intervention) contribute to engagement, as can a peripheral cue (see Figure 1 in Chapter 1). While the intervention did not touch upon all aspects of the CRKM, findings from both the focus groups and the quantitative analysis supported the model’s central tenet that deeper levels of engagement led to greater conceptual change.

Findings 1 and 3 gave evidence that engaged students decreased the number of errors they made more than less engaged students did. In Finding 1, the more engaged Full Group improved significantly more than either of the less engaged Partial or Control Groups. Further, Finding 3 directly tested the link between engagement and a decrease in errors and found a significant relationship. It also makes good common sense that students will only change their conceptions when deeply engaged rather than superficially active.

Two of the characteristics of the learner discussed in the CRKM are motivation and social context, which are relevant to Findings 5-7. In Mr. G’s Full Group, there was a class culture of excitement and engagement that led to greater motivation and increased self-confidence as well as a compelling social context created by the wall chart and positive peer pressure. This environment created an ideal setting that helped the learner interact more intensely with the message (the intervention), leading to deeper engagement and greater conceptual change. In contrast, in Mr. M’s Partial Group, only a few students engaged meaningfully with the intervention, perhaps due to other personal characteristics such as Personal Relevance, Dissatisfaction, or Need for Cognition. These characteristics could be studied in future research.
The CRKM also posits that a convincing message (the intervention) should be comprehensible, coherent, plausible, and rhetorically compelling. These characteristics were intentionally designed into the intervention, especially the videos, with their focus on augmented activation, refutation text, clear examples, and catchy process mnemonics. While the students were not directly asked about these characteristics in the focus groups, there was evidence that they felt these qualities existed in the intervention. For example, in the focus groups, students mentioned how they were surprised when they answered the opening question incorrectly, thus having their misconceptions activated. They cited the usefulness of the videos overall and deemed them clearly constructed and helpful.

Lastly, the CRKM adds the “peripheral cue,” a compelling, attractive source or simple message, as a route to engagement. In the intervention, the process mnemonic served as the peripheral cue for each MC. On their own, process mnemonics can lead to shallow understanding, superficial change, and misapplication, such as the classic problems with application of the PEMDAS mnemonic discussed in Chapter 2. However, when combined with a properly designed message, the process mnemonic can serve as another route to draw students in to deeper engagement. In the focus groups, students mentioned the usefulness of some of the process mnemonics in the videos, such as “E-X-P-A-N-D,” finding the “Fancy One,” “ENDOSA,” and “Married vs. Neighbors” (see Appendix I). When properly integrated, these peripheral cues can lead to greater conceptual change.

Discussion of Findings in Relation to the Literature Review

Chapter 2 of this study presented a literature review with the following threads: diagnosing misconceptions, types of mathematical misconceptions, conceptual change models, mathematics classroom interventions, spaced and interleaved learning, and process mnemonics.
The literature review informed all aspects of the intervention’s design, including which misconceptions to address, how to design the pre-test instrument and the various components, how to foster conceptual change, and how to space the intervention across time. Findings from the intervention both confirmed previous work discussed in the literature and began to fill some of the identified gaps.

**Diagnosing misconceptions.** The literature evaluated a number of design options for diagnostic assessments, including multiple choice tests with distractors, free-response questions, and interviews (Almog & Ilany, 2012; Poon & Leung, 2010; Russell et al., 2009). However, none of the existing diagnostic instruments were designed to target the specific misconceptions in this study. To fill this gap, after examining the advantages and disadvantages of each method, the researcher created a diagnostic instrument that employed tightly focused free-response questions designed to elicit a misconception response. In addition, multiple-choice questions with distractors were used in the quizzes and embedded pop-up video questions.

**Types of misconceptions.** The literature also revealed that certain mathematics misconceptions are quite common and exceptionally persistent, with teachers as far back as the early 1900s noting the same types of errors (Grossman, 1924; Smith, 1946; Webber, 1929). Based on the literature review, the misconceptions chosen for the intervention were use of notation, use of the negative sign, rules of positive and negative exponents, distribution errors, and simplifying algebraic fractions. The diagnostic pre-test and subsequent written work confirmed that students in the intervention demonstrated all of the common misconceptions identified in the literature.

**Conceptual change models.** The strong support in the literature for using augmented activation (AA) and refutational text (RT) to create cognitive conflict (Guzzetti et al., 1993;
Tippett, 2010) led to their incorporation in the intervention’s videos. Like the interventions in the literature, the study’s overall success at reducing misconception errors gave evidence that AA/RT was an effective way to induce students to change their conceptions. While it is difficult to separate the effects of AA/RT from the rest of the components in the intervention, the analysis for RQ3 showed that the video predictor was the most significant of the three components in the intervention. Students who had their prior misconceptions activated and refuted in the videos had begun the process of conceptual change and could then use the quizzes and slips to practice and internalize the new, correct conception. Those who did not watch the videos missed the conceptual change lesson which made it more difficult to change the way they thought about the problem.

**Mathematics classroom interventions.** Although there is a rich body of existing research that categorizes students’ algebraic misconceptions, there are very few classroom interventions designed to remediate such misconceptions and even fewer that employ conceptual change models to do so (Adams, 1998; Bokhove & Drijvers, 2012; Koparan et al., 2010; Kramarski, 2004; Moschkovich, 1998; Russell et al., 2009; Van Dooren et al., 2006). Only two studies used refutational text as part of a math classroom intervention (Gill et al., 2004; Lem et al., 2015) and only one (Gill et al., 2004) that used a warm conceptual change model. By studying warm conceptual change theory incorporating AA/RT in a mathematics classroom setting, the intervention addressed this obvious gap in the literature.

**Spaced and interleaved learning.** The literature consistently shows that spacing learning is a simple, robustly effective technique that can increase long term retention (Dunlosky et al., 2013; Kapler et al., 2015; Mettler et al., 2016). Intervention group students were required to complete three different practice slips correctly on three different days, which built spaced
learning into the intervention. Interleaved learning can also help students retain information over the long term by mixing topics together (Rohrer & Taylor, 2007), although the research is mixed. Students often do worse in interleaved practice sessions and often tend not to like the way interleaved practice confuses them, even if they end up retaining more in the long term (Taylor & Rohrer, 2010). Focus groups showed the same effect in this study, with students reporting they did not like interleaved learning even though they understood how it could be helpful. The literature also suggested that a mixed practice schedule might be best, with massed learning followed by interleaved practice (Carpenter & Mueller, 2013). This was the method employed in the intervention, where students first practiced the techniques in isolation (massed) until they were mixed together in the 8th category (interleaved). This proved quite challenging and frustrating for students, confirming previous work in the literature.

**Process mnemonics.** The final thread in the literature review dealt with the benefits and drawbacks of process mnemonics. The literature indicated that process mnemonics could be a useful tool in helping students retain learning over time (Manalo, 2002), though there were only two studies that used process mnemonics in a math classroom (Manalo et al., 2000; Test & Ellis, 2005). The intervention addressed this gap in the literature by adding a process mnemonic to each misconception category to help students remember what to do over the long term. In focus groups, students said the mnemonics (such as “Cross the Line, Change the Exponent’s Sign”) were very useful in helping them remember the proper technique. Although the literature warns that process mnemonics can be misapplied (Hewitt, 2012), no evidence surfaced in any student work in this intervention that showed misapplication of any of the process mnemonics.
Implications for Current Practice

The findings of this study clearly indicate that students benefited from this intervention. The recommendation is for all Geometry and non-honors Algebra 2 classes at the researcher’s school to deploy the intervention during the winter term, with a possibility of repeating the intervention in the spring to further solidify learning. All future deployments should put into practice the best practices uncovered in this research, including:

- Making a convincing case for the long-term benefits of the intervention at the outset to increase students’ intrinsic motivation.
- Creating a strong start by assigning homework on the first night to unlock one MC.
- Dedicating 5-7 minutes at the start of each class to work on the intervention.
- Fostering a strong class culture of enthusiasm and positive peer pressure.
- Providing ½ class period on the kickoff day followed by one full class period at about one week and one at about three weeks to keep momentum going.
- Allowing the intervention to span an entire 10-week term, which would allow slower students to finish.

As more students participate in the intervention, the language of the process mnemonics should be socialized among all math faculty so they can reinforce the ideas in the videos, which will be made available for students to re-watch at any time.

Limitations of the Study

Like all research, this study had limitations that must kept in mind when interpreting the results. Key limitations included statistical constraints, methods of measuring key constructs, implementation issues, and time constraints. Each of these limitations is discussed in greater detail below.
First, the design of the study limited the statistical conclusions that could be drawn. Since intact classes were assigned to treatment groups, this was only a quasi-experimental study instead of a random assignment of students to groups. This limited the ability to generalize to a larger population and made conclusions about causality more tentative. Further, since each teacher either taught all intervention or all control group classes, there was no way to control for the variability introduced by the teacher. By assigning classes to treatment groups this way, the effect of the teacher was confounded with the effect of the intervention such that it was not possible to determine the impact of a teacher on post-test scores through means unrelated to the intervention. Another statistical issue was the nature of the convenience sample drawn from the researcher’s own unusual school setting (a private boarding school). However, as discussed in Chapter 3, there is strong reason to believe that algebra students at this school are similar to algebra students at other schools and that all share the same misconceptions. Other possible statistical limitations included assessing the reliability and validity of the researcher-written instrument and the relatively small sample size which limited the power of statistical tests.

The focus group format was limiting in some ways. Some students may not have been comfortable expressing their true opinions in front of peers, especially if they ran contrary to the opinions of the more vocal students. Some students chose not to participate at all. While each group had the same scripted questions, the conversation in each group ended up taking different paths, so not all groups discussed the same topics. The focus groups were not as in depth as individual student interviews would have been, though they did reveal each class’s culture relative to the intervention.

Deep engagement, the most crucial construct in the CRKM, was quite difficult to measure in practice. Merely counting the activity level of each student was not adequate, as
students could do many slips and quizzes without ever thinking deeply about the misconception. One student, Einstein, did just that, completing 24 quizzes and 32 practice slips while earning only three “Yes!” indicators. His teacher, Mr. G, noted that Einstein’s superficial level of engagement did not lead to a great deal of progress. In contrast, one of the most successful students, Chebyshev, completed only 12 quizzes and 34 practice slips while earning seven “Yes!” indicators. One difference between them was that Chebyshev had one of the highest number of segments watched (80) while Einstein watched a below average number of video segments (56). As a result of this difficulty, instead of measures that counted activity levels, only activity with a successful result was counted: completely correct video pop-up questions, quizzes, and slips. However, measuring engagement this way was still problematic, since it was possible for a student to be engaging deeply but not achieving the real conceptual change that would have led to decreased errors on the post-test. In the future, a measure of student self-report of engagement might be useful to include, though students are not necessarily accurate self-reporters.

Another limitation of the study was that Canvas could not prevent a student from skipping past a video once the link was clicked so that they proceeded directly to the quizzes and practice slips. As a result, it cannot be assumed that everyone who earned a “Yes!” actually watched the video designed to initiate conceptual change. In the future, technology may allow for a way to require the entire video be watched before other aspects of the intervention are unlocked. Teachers can also make a stronger case to students about the importance of the videos.

Finally, the study was limited by time. Given that only two Full Group students completed the entire intervention before the term ended, it was evident that more time was
needed. It is plausible that the differences among the groups would have been even more
dramatic had all students in the Full Group finished the intervention.

**Significance of the Study**

Students continue to struggle with math proficiency. In October 2017, New Hampshire’s DOE reported that the percentage of students scoring “proficient” (at grade level) or higher on the 2016-2017 Smarter Balanced assessments dropped in New Hampshire (NH Department of Education, n.d.). In grades 3-8, math proficient scores were down 1-3%, ranging from 45% of 8th graders to 55% of 3rd graders. In 11th grade, 96% of the New Hampshire’s class of 2017 took the SAT but only 46% met the math benchmark, which is the score associated with a 75% chance of earning at least a C in an entry-level, first-semester college-level course in algebra or beyond (“SAT suite of assessments 2017 NH,” n.d.). While standardized tests themselves are arguably flawed and difficult to interpret, it is nonetheless discouraging that roughly half of this state’s younger students are failing to meet grade level expectations and only half of our graduating seniors are prepared to even have a chance at earning a C in an entry level college math course.

In the long term, we must re-examine mathematics education from the earliest years so that students learn mathematics correctly and joyfully and are prepared for future studies. In the meantime, however, it is vital that students who have not learned proper mathematics in middle school be helped over their misconceptions so they have a chance to succeed in mathematics, and by extension, reach their full potential in college and the workforce. This intervention is one way to help students overcome their persistent misconceptions, gain confidence, and develop a more positive, growth-oriented mindset towards mathematics. By deploying interventions such as this one, math educators can help students not only graduate from college, but also to
maximize their lifetime earnings, contributions to society, and the global competitiveness of the U.S (Rose & Betts, 2001).

Conclusion

The purpose of this study was to assess the effectiveness of a warm conceptual change based intervention on remediating algebraic misconceptions held by Algebra 2 students at a private secondary school in New England. The following four research questions guided this study:

1. To what extent were there significant differences over time in the misconception error scores (MES) on a test of algebraic misconceptions among a control group and two intervention groups?

2. For each MC, to what extent were there significant differences in the change in MES among students in the control group and successful students in the intervention group, after excluding students who earned a “Yes!” on the pre-test for that MC?

3. To what extent did engagement predict the depth of conceptual change, as measured by the change in MES in the two intervention groups?

4. How did participating students experience the intervention?

The researcher employed a quasi-experimental, pre-test post-test, control group design as well as student focus groups. There were three groups of students: a Control group, a Full intervention group, and a Partial intervention group. All students took a pre-test to measure their baseline level of algebraic skills and demonstration of misconceptions. After the 8-week intervention, students in both groups took an identical post-test and the results were analyzed statistically. Focus groups were conducted and coded for themes using an inductive method.
The results of the quantitative and qualitative analyses showed the intervention was worthwhile because it benefited students more than taking the same Algebra 2 class without the intervention. Full Group students made significantly fewer misconception errors after the intervention than either the Partial or Control Groups. Further, the more deeply a student engaged in the intervention, the fewer misconception errors he or she made. The intervention’s effectiveness at reducing misconception errors for each MC varied from highly significant to non-significant, so targeted efforts can be made to improve the delivery for the least significant MCs. It was clear that extrinsic, intrinsic, and social motivation were crucial to creating student engagement in the intervention, and that a class culture of excitement and engagement in the intervention was vital to sustained progress, conceptual change, an increase in self-confidence, and a more positive attitude towards math. Students felt the intervention created a great deal of time pressure, both to make progress daily and to finish the intervention in eight weeks. Finally, although students recognized the benefits of spaced and interleaved learning, they disliked it in practice. Taken together, the findings support the premises of the CRKM and show real progress can be made in helping students change their misconceptions by using this intervention.

**Future Work**

With its inherent complexity, this project is rich with opportunities for future research, improvements, and deployments. All the avenues of future work discussed below would benefit students or advance the knowledge of the field related to conceptual change or overcoming mathematical misconceptions.

**Future research.** There are many directions future research could take, depending on whether the researcher was interested in a more robust statistical study, further testing the
CRKM, seeking to better understand the student experience, or investigating long-term retention.

There are a number of options for researchers interested in improving the statistical robustness of this study. A larger sample size could be used to increase the power of statistical tests. Students could be randomly assigned to groups to have a fully experimental design. By adding a control group for each teacher, the confounding of teacher and intervention could be prevented. An experimental design could be conducted in which students are assigned only certain components of the intervention to explore their relative effectiveness. Finally, the intervention could be deployed in a wide range of schools to assess the external validity of the results.

To better understand the construct of engagement, the researcher could add surveys or questionnaires for students to self-assess engagement or replace focus groups with student interviews. Another possibility is the direct observation of students while doing the work, possibly employing a think-aloud protocol to gain insight into the student’s thinking and depth of engagement.

If the researcher wanted to study other aspects of the CRKM, there are multiple areas available to explore. Other characteristics of the learner in the CRKM could be studied, such as Need for Cognition, Dissatisfaction, and Personal Relevance. The researcher could target data-gathering around student’s opinions of the message’s characteristics as theorized in the CRKM (comprehensible, coherent, plausible, rhetorically compelling). The role of the peripheral cue could be studied in more depth, perhaps comparing student reactions to an intervention with and without the cue or substituting a different cue, such as having the rector of the school talk to the class about the importance of the intervention.
Another vital area for research is to test the longitudinal effect of the intervention, since it is only worthwhile if the benefits are long-lasting. For this study, the researcher could administer the post-test again to the same cohort of students after one year (in February 2018). Results could be compared to each student’s original pre-test and post-test to see if any gains were retained.

Finally, the intervention could be rewritten to address an entirely different set of misconceptions. It is reasonable to expect that the same intervention structure could work to change conceptions across any field of study, such as in Physics where there is a well-established set of common misconceptions.

**Future improvements.** Through teacher observations, student focus groups, and teacher interviews, the following recommendations to improve the intervention for future deployments were gathered:

- Add a paper worksheet for MCs 1-7 with instruction, process mnemonics, and practice problems that would be handed out after a student unlocked a category online. Refutational text has been shown to work better when combined with worksheets (Korur et al., 2015).
- Replace MC4, which showed the smallest improvement, with a new category directly related to the misapplication of PEMDAS.
- Modify slips for MCs 1-7 so that they require students to list the relevant process mnemonic before doing the problem. Such “self-explanation prompts” in which students first choose the relevant principle or strategy has been shown to effectively and efficiently increase both near and far transfer of learning (Atkinson, Margaret Merrill, & Renkl, 2003).
• Add self-explanation prompts to the MC8 slips so that students must first choose the relevant techniques from check boxes before doing the work. Due to space and time constraints, the number of problems on each slip may need to be reduced to two from three.

**Future deployments.** Given the evidence that the intervention is effective, it should be deployed widely to reach as many students as possible so they can reach their full potential in math. To begin this process, the researcher created a No-to-Yes Project website (notoyes.org) that contains the intervention’s materials. Any educator can create a “Free to Teachers” Canvas account (“Canvas free for teachers account: Registration and login,” n.d.) to access the videos and quizzes to use with their own students. The next step is to seek out as broad an audience as possible to try the intervention, perhaps starting with other local high schools.

If teaching in elementary and middle schools could be changed to ensure students never formed these misconceptions in the first place, this intervention would not be needed. To achieve this desirable state, the intervention could be deployed to schools of education. If future teachers are explicitly taught about these common, persistent errors, they could both remediate any of their own misconceptions as well as help their future students avoid forming them. This would benefit math students immensely and create a new generation of students prepared to succeed in high school mathematics and beyond.
References


Combining operations (distributive laws). (n.d.). Retrieved August 16, 2016, from
http://brownmath.com/alge/distrib.htm


Construct Validity. (n.d.). Retrieved May 7, 2016, from
http://www.socialresearchmethods.net/kb/constval.php


https://doi.org/10.1080/00220671.2014.946121


President’s Council of Advisors on Science and Technology. (2012). Engage to excel producing one million additional college graduates with degrees in science, technology,
engineering, and mathematics. Executive Office of the President, President’s Council of
https://doi.org/10.1037/edu0000001
*Instructional Science, 35*(6), 481–498. https://doi.org/10.1007/s11251-007-9015-8
graduation, and earnings. Retrieved from
http://www.ppic.org/content/pubs/report/R_701JBR.pdf
algebra: Results from an experimental pilot study. *Behavior Research Methods, 41*(2),
SAT suite of assessments 2017 NH. (n.d.). Retrieved from
https://reports.collegeboard.org/pdf/2017-new-hampshire-sat-suite-assessments-annual-
report.pdf
interactions to foster conceptual change: Designing and evaluating an instructional

https://doi.org/10.1002/sce.20037


https://doi.org/10.1207/s15326985ep4002_5


Unintended consequences: More high school math, science linked to more dropouts. (2014). *States News Service*.

https://doi.org/10.1016/j.learninstruc.2004.06.019


https://doi.org/10.1016/j.learninstruc.2004.06.014


https://doi.org/10.2307/3027549

https://doi.org/10.1111/j.1949-8594.2012.00136.x


Appendices

Appendix A: Teacher Recruitment Email

Dear Algebra 2 teachers,

I hope your fall term is off to a great start! As you may know, I have finished my coursework for my doctoral work at Northeastern and am now ABD, or “all but dissertation.” This is where you come in.

I’m inviting you to participate in my research study on the No-to-Yes intervention, tentatively titled “The Impact of an Intervention on the Persistence of Algebraic Misconceptions among Secondary School Algebra Students.” The purpose of the study is to evaluate the effectiveness of the No-to-Yes intervention on overcoming the persistent algebraic misconceptions that many of our students have. The intervention will have a Canvas component (videos and practice quizzes) and a classroom component (practice slips).

Based on teacher interest, some classes will be in the experimental group and some in the control group. Requirements during the 6-8 week winter term intervention include:

- Control group: Give a pre-test in November and a post-test in February or March
- Experimental group:
  - Attend training in early November (about 30 min), at a convenient time and place
  - Give the pre-and post-test
  - Oversee the class portion of the intervention, which consists of giving 5-10 minutes daily for students to complete practice slips and logging any questions students ask about the misconceptions
  - Allow me to conduct a focus group in your classes with you present (no more than 30 minutes) to debrief the intervention. Parental consent will be obtained for these focus groups.

** Please note that I will be doing ALL the preparation, copying, grading, and chart updating; basically you just need to allow time for slips, log questions, and attend the focus group **

Your decision to participate in this study is completely voluntary and your identity in the study (including your name and student names) will be completely confidential. Only pseudonyms will be used. There are no known risks for participating. Please note that if the intervention should be successful, control group teachers will be highly encouraged to administer the intervention in the spring term.

This work should benefit you and your students. First, the intervention should reduce the frequency that students make the targeted errors and strengthen their growth mindset, making your job easier. Second, the intervention gives you a shared vocabulary for talking about the errors. Finally, the intervention framework can be customized for other kinds of problems you identify.

Please reply to this email to let me know if you are interested in participating. I am happy to answer any questions you may have. Thanks so much for considering!

Sue
Appendix B: Teacher Signed Consent Form – Experimental Group

Northeastern University, College of Professional Studies

You are being invited to take part in a research study. This form will tell you about the study, but the researcher will explain it to you first. You may ask the researcher any questions or concerns that you have. When you are ready to make a decision, you may tell the researcher if you would like to or not to participate. You do not have to participate if you do not want to. If you do decide to participate, you will be asked to sign this statement. Upon signature, you will receive a copy to keep for yourself.

Investigator: Dr. Kelly Conn, Principal Investigator, and Susan Zielinski, Student Investigator

Title of Project: The Impact of an Intervention on the Persistence of Algebraic Misconceptions among Secondary School Algebra Students

Why are you being asked to take part in this research?

You are being asked to participate in this study because you are a non-Honors Algebra 2 teacher at the researcher’s school and have expressed interest in participating in the experimental group.

Why is this research being done?

The purpose of this research study is to evaluate how the No-to-Yes intervention impacts the frequency of errors based on algebraic misconceptions in our students. The goal is to help students get better at algebra and hence be more successful in academia and beyond.

What will I be asked to do?

Teachers in the experimental group will:

- Attend training in early November at a convenient time and place for you.
- Give the pre-and post-test give a pre-test in November and a post-test in February or March.
- Oversee the class portion of the intervention following study protocols, which consists of giving 5-10 minutes daily for students to complete practice slips and logging any questions students ask about the misconceptions
- Allow me to conduct a focus group with your students during regular class time (with you present) to debrief the intervention. The focus group will last no longer than 30
minutes and will be audiotaped for transcription purposes only. Parental consent will be obtained prior to the focus group.

Where will this take place and how much time will it take?

The training will last no more than 30 minutes and will occur at a convenient time and place for you. The intervention, testing, and focus group will take place in your classroom during regular class time. The pre- and post-test will both take approximately 30 minutes each. The classroom portion of the intervention should take 5-10 minutes daily over 6-8 weeks in the winter term. Finally, the focus group will take no more than 30 minutes.

Will there be any risk or discomfort to me?

There is no foreseeable risk or discomfort.

Will I benefit by being in this research?

There will be no direct benefit to you for participating in this study. The potential benefits for your students include improved algebra skills that will lead to greater success in math classes, greater participation in STEM fields, and even increased future earnings.

Who will see the information about me?

As a participant in this research, your part in this study will be confidential. Only the researchers will see the collected data and information about you. No reports or publications will use information that can identify you or your classes. A pseudonym will be used to protect your identity. The data collected for this study will be kept by the researcher in password protected files and will not be shared with others. Written work will be kept in the researchers’ locked desk drawer or home office. All documents and data containing real names including audiotapes will be destroyed within six months of the final approval of the dissertation.

In rare instances, authorized people may request to see research information about you and others involved in this study. This is done only to be sure that the research is done properly. The researcher would only permit people who are authorized by organizations such as the Northeastern University Institutional Review board to see this information.

Can I stop my participation in this study?

Your participation in this research study is completely voluntary. You do not have to participate if you do not want to. Even if you begin the study, you may quit at any time. You may also refuse to answer any interview questions. If you do not participate, or decide to resign, you will not suffer any negative consequences.
Who can I contact if I have questions or problems?

Student Investigator: Susan Zielinski, Doctor of Education Student
325 Pleasant St.
Concord, NH 03301
603-969-2720
Email: zielinski.s@husky.neu.edu

Principal Investigator: Dr. Kelly Conn, Assistant Teaching Professor
College of Professional Studies
Northeastern University
360 Huntington Avenue, 20 BV
Boston, MA 02115
857-205-9585
Email: k.conn@northeastern.edu

Who can I contact about my rights as a participant?

Inquiries regarding your rights as a participant may be referred to Nan C. Regina, Director, Human Subject Research Protection, 960 Renaissance Park, Northeastern University, Boston, MA 02115. Tel: 617-373-4588, email: irb@neu.edu. You may call anonymously if you wish.

Will I be paid for my participation?

There is no compensation for being a participant in this research study.

Will it cost me anything to participate?

There is no cost to participate in this research study.

I have read, understood, and had the opportunity to ask questions regarding this consent form. I fully understand the nature and character of my involvement in this research as a participant and the potential risks. I agree to participate in this study on a voluntary basis.

____________________________________________________  __________________
Research Participant (signature)  Date

____________________________________________________
Research Participant (printed)

____________________________________________________  __________________
Researcher who explained the study to the participant above and obtained consent (signature)  Date

____________________________________________________
Researcher Name (printed)
Appendix C: Focus Group Protocol

**Purpose:** To obtain student feedback on the NTY intervention. What worked well? What didn’t work well? What could have been better?

I will conduct a focus group session in every experimental group classroom (3-4 classrooms) with the classroom teacher as observer. All students (typically 8-10 in each class) will be asked for verbal consent. The location will be the students’ regular classroom with chairs arranged in a circle.

**Introduction:** “Thank you for allowing me to join your class today! My name is Ms. Zielinski, and I am a student at Northeastern University working on my doctoral dissertation. My study is about the No-to-Yes project, which you just finished. The project will help me understand how to help students make fewer algebra errors.

Today I am conducting a group discussion that should take approximately 30 minutes. We’ll be using the talking circle process that you all have practiced in your dorms. Remember the rules: talk only when you have the talking piece, respect others’ opinions, and it’s OK to pass. We’ll also have time for free discussion.

Your participation is voluntary. If you do not wish to participate, you may pass when the talking piece comes around to you. Taking part in the discussion indicates your agreement to participate. Responses will be kept confidential and your real name will not appear in the final write-up. There are minimal risks associated with this discussion. I am recording the discussion so it can be transcribed, and I will destroy the original recording after the final report is approved. Finally, please do not share anything you are uncomfortable discussing with the group.

If you would like a copy of this information, please let me know and I will email it to you. If you have any questions or concerns you can contact Ms. Zielinski at 603-969-2720 or zielinski.s@husky.neu.edu. You may also contact Dr. Kelly Conn, the Northeastern University advisor for this study, at 857-205-9585 or k.conn@northeastern.edu. If you have questions but want to talk to someone else who is not a part of the study, you may contact Nan Regina, Director, Human Subject Research Protection, at 617-373-4588, or n.regina@neu.edu. You may call anonymously if you wish.”

**Questions:** “For each question, we will go around the circle with the talking piece and then open up the discussion. I may ask follow-up questions while a student has the talking piece. To remind you, the No-to-Yes intervention was made up of the pre- and post-test, Canvas videos and quizzes, practice slips, and the classroom wall chart.”

1. Can you share an example of what you liked about the No-to-Yes project?
2. Can you share an example of what you didn't like about the No-to-Yes project?
3. Can you describe how you used the Canvas videos and quizzes during the project?
4. How would you recommend we improve the No-to-Yes project?

**Closing:** “Thank you for participating! I will be analyzing the data for common themes and suggestions. Please contact me if you have any questions about the study or the discussion.”
Appendix D: Student Instructions

By trying hard and engaging in these modules, you will be able to fix some of the most common errors algebra students make. It takes effort to change your brain, but you’ll see the payoff in better grades, lower frustration levels, and (yes!) even more enjoyment of math.

Your pre-test results show you which modules you need to work on. Look at your row on the wall chart in your class. If the module is labeled “Yes!” you got all four related questions correct on the pre-test - congrats! If it has an orange sticky note, you missed some of the questions, so you need to do more work, starting with unlocking the module online. If it’s white (blank) or has dates written in it, you’ve already unlocked it online and can work on the Nocoos! Slips in class.

To earn a "Yes!” for a module, you need to:

1. Complete the online (Canvas) portion to remove the orange sticker
   - First watch the video all the way through (as many times as you want), answering any pop-up questions. The first time watching, you can’t skip ahead! (Here’s some help with logging in.) Then,
   - Earn 100% on the online practice quiz
     - Take each practice quiz as many times as you want
     - You’ll get different questions each time
     - If you get a question wrong, scroll down & study the explanation, then try again

2. After the module is unlocked online, complete the paper Nocoos! Slips for this module in class
   - For each module, you must get 3 Nocoos Slips completely correctly to earn a “Yes!”
   - These 3 Nocoos slips must be completed on 3 different days...and
   - you must get 3 different versions correct (there are 4 versions of each slip)...and
   - You can only complete ONE slip per module per day!
   - So, if you have 4 modules unlocked (orange slip removed), you can do 4 slips in class, one for each module!

If at any time you are stuck or keep getting problems wrong, you can:
   - Watch the video again
   - Take more practice quizzes
   - Ask your Algebra 2 teacher or Ms. Zielinski for help

You should be working on more than one module at a time, and modules can be completed in any order (though it may be easier to go in order).
Appendix E: NTY Test Instrument (Pre- and Post- Test)

Note: instead of category numbering, students received a test numbered #1-28.

No-to-Yes!

Simplify, if possible, leaving no parentheses and no negative exponents. Any answer with fractions should contain only one simplified fraction. If no simplification is possible, write “None.” Show your work. No calculators. You can do this!!

<table>
<thead>
<tr>
<th>5b. (\frac{2x+6}{2x-10})</th>
<th>5a. (\frac{6+a}{-12a})</th>
</tr>
</thead>
<tbody>
<tr>
<td>4a. ((3 + x)^2)</td>
<td>6b. (\frac{5}{a} \cdot \frac{6b}{a})</td>
</tr>
<tr>
<td>6a. (\frac{3}{a} + \frac{3}{b})</td>
<td>2b. ((2x^2)^4)</td>
</tr>
<tr>
<td>7a. ((3x)^{-1})</td>
<td>5c. (\frac{4}{x+4})</td>
</tr>
<tr>
<td>2a. (x^2(x^4) + x^3)</td>
<td>6c. (\frac{5}{a} \div \frac{6a}{b})</td>
</tr>
<tr>
<td>3c. (-(6x^2 - 5x + 3) - 2)</td>
<td>7c. ((-2x)^{-3})</td>
</tr>
<tr>
<td>4b. (\sqrt{x^2 + y^2})</td>
<td>6d. (\frac{3}{a} \cdot \frac{b}{2} \div \frac{3}{a} \div 4)</td>
</tr>
<tr>
<td>7b. (-\frac{1}{2x-3})</td>
<td>5d. (\frac{3x+6xy}{9x^2+30xy})</td>
</tr>
<tr>
<td>7d. (-\frac{3x^{-4}y^2}{(2x)^{-3}y^{-5}})</td>
<td>4c. (3(2ab)(4ab))</td>
</tr>
<tr>
<td>2c. (\frac{4x^6}{x^2} + 5x^4)</td>
<td>2d. (\left(\frac{x^4x^2}{9y^2}\right)(2y^2)^3)</td>
</tr>
<tr>
<td>3d. (-2(-(3x - 5)) - 10 - (3 - x))</td>
<td>4d. (2(x + 1)(3x))</td>
</tr>
</tbody>
</table>

Evaluate, leaving no parentheses:

<table>
<thead>
<tr>
<th>1a. (x^2), when (x = -2)</th>
<th>1c. (x^2 - x^3), when (x = -2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3b. (a - b), if (a = -3x - 6) and (b = -3x^2 - 4x + 2)</td>
<td>3a. (a - b), if (a = 3x + 6) and (b = 4x - 2)</td>
</tr>
<tr>
<td>1b. (-x^2), when (x = -2)</td>
<td>1d. (-2x^3 + x^2), when (x = -2)</td>
</tr>
</tbody>
</table>
Appendix F: Instructions for administering the No-To-Yes Diagnostic

Before you hand out the test, read the instructions at the top aloud. Tell students the test should take about 20 minutes and that most of the problems can be answered quickly, so they should not take a lot of time on any one problem.

Write on the board as a reminder:

No (  )
No blanks
No negative exponents

At 20 minutes, announce that they have 10 minutes remaining.
At 30 minutes, announce that they may have 5 final minutes to answer the rest quickly.
At 35 minutes, collect them whether or not they have finished.

As you collect the completed diagnostics, ensure they have written their name, block, and teacher at the top. Quickly scan for blanks and parentheses and ask students to fix these. A guess is better than a blank for diagnosing misconceptions!

Place them in my mail cubby and I will grade them ☺.

Thanks so much,

Sue
### Appendix G: Sample of the NTY Test Grading Rubric

<table>
<thead>
<tr>
<th>MC2: Rules of Exponents</th>
<th>Simplify each expression, if possible, leaving no parentheses:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct Answer</td>
<td>6x^6</td>
</tr>
<tr>
<td>Problem</td>
<td>2a. x^2 (x^4) + x^2</td>
</tr>
<tr>
<td></td>
<td>A. Add exponents when multiplying terms with the same base (not multiply)</td>
</tr>
<tr>
<td></td>
<td>B. Combine only like terms (same variable and exponent) when adding and subtracting</td>
</tr>
<tr>
<td></td>
<td>2b. (2x^2)^3</td>
</tr>
<tr>
<td></td>
<td>C. Distribute exponents over multiplication</td>
</tr>
<tr>
<td></td>
<td>D. Multiply exponents when raising a power to a power (not add)</td>
</tr>
<tr>
<td></td>
<td>E. Do not confuse exponentiation with multiplication: a^b != ab</td>
</tr>
<tr>
<td></td>
<td>2c. 4x^6 + 5x^4</td>
</tr>
<tr>
<td></td>
<td>F. Subtract exponents when dividing terms with the same base (not divide)</td>
</tr>
<tr>
<td></td>
<td>B. Combine only like terms (same variable and exponent) when adding and subtracting</td>
</tr>
<tr>
<td></td>
<td>2d. ( \left( \frac{x^2}{(3y^2)^2} \right)^2 )</td>
</tr>
<tr>
<td></td>
<td>Combination of rules A-F (except B)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Misconception Error</th>
<th>Other Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>x^8 + x^2 (Error A)</td>
<td>2^2 x^6 (Not simplified)</td>
</tr>
<tr>
<td>x^9 (Errors A &amp; B)</td>
<td>2^2 x^6 (Not simplified)</td>
</tr>
<tr>
<td>x^11 (Error B)</td>
<td>2^2 x^6 (Not simplified)</td>
</tr>
<tr>
<td>x^6 + x^3</td>
<td>2 in numerator (Error C)</td>
</tr>
<tr>
<td></td>
<td>x^5 in numerator (Error D)</td>
</tr>
<tr>
<td></td>
<td>6 in numerator (Error E)</td>
</tr>
<tr>
<td></td>
<td>Combinations</td>
</tr>
<tr>
<td>6x^6</td>
<td>4x^2 + 5x^4 (Error F)</td>
</tr>
<tr>
<td></td>
<td>4x^4 + 5x^4 (Error B)</td>
</tr>
<tr>
<td></td>
<td>9x^7 (Errors F and B)</td>
</tr>
<tr>
<td>9x^2</td>
<td>x^8 (Error A)</td>
</tr>
<tr>
<td></td>
<td>3 in numerator (Error C)</td>
</tr>
<tr>
<td></td>
<td>y^3 (Error D or F)</td>
</tr>
<tr>
<td></td>
<td>6 in numerator (Error E)</td>
</tr>
<tr>
<td></td>
<td>Combinations</td>
</tr>
<tr>
<td></td>
<td>Variables not simplified</td>
</tr>
</tbody>
</table>

*Note: The table above represents a sample grading rubric for the MC2: Rules of Exponents question from the NTY (National Training Examination) test.*
Appendix H: Screen Shots from Video

1. If \( x = -3 \), then evaluate: \(-x^2\). Did you answer 9? ???

2. \( \text{Nooooooo!!} \)
   
   If \( x = -3 \), does \(-x^2 = 9\)?
   
   Plug in -3 like this:
   
   \[-(x)^2 \text{ becomes } -(3)^2 = -9\]
   
   Which is NOT 9!

3. If you answered "9," pay close attention to the ideas that follow!!

4. Many students think:
   
   It makes a negative inside a variable squared.

5. Many students think:
   
   Nooo! You need to Reveal the Hidden Parentheses!

6. Try Again! If \( x = -2 \), evaluate:
   
   \(-x^2\).

7. So what's the secret?
   
   EVERY time you plug into an expression,
   
   REVEAL THE HIDDEN PARENTHESES

8. That's all there is to it!
   
   1. REVEAL THE HIDDEN PARENTHESES
   
   Yes! Negative Inputs CONQUERED
   
   Then EVALUATE
   
   (Don't forget about following the proper order of operations)
Appendix I: Process Mnemonics in Online Videos

1. Negative Inputs

**REVEAL THE HIDDEN PARENTHESES**

2. Exponent Rules

**E-X-P-A-N-D**

3. Distribute the Negative

**DTN**

**REVEAL THE HIDDEN PARENTHESES**

4. Distribution Rules

**ENDOSA!**

**E-X-P-A-N-D**

(ENOSA = Exponents Never Distribute Across Subtraction or Addition)

5. Random Canceling

6. Fraction Rules

7. Negative Exponents

**CROSS THE LINE**

- Change Exponent’s Sign +

**Married? OR Neighbors?**

**E-X-P-A-N-D**
Appendix J: Sample Online Quiz Question and Feedback

Sample Question:

Quiz Instructions

Use what you learned in the video to answer the 4 questions in this quiz. If you miss any questions, scroll down and study the explanations given, and then try again! You need to score 100% on a quiz to unlock this module and remove the orange sticker on the wall chart.

You can try as many times as you like and you can come back later to practice again.

Good luck!

Question 1

Simplify, if possible, leaving no parentheses: \(x \cdot x^3 + x^4\)

○ \(x^7\)

○ \(x^8\)

○ \(x^4 + x^4 = 2x^4\)

○ \(x^3 + x^4\)

Feedback:

You Answered

Nooo! There are two errors here. First, remember that when you multiply terms with the same base, you add the exponents. So, \(x^1 \cdot x^3 = x^4\).

If you’re not sure, e-x-p-a-n-d the problem (or a simpler one) like in this example:

\(x^3 \cdot x^4 = x \cdot x \cdot x \cdot x \cdot x \cdot x = x^7\)

Also, you cannot add the exponents when adding two unlike terms.

If you’re not sure, plug in \(x=2\) to test your answer.
Appendix K: Examples of In-Class Practice Slips

This page was cut into four separate versions of the practice slip.

**CODE NAME:**
5. *(Version A)* Simplify if possible, leaving no parentheses:

5a. \( \frac{d}{10+d} \)

5b. \( \frac{6y+6}{6y} \)

5c. \( \frac{12+6x}{2+6x} \)

5d. \( \frac{2xy+5y}{30y+12x^2y} \)

**CODE NAME:**

5. *(Version B)* Simplify if possible, leaving no parentheses:

5a. \( \frac{16a+2b}{2a+4} \)

5b. \( \frac{a}{a+b} \)

5c. \( \frac{7+x}{14x} \)

5d. \( \frac{3x+6xy}{3x^2+12xy} \)

**CODE NAME:**

5. *(Version C)* Simplify if possible, leaving no parentheses:

5a. \( \frac{a+2}{-2a} \)

5b. \( \frac{3x^2-9}{12x^2+9} \)

5c. \( \frac{2b}{2b+a} \)

5d. \( \frac{6y+12xy}{24y^2+30xy} \)

**CODE NAME:**

5. *(Version D)* Simplify if possible, leaving no parentheses:

5a. \( \frac{-8}{8x+4} \)

5b. \( \frac{6x+4}{6x} \)

5c. \( \frac{3y+6x}{9y^2+14xy} \)

5d. \( \frac{2y^2+12x^2y}{24y^2+30xy^2} \)
Appendix L: Sample Wall Chart

**The Wall Chart**

- "Yes!" earned on pre-test
- Orange = locked; must do online component
- White = unlocked; ready to do slips in class
- 2 slips completely correct
- 3 slips correct = green "Yes!"

Note: Since Apollonius earned the first 7 MCs (turning them green), she has unlocked the 8th MC.
Appendix M: Assumption Checking for Hypothesis 1

For a repeated measures ANOVA model to be valid, the assumptions that must be checked include a lack of outliers, normality for every combination of time and group, and homogeneity of variances (Laerd Statistics, 2015d). The boxplot in Figure M1 shows there were no outliers in the data greater than 3 standard deviations away from the mean, though there were observations in the Full Group greater than 1.5 standard deviations from the mean. There were also no studentized residuals greater than 3.0.

Figure M1: Boxplots of pre- and post-test misconception error scores by group.

As assessed by Shapiro-Wilk's test of normality ($p > .05$), the data was normally distributed for every combination of time and group except for MESpre-Partial, which was just under the significance level ($p = .039$). Non-normality has “negligible consequences on type-I and type-II errors unless the populations are highly skewed, the $n$’s are very small, or one-tailed tests are employed” (Glass & Hopkins, 1996). Since none of the caveats apply, the test is valid
for this mild deviation from normality. There was not homogeneity of variances \( (p > .05) \) for MESpost as assessed by Levene's test of homogeneity of variances \( (p = .001) \). However, according to McGuinness (2002), all that is required for the analysis to be reliable “is approximate normality of the observations and equality of variances. Only when there is a single, large variance, or marked non-normality are there likely to be substantiated problems” (p. 687). Since there was not markedly non-normal data nor one single, large variance, the analysis may proceed. There was homogeneity of covariances \( (p = .077) \) as assessed by Box’s M test. See Table M1 for SPSS output for checking these assumptions.

Table M1

| Tests of Normality, Equality of Error Variances, and Equality of Covariances |
|-----------------------------|-----------------------------|
|                            | Kolmogorov-Smirnov\(^a\)    | Shapiro-Wilk                   |
|                            | ThreeGroups                | Statistic | df | Sig.  | Statistic | df | Sig.  |
| **MESpre**                 |                            |           |    |       |           |    |       |
| Control                    | .128                       | 26        | .200\(^*\) | .954 | 26 | .287 |
| Full                       | .230                       | 20        | .007 | .917 | 20 | .088 |
| Partial                    | .189                       | 21        | .050 | .902 | 21 | .039 |
| **MESpost**                |                            |           |    |       |           |    |       |
| Control                    | .139                       | 26        | .200\(^*\) | .972 | 26 | .681 |
| Full                       | .190                       | 20        | .056 | .955 | 20 | .445 |
| Partial                    | .130                       | 21        | .200\(^*\) | .957 | 21 | .461 |

\(^a\). This is a lower bound of the true significance.

a. Lilliefors Significance Correction

### Levene's Test of Equality of Error Variances\(^a\)

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>df1</th>
<th>df2</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MESpre</strong></td>
<td>1.270</td>
<td>2</td>
<td>64</td>
<td>.288</td>
</tr>
<tr>
<td><strong>MESpost</strong></td>
<td>7.799</td>
<td>2</td>
<td>64</td>
<td>.001</td>
</tr>
</tbody>
</table>

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + ThreeGroups
   Within Subjects Design: Time

### Box's Test of Equality of Covariance Matrices\(^a\)

<table>
<thead>
<tr>
<th></th>
<th>Box’s M</th>
<th>F</th>
<th>df1</th>
<th>df2</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Box’s M</strong></td>
<td>11.934</td>
<td>1.898</td>
<td>6</td>
<td>76054.254</td>
<td>.077</td>
</tr>
</tbody>
</table>

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + ThreeGroups
   Within Subjects Design: Time
Once the assumptions were evaluated, the ANOVA was interpreted. Tables M2 and M3 contain the SPSS output generated from testing the simple main effects of Group and Time.

Table M2

**ANOVA Output for the Simple Main Effects of Group by Time**

*Tests of Between-Subjects Effects*

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>31.630²</td>
<td>2</td>
<td>15.815</td>
<td>.660</td>
<td>.520</td>
<td>.020</td>
</tr>
<tr>
<td>Intercept</td>
<td>14350.902</td>
<td>1</td>
<td>14350.902</td>
<td>598.951</td>
<td>.000</td>
<td>.903</td>
</tr>
<tr>
<td>ThreeGroups</td>
<td>31.630</td>
<td>2</td>
<td>15.815</td>
<td>.660</td>
<td>.520</td>
<td>.020</td>
</tr>
<tr>
<td>Error</td>
<td>1533.444</td>
<td>64</td>
<td>23.960</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>16223.000</td>
<td>67</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>1565.075</td>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. R Squared = .020 (Adjusted R Squared = -.010)

*Tests of Between-Subjects Effects*

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>521.579³</td>
<td>2</td>
<td>260.790</td>
<td>12.116</td>
<td>.000</td>
<td>.275</td>
</tr>
<tr>
<td>Intercept</td>
<td>4277.869</td>
<td>1</td>
<td>4277.869</td>
<td>198.750</td>
<td>.000</td>
<td>.756</td>
</tr>
<tr>
<td>ThreeGroups</td>
<td>521.579</td>
<td>2</td>
<td>260.790</td>
<td>12.116</td>
<td>.000</td>
<td>.275</td>
</tr>
<tr>
<td>Error</td>
<td>1377.525</td>
<td>64</td>
<td>21.524</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>6563.000</td>
<td>67</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>1899.104</td>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. R Squared = .275 (Adjusted R Squared = .252)

**Multiple Comparisons**

<table>
<thead>
<tr>
<th>(I) ThreeGroups</th>
<th>(J) ThreeGroups</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td>Tukey HSD</td>
<td>Control</td>
<td>Full</td>
<td>6.78 $$^*$</td>
<td>1.380</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Partial</td>
<td>3.24</td>
<td>1.361</td>
<td>.052</td>
</tr>
<tr>
<td></td>
<td>Full</td>
<td>Control</td>
<td>-6.78 $$^*$</td>
<td>1.380</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Partial</td>
<td>-3.54 $$^*$</td>
<td>1.450</td>
<td>.045</td>
</tr>
<tr>
<td></td>
<td>Partial</td>
<td>Control</td>
<td>-3.24</td>
<td>1.361</td>
<td>.052</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Full</td>
<td>3.54 $$^*$</td>
<td>1.450</td>
<td>.045</td>
</tr>
</tbody>
</table>
Table M3

**ANOVA Output for the Simple Main Effects of Time by Group**

*Tests of Within-Subjects Contrasts*  
*Measure: MES*

<table>
<thead>
<tr>
<th>Source</th>
<th>Type</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
<th>Noncent. Parameter</th>
<th>Observed Power b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Linear</td>
<td>204.019</td>
<td>1</td>
<td>204.019</td>
<td>35.059</td>
<td>.000</td>
<td>.584</td>
<td>35.059</td>
<td>1.000</td>
</tr>
<tr>
<td>Error(Time)</td>
<td>Linear</td>
<td>145.481</td>
<td>25</td>
<td>5.819</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>a. ThreeGroups = Control</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>b. Computed using alpha = .05</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Tests of Within-Subjects Contrasts*  
*Measure: MES*

<table>
<thead>
<tr>
<th>Source</th>
<th>Type</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
<th>Noncent. Parameter</th>
<th>Observed Power b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Linear</td>
<td>507.524</td>
<td>1</td>
<td>507.524</td>
<td>87.901</td>
<td>.000</td>
<td>.815</td>
<td>87.901</td>
<td>1.000</td>
</tr>
<tr>
<td>Error(Time)</td>
<td>Linear</td>
<td>115.476</td>
<td>20</td>
<td>5.774</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>a. ThreeGroups = Partial</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>b. Computed using alpha = .05</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Tests of Within-Subjects Contrasts*  
*Measure: MES*

<table>
<thead>
<tr>
<th>Source</th>
<th>Type</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
<th>Noncent. Parameter</th>
<th>Observed Power b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Linear</td>
<td>837.225</td>
<td>1</td>
<td>837.225</td>
<td>124.009</td>
<td>.000</td>
<td>.867</td>
<td>124.009</td>
<td>1.000</td>
</tr>
<tr>
<td>Error(Time)</td>
<td>Linear</td>
<td>128.275</td>
<td>19</td>
<td>6.751</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>a. ThreeGroups = Full</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>b. Computed using alpha = .05</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix N: Assumption Checking for Hypothesis 3B

For Hypothesis 3B, the continuous dependent variable was MESchange and the continuous independent variables were TotalVideo, Total4Quiz, and TotalYesSlips. All assumptions for multiple regression were met as detailed below.

A Durbin-Watson statistic of 1.55 showed independence of residuals (see Table N2). There was no multicollinearity since there were no individual correlations greater than 0.7 (maximum of 0.662) (see Table N1) and no VIF greater than 2.002 (see Table N2).

Table N1

| Correlations between Variables in the Multiple Regression for Hypothesis 3B |
|-----------------|--------|--------|--------|--------|
|                 | MChange | TotalYesSlips | TotalVideo | Total4Quiz |
| Pearson Correlation | 1.000  | -.305  | -.415  | -.374  |
| TotalYesSlips     | -.305  | 1.000  | .662   | .557   |
| TotalVideo        | -.415  | .662   | 1.000  | .519   |
| Total4Quiz        | -.374  | .557   | .519   | 1.000  |

<table>
<thead>
<tr>
<th>Sig. (1-tailed)</th>
<th>MChange</th>
<th>TotalYesSlips</th>
<th>TotalVideo</th>
<th>Total4Quiz</th>
</tr>
</thead>
<tbody>
<tr>
<td>MChange</td>
<td>.</td>
<td>.026</td>
<td>.003</td>
<td>.008</td>
</tr>
<tr>
<td>TotalYesSlips</td>
<td>.026</td>
<td>.</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>TotalVideo</td>
<td>.003</td>
<td>.000</td>
<td>.</td>
<td>.000</td>
</tr>
<tr>
<td>Total4Quiz</td>
<td>.008</td>
<td>.000</td>
<td>.000</td>
<td>.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>MChange</th>
<th>TotalYesSlips</th>
<th>TotalVideo</th>
<th>Total4Quiz</th>
</tr>
</thead>
<tbody>
<tr>
<td>MChange</td>
<td>41</td>
<td>41</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>TotalYesSlips</td>
<td>41</td>
<td>41</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>TotalVideo</td>
<td>41</td>
<td>41</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>Total4Quiz</td>
<td>41</td>
<td>41</td>
<td>41</td>
<td>41</td>
</tr>
</tbody>
</table>

As shown in Figure N1, there was homoscedasticity of residuals as assessed by visual inspection of a scatterplot of the studentized residuals against the unstandardized predicted values.
Further, there were no studentized residuals greater than 3 standard deviations from the mean, no leverage points higher than 0.189 (under the than safe value of .2), and no highly influential points over .69 (under the safe value of 1.0). As shown in Figure N2, the errors (residuals) were approximately normally distributed as assessed by visual inspection of the Normal P-P Plot of the regression standardized residuals.

**Figure N2:** Normal P-P plot of the regression standardized residuals for Hypothesis 3B.
Linearity was shown in two ways. First, there was a linear relationship between the dependent and independent variables collectively, assessed by visual inspection of a scatterplot of the studentized residuals against the unstandardized predicted values (see Figure N1). Second, as shown in Figure N3, there was a linear relationship between each of the predictor variables and the dependent variable assessed by visual inspection of a series of partial regression plots of each independent variable against the dependent variable.

*Figure N3:* Partial regression plots of each independent variable vs. the dependent variable MESchange

Once all the assumptions were assessed, the regression was interpreted as described in Chapter 4.
Summary of Multiple Regression Results for Hypothesis 3B

### Model Summary

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of Estimate</th>
<th>Durbin-Watson</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.455</td>
<td>.207</td>
<td>.143</td>
<td>3.392</td>
<td>1.552</td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), Total4Quiz, TotalVideo, TotalYesSlips
b. Dependent Variable: MChange

### ANOVA

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Regression</td>
<td>3</td>
<td>37.050</td>
<td>3.219</td>
<td>.034</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>37</td>
<td>11.509</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>40</td>
<td>536.976</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Dependent Variable: MChange
b. Predictors: (Constant), Total4Quiz, TotalVideo, TotalYesSlips

### Coefficients

<table>
<thead>
<tr>
<th>Model</th>
<th>B</th>
<th>Std. Error</th>
<th>Beta</th>
<th>t</th>
<th>Sig.</th>
<th>95.0% Confidence Interval for B</th>
<th>Correlations</th>
<th>Collinearity Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Upper Bound</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Zero-order</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Partial</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Part</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Tolerance</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VIF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>-5.015</td>
<td>1.109</td>
<td>-4.524</td>
<td>.000</td>
<td>-7.262</td>
<td>-2.770</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TotalYesSlips</td>
<td>.016</td>
<td>.102</td>
<td>.032</td>
<td>.153</td>
<td>.879</td>
<td>-1.191</td>
<td>.222</td>
</tr>
<tr>
<td></td>
<td>TotalVideo</td>
<td>-.098</td>
<td>.062</td>
<td>-.319</td>
<td>-1.584</td>
<td>.122</td>
<td>-2.224</td>
<td>.207</td>
</tr>
<tr>
<td></td>
<td>Total4Quiz</td>
<td>-.326</td>
<td>.264</td>
<td>-.226</td>
<td>-1.243</td>
<td>.222</td>
<td>-0.864</td>
<td>-.374</td>
</tr>
</tbody>
</table>

a. Dependent Variable: MChange