GRADE DELAYED: COLLEGE ALGEBRA STUDENTS’
REFLECTION, ANALYSIS, AND RECIPIENCE
OF NARRATIVE FEEDBACK ON EXAMS

A Thesis
Presented
To the Faculty of
California State University, Chico

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
in
Mathematics Education

by
© Allison J. McConnell
Fall 2020
GRADE DELAYED: COLLEGE ALGEBRA STUDENTS’
REFLECTION, ANALYSIS, AND RECIPIENCE
OF NARRATIVE FEEDBACK ON EXAMS

A Thesis

By

Allison J. McConnell

Fall 2020

APPROVED BY THE DEAN OF GRADUATE STUDIES:

______________________________

Sharon Barrios, Ph.D.

APPROVED BY THE GRADUATE ADVISORY COMMITTEE:

______________________________

Pamela A Morrell
Graduate Coordinator

______________________________

Brian Lindaman, Ph.D., Co-Chair

______________________________

Christine “Christin” Herrera, Ph.D., Co-Chair
PUBLICATION RIGHTS

No portion of this thesis may be reprinted or reproduced in any manner unacceptable to the usual copyright restrictions without the written permission of the author.
ACKNOWLEDGEMENTS

Throughout the writing of this thesis, I received enormous support and assistance. I would like to thank my thesis committee co-chairs Drs. Christine “Christin” Herrera and Brian Lindaman. Throughout the writing process, both provided invaluable and consistent encouragement, resources, and direction, all of which has strengthened this work far beyond its initial potential. They always kept their figurative “door” open whenever I needed guidance. Thanks to Dr. Kathy Gray as well – despite not being on my committee, she provided crucial guidance and feedback on my statistical analysis.

I would like to thank my professors in my Mathematics Education master’s and undergraduate programs for a rich and inspiring learning experience and for constant encouragement. In particular, I would like to thank Dr. Yuichi Handa as my former professor turned friend and life coach for the advice about “showing up” to the thesis as a daily practice – I do not know what the status of my thesis would be at this time of completion had I not been privy to Dr. Handa’s wisdom and insight. Thank you to Dr. Jorgen Berglund and Dr. Handa once more for sparking my passion for mathematics education in my undergraduate program and for modeling excellent teaching practice. I would also like to thank Dr. Krista Strand for her emotional support and authenticity, as well as the expertise she shared in her course and beyond. She is a stellar model of compassion in education, and I feel so fortunate to know her.

I would also like to thank my family, friends, and colleagues for their support and consistent check-ins, which served as reminders to keep going and see the process through to the end. Special thanks to my spouse, Jay Gatton, for endless patience, understanding, and partnership throughout this process.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Publication Rights</th>
<th>iii</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgements</td>
<td>iv</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>v</td>
</tr>
<tr>
<td>List of Tables</td>
<td>vii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>viii</td>
</tr>
<tr>
<td>Abstract</td>
<td>ix</td>
</tr>
</tbody>
</table>

## Chapter I Introduction to the Study

1. Background ....................................................... 1
2. Statement of the Problem .................................... 8
3. Purpose of the Study .......................................... 12
4. Theoretical Bases and Organization .......................... 13
5. Limitations and Delimitations of the Study ................ 14
6. Definition of Terms ............................................ 15

## Chapter II Survey and Review of Literature

1. Defining Mathematical Proficiency .............................. 17
2. Constructivist Learning Theory ................................ 19
3. Feedback as Grades ............................................. 22
4. Feedback over Grades .......................................... 30
5. Feedback in the Mathematics Classroom – Forms and Effectiveness ........................................ 38
6. Conclusion ....................................................... 45

## Chapter III Methodology

1. Introduction .................................................... 47
2. Research Questions ............................................. 50
3. Research Design and Worldview ................................ 50
4. Research Approach .............................................. 52
5. Population ....................................................... 55
6. Treatment ......................................................... 56
7. Data Analysis Procedures ..................................... 59

## Chapter IV Results and Discussion

1. Results and Discussion ......................................... 66
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantitative Results and Discussion</td>
<td>68</td>
</tr>
<tr>
<td>Qualitative Results and Discussion</td>
<td>84</td>
</tr>
<tr>
<td>Chapter V Summary, Conclusions, and Recommendations</td>
<td>97</td>
</tr>
<tr>
<td>Research Question 1: Comparing Scores</td>
<td>98</td>
</tr>
<tr>
<td>Research Question 2: Comparing Rationale</td>
<td>104</td>
</tr>
<tr>
<td>Research Question 3 – Student Perception and Experience</td>
<td>109</td>
</tr>
<tr>
<td>Future Directions</td>
<td>111</td>
</tr>
<tr>
<td>Summary</td>
<td>114</td>
</tr>
<tr>
<td>References</td>
<td>118</td>
</tr>
<tr>
<td>Appendix</td>
<td>128</td>
</tr>
<tr>
<td>Exam 1</td>
<td>129</td>
</tr>
<tr>
<td>Exam 1 Reflection and Analysis Process (RAP) Assignment</td>
<td>133</td>
</tr>
<tr>
<td>Exam 1 Rubric and Scoring Sheet for Students*</td>
<td>134</td>
</tr>
<tr>
<td>Exam 2</td>
<td>135</td>
</tr>
<tr>
<td>Exam 2 Reflection and Analysis Process (RAP) Assignment</td>
<td>137</td>
</tr>
<tr>
<td>Exam 2 RAP Assignment (Screenshots of Google Form)</td>
<td>138</td>
</tr>
<tr>
<td>Exam 2 Rubric and Scoring Sheet for Students</td>
<td>141</td>
</tr>
<tr>
<td>Error Analysis Reference Guide for Exam 1 and Exam 2</td>
<td>142</td>
</tr>
<tr>
<td>Informed Consent Letter</td>
<td>143</td>
</tr>
<tr>
<td>IRB Approval Letter</td>
<td>145</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Epistemological Beliefs and Corresponding Levels of Belief (Hofer &amp; Pintrich, 1997 as cited in Steiner, 2007)</td>
</tr>
<tr>
<td>2</td>
<td>4-Point Rubric for Scoring a Single Problem or Task (adapted from CPM Educational Program, 2013)</td>
</tr>
<tr>
<td>3</td>
<td>List of Quantitative and Qualitative Data</td>
</tr>
<tr>
<td>4</td>
<td>Categories and Subcategories of Survey Responses for Question 1</td>
</tr>
<tr>
<td>5</td>
<td>Exam 1 Concept 1 Summary Statistics ($n = 34$)</td>
</tr>
<tr>
<td>6</td>
<td>Exam 1 Concept 2 Summary Statistics ($n = 34$)</td>
</tr>
<tr>
<td>7</td>
<td>Exam 1 Concept 3 Summary Statistics ($n = 34$)</td>
</tr>
<tr>
<td>8</td>
<td>Exam 2 Concept 1 Summary Statistics ($n = 27$)</td>
</tr>
<tr>
<td>9</td>
<td>Exam 2 Concept 2 Summary Statistics ($n = 27$)</td>
</tr>
<tr>
<td>10</td>
<td>Exam 2 Concept 3 Summary Statistics ($n = 27$)</td>
</tr>
<tr>
<td>11</td>
<td>Percent of Each Type of Student Response from Survey Question 1</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

FIGURE                                PAGE
1  Working Model of How Epistemological Theories Influence Classroom Learning (Hofer, 2001) ........................................................... 26
2  Chronological Timeline for Each Piece of Data Collected .............................. 54
3  Categories of Math Placement for Incoming First-Year CSU Students ............... 56
4  A Comparison of Scatterplots without “Jitter” (left) and with “Jitter” (right) ..... 61
5  Exam 1 Concept 1 Score Comparison Scatterplot and Line Graph ..................... 69
6  Exam 1 Concept 2 Score Comparison Scatterplot and Line Graph ..................... 71
7  Exam 1 Concept 3 Score Comparison Scatterplot and Line Graph ..................... 73
8  Exam 2 Concept 1 Score Comparison Scatterplot and Line Graph ..................... 77
9  Exam 2 Concept 2 Score Comparison Scatterplot and Line Graph ..................... 80
10 Exam 2 Concept 3 Score Comparison Scatterplot and Line Graph .................... 82
ABSTRACT

GRADE DELAYED: COLLEGE ALGEBRA STUDENTS’ REFLECTION, ANALYSIS, AND RECIPIENCE OF NARRATIVE FEEDBACK ON EXAMS

by

© Allison J. McConnell 2020

Master of Science in Mathematics Education

California State University, Chico

Fall 2020

This mixed-methods study explored thirty-four students’ reflection, error analysis, and recipience of feedback in a College Algebra course at a large public university in Northern California. The Reflection and Analysis Process (RAP) assignment was designed to emphasize reflection and de-emphasize grades. For the RAP, students self-scored each concept, provided rationale for these self-scores, and used instructor feedback to analyze errors and provide corrections.

Results indicated that student scores were not correlated with instructor scores, with a consistent trend for students to self-score in the middle range of the rubric. In the rationales, students who earned lower instructor scores were prone to minimizing the severity of their errors, and self-scored based on perceived understanding at the time of the RAP rather than when they took the exam. Students earning higher instructor scores indicated they could have better explained the mathematics or shown more work. Students with the largest discrepancies between their self-score and instructor score were more likely to refer to feelings and behaviors, while
students whose scores aligned with the instructor relied more on their exam work and used more mathematical vocabulary in their rationales.

It is recommended that practitioners implementing the RAP leverage technology for more efficient and consistent narrative feedback, aim for open-ended assessment items, implement a system of single-point rubrics, include practice on self-scoring prior to a RAP-type assignment, and frequently highlight exemplars of errors in class discussions to normalize struggle and model meaningful reflection.

*Keywords: grading, gradeless, feedback, reflection, error analysis, mistakes, developmental mathematics, assessment, exam*
CHAPTER I

INTRODUCTION TO THE STUDY

Background

A key pillar of the learning process for students is the feedback they receive from their teachers, and methods of feedback vary significantly among settings. Feedback in its many forms can be one of the most powerful influences on student learning. Over the last century, we have seen a surge of research about instructional theory and practice (Ertmer et al., 2013) which has showcased a shift of educators toward constructivist teaching and learning philosophies and away from more behaviorist philosophies (Karagiorgi & Symeou, 2005). For mathematics instruction specifically, recent research emphasizes the need for a robust and multi-faceted construction of mathematical proficiency in students and the need for learning experiences that facilitate it. This study will focus on one method to fill the need for such learning experiences through course assessments, written feedback, and student reflection.

Educational Philosophy and Learning Theory

The underlying educational philosophy of this study is constructivism. Mathematicians actively construct knowledge by “doing,” or actively engaging with mathematics hands-on through experimentation and self- and community-evaluation (Confrey, 1990). Constructivists reject rigidity and narrow “facts” passed down from external authority, which in an academic environment is the teacher or even other students. Constructivist teachers reject the role as an arbiter of fact and truth (Confrey, 1990); instead, they value epistemic heterogeneity and epistemic authority of the students. Epistemic heterogeneity is the notion that knowledge and learning can be exhibited in many valid forms while epistemic authority has to
do with who has the responsibility of expertise and knowing. The concept of student agency describes when the student feels epistemic authority over developing knowledge and expertise themselves.

Constructivist educators strive to validate student thought and act as an assistant and facilitator to their students to overcome weaknesses and limitations in their constructions. They avoid simply telling learners discrete facts directly in favor of students developing their constructions through experience. Confrey (1990) emphasizes that the instructor’s role is to promote student autonomy and agency (i.e., help students seize their epistemic authority over the content) and develop students’ reflective processes with continuing opportunities to “re-view” previous work, always reflecting on growth and progress throughout a course of study.

With the advent of Common Core (2010) and the surge of research in the 1990s that advocate for constructivist education and challenge behaviorist practices, it is much more prevalent nowadays for a teacher to follow principles of constructivism in their lessons. In fact, the reader may notice the parallels between the principles of strong knowledge constructions (Confrey, 1990) and the eight Standards of Mathematical Practice in the CCSS (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) that serve as one definition of mathematical proficiency.

**Defining Mathematical Proficiency**

There are several sources that define mathematical proficiency (e.g., Confrey, 1990, Kilpatrick et al., 2009; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). The overall themes include conceptual reasoning, procedural fluency, modeling mathematics through multiple representations, problem-solving competence and strategy, and the need for a productive disposition about mathematics. Furthermore, Confrey
(1990) posits that powerful constructions of mathematical proficiency should function across a variety of topics, converge among multiple forms and contexts of representation, be adequately understood such that they can be reflected upon and clearly described, bejustifiable and defendable, and have the power to be integrated with past and future constructions. Like physical tools, mathematics is a tool that if overly specialized or inaccessible, is abandoned (Confrey, 1990). The most long-lasting tools can be used consistently over time for a purpose that serves the user; the cognitive and problem-solving aspects of mathematics are the tools which are transferrable and lasting (Confrey, 1990).

If educators wish to develop such a rich mathematical proficiency and constructions of knowledge for all students, then educators’ choices of their student learning experiences must be highly strategic. The learning experiences should develop each aspect of mathematical proficiency interdependently with one another in a way that supports all students’ learning.

**Personal and Epistemological Beliefs**

One of the key aspects included in all definitions of mathematical proficiency is a productive disposition to learn mathematics. Unfortunately, research indicates that students generally have a very unproductive disposition about mathematics. Steiner (2007) found that students were generally not very confident about their ability to solve mathematics problems, and students held, what Steiner refers to as “nonavailing beliefs” about their mathematical abilities. Furthermore, the study concluded that a positive self-concept positively influenced how students performed on their exams. Steiner’s (2007) conclusions are echoed by many other researchers (e.g., Kohn, 2004; Barnes, 2015; Boaler, 2016).

It follows that for students who lack confidence in mathematics, if or when their low confidence translates to lower performance on exams as indicated by a grade or score, such low
performance could result in self-fulfilling affirmations contributing to their nonavailing beliefs about mathematics. Steiner (2007) measured these students’ beliefs about knowledge and learning on a scale of low-end to high-end epistemological beliefs with the low-end representing beliefs that are least productive for deep learning and the high-end representing beliefs that are most productive.

Steiner (2007) concluded that there is a significant need for more experiences in college mathematics courses that challenge students’ nonavailing personal and epistemological beliefs. These learning experiences need to promote the development of a more positive self-conception surrounding mathematics, contribute to a more productive disposition about learning and doing mathematics, and progress students’ epistemological beliefs to the higher end of the scale. Other sources, such as Hofer’s (2001) work about personal epistemologies’ and their implications for learning and teaching, support Steiner’s conclusions.

**Current Issues with Assessment**

While the push for more constructivist mathematics instruction has transformed instruction to become more robust and inclusive, our assessment and feedback practices have been slower to change. Feedback is often provided, if not *purely* through grades, points, and scores, then at least with an often-inadvertent *emphasis* on grades, points, and scores. Furthermore, assessments tend to be weighted more heavily grade-wise than other types of assignments resulting in immense pressure for students to perform well on the assessment without necessarily providing accurate evidence of their learning.

Students tend to engage in an intense task of studying before an exam, often by “cramming” and memorizing discrete pieces of information in one or a few sittings ranging from a few hours to a few days before the test takes place. The belief that knowledge is defined as the
accumulation of discrete pieces of information is one that is considered less productive for deep mathematics learning when contrasted with the more productive belief that knowledge consists of highly interrelated concepts (Hofer, 2001; Steiner, 2007). This typical method of study as the memorizing of facts is this knowledge-as-accumulation-of-fact belief played out in practice. Ideally, for the most robust mathematical proficiency, we want students to think of knowledge as complex, evolving, and justified through evidence, logic, and reason.

Given the dramatic differences between learning experiences and testing environments, students often struggle to integrate the two. A test or quiz tends to take place in a single class period and may only be revisited once more briefly. This is notable as a missing component of one of the aspects of Confrey’s characteristics of constructivist educators (i.e., that students should be offered opportunities to reflect on their growth and progress and re-view their past work often). Since students with epistemological beliefs that are less productive for deep learning perceive teachers or other experts as the source of knowledge, that authority can translate to how the student should feel about themselves as a learner of mathematics. Low grades tend to correlate with less favorable feelings about oneself and often feelings of helplessness to improve one’s grade and therefore the understanding that is theoretically correlated with it (Selby & Murphy, 1992).

Given how classroom tasks and pedagogical practices influence student learning, assessment and feedback should not be treated as classroom tasks that are disconnected from other learning activities, and their focus should not revolve around a grade or score. The typical assessment process inherently discourages students from reflecting upon their learning because it focuses their narrowed attention almost entirely upon an extrinsic motivator of a grade that is assigned by an outside authority.


**Current Issues with Grades**

Research shows that extrinsic motivators narrow students’ focus and incentivize students to produce work when faced with monotonous tasks. Extrinsic motivators are an inhibitor of the deep and creative constructions of knowledge we aim to develop (Duncker, 1945; Pink, 2009). Yet research shows that while most teachers believe, at least in theory, that grades should reflect understanding, teachers in practice tend to assign grades to incentivize behaviors – the grades reflect the extent to which students complied and completed tasks (Randall & Engelhard, 2010). The threat of earning a low grade on exams can very literally activate afight-or-flight response, which undercuts students’ ability to think in higher order ways. This can exacerbate students’ low-level epistemological beliefs as consisting of discrete and unrelated pieces of information because they are driven to memorize so they can produce the information on the exam. This clearly contradicts the goals of robust mathematical proficiency, specifically the critical thinking and creation aspects since students are so focused on complying with what they think the teacher wants to see.

Grades, especially when they are deficit oriented by taking away credit when students make mistakes and errors, have been shown to discourage experimentation and creativity during students’ learning experiences (e.g., Pink, 2009; Selby & Murphy, 1992 as cited in Guskey, 2011). And furthermore, "...[N]o research supports the idea that low grades prompt students to try harder. More often, low grades prompt students to withdraw from learning. To protect their self-images, many students regard the low grade as irrelevant or meaningless. Others may blame themselves for the low grade but feel helpless to improve.” (Selby & Murphy, 1992 as cited in Guskey, 2011). Deficit-based grading discourages student agency as students must constantly defer to the teacher as the expert for fear of being “marked down” – the teacher is the one
responsible for their grade, and the grade is the only measure of learning that remains once a course is completed.

When one considers the power that a teacher has to influence the outcome of the grade, one must also consider the equity implications. As previously mentioned, teachers tend to assign grades that incentivize desired behaviors as a component of the grades. For students from populations that are underrepresented as education professionals, this can result in huge inequities because the grading practices are so often based on behavior which is, in turn, based on normative expectations of the dominant culture. Low grades can act as barriers when it comes to accessing opportunities in the future and yet do not necessarily reflect a students’ competence or understanding of the course content. Low grades can lead to institutions tracking students into lower-level classes where they almost always stay permanently. In these lower-level classes, students tend to lack access to high-quality education and high-level content and can be prevented from qualifying for scholarly and extra-curricular pursuits. Historically, low numeric measures have been used as evidence against marginalized communities to imply a group is inferior (Gould, 1981).

**Feedback over Grades**

If all aforementioned conditions apply – that is, (1) students believe that knowledge is acquired from external authority and experts; (2) students believe that grades are a measure of how well they understand the mathematics; and yet (3) teachers are, in practice, assigning grades to incentivize desired and normative behaviors rather than assigning grades that reflect acquisition of knowledge – there exists an egregious disconnect about what students learn and what teachers want students to learn. Feedback is often considered almost inextricably linked with grades, and yet research shows grades and scores detract from students accepting written
feedback (Hattie, 2009, 2012). According to Hattie (2009, 2012) in meta-analyses of over 800 studies of learning, feedback is one of the most powerful influences. As such, it is very important that educators consider the effectiveness of their feedback practices.

Barnes’ (2015) recommends one model of giving effective feedback, the SE2R method, in which the instructor summarizes what the student showed, explains which concepts and standards the student has mastered, redirects the student as to what they need to do to make sense of the concepts that have not been mastered, and then encourages the student to resubmit the assignment once corrected. Several researchers in the field advocate for entirely gradeless classrooms (e.g., Boaler, 2016; Sackstein, 2015a), and Barnes is no exception. As part of the SE2R model, Barnes advocates for grades to not be assigned as part of an assignment at all and instead encourages teachers to set the expectation that students must resubmit until they demonstrate mastery. If and when grades must be assigned, depending on the administrative requirements, Barnes (2015) advocates for student-instructor conferences in which the student assigns themself a grade by defending their work using evidence of their progression with work samples or a portfolio of all work completed during the course. Of course, this practice certainly aligns with the assumed definitions of strong mathematical constructions and high-level epistemological beliefs. However, this study proposes a method of feedback that works within the parameters of institutional requirements of the instructor to assign grades but that still promotes student reflection and agency.

Statement of the Problem

If the goal of a mathematics class is constructing deep knowledge and understanding, then carefully selected classroom activities must contribute to the development of that knowledge and high-level epistemological beliefs. If the goal of a final grade is to reflect
students’ overall mathematical proficiency, then the learning experiences that contribute to the final grade should accurately reflect the mathematical proficiency supposedly demonstrated on it. And, if feedback is to be accepted by the students in a way that is conducive to their learning and development of productive epistemological beliefs, then systems must be in place that require students to grapple with the feedback and learn from it before a grade is ever assigned (if a grade has to be assigned at all). In the literature, there exists a call for more learning experiences that contribute to the development of high-level epistemological beliefs (e.g., Steiner, 2007; Barnes, 2015; Boaler, 2016).

Hence, this research focuses on one method of an assessment and feedback system focused on quarterly exams in a College Algebra class with a developmental math corequisite course. The method, which from here on out will be referred to as the Reflection and Analysis Process (RAP), delays any scores or grades and only provides narrative feedback to students when they initially receive an exam back from the instructor. The RAP is one component of a larger assessment system with the potential to embody the shift of education to develop more critical thinking, creativity, and innovation while working within the parameters of exams, which are often required by administration.

The RAP assignment tasks students to re-submit their original tests along with their completed RAP assignment. The assignment requires students to: (1) copy original work on a separate piece of paper and correct any mistakes and errors referenced in the feedback on the copied problem; (2) describe errors and mistakes completely using an Error Analysis Reference Guide to determine the type of error or mistake they made; (3) summarize their original and corrected thought process, including how they can avoid the mistake or error in the future (suggestions were provided in the Error Analysis Reference Guide); and (4) use the provided 4-
point rubric to self-assess their level of understanding as defined by the characteristics for each level. Students also completed surveys regarding how they prepared for each exam and their opinions about the experience. All of these materials can be found in the appendix for reference.

The qualitative feedback for this study roughly followed Barnes’ (2015) SE2R method of feedback which encourages gradeless feedback on assignments. However, while principles of the gradeless movement inspired the creation of the RAP, the RAP assignment and overall assessment structure in this study does assign an instructor grade once students have engaged in the reflection process. The RAP aims to deemphasize the grade of the exam and re-engage students with the learning goals that were assessed by an exam turning it into an assessment of learning (i.e., summative assessment) and for learning (i.e., formative assessment). Students complete the RAP assignment after they have taken their exam and have received this purely narrative feedback from the instructor. The RAP strategically avoids scores and grades until students have already reflected on their work and interacted with the feedback to promote reciepline of the feedback while still working within the parameters around grades that are set by most educational institutions.

By the time students complete the RAP, they have already experienced both collaborative and individual learning experiences of the concepts through several different lessons and thought-provoking lines of questioning. Through these learning experiences, students have received an abundance of informal, formative feedback before an exam takes place. After exams, students receive more formal feedback that their teacher has tailored to their individual needs. Through giving specific, problem-level feedback, the educator gathers information on two levels: (1) the individual students and (2) the class as a whole. Teachers can then use that information to adjust class instruction based on observations and patterns of where students are
in their learning. The exam and RAP is intended to allow the students to identify through questioning on the exam what they were and were not able to do, and therefore integrates the exam into the other learning experiences of the class rather than separating the assessment from all the other learning experiences that have taken place.

The specific, narrative feedback for each problem solution is intended to further facilitate each student’s development of their own learning. Additionally, the feedback provides a guide with scaffolding questions for each of them to develop their understanding, and hopefully, the narrative form feedback lowers some of the psychological barriers that naturally arises when one receives critical feedback alongside points and scores (in particular, a loss of points). In this way, it is hoped that the students will be better equipped to receive and accept the feedback rather than reject and dismiss it as is common when receiving feedback of a more critical nature. Aside from the fact that students must complete the RAP assignment, compliance and behavior is not measured in the grade. The RAP is not an assignment separate from the exam to be entered into a gradebook, nor does it alter any grades earned on the exam as is often the case with test-correction assignments. Students are not simply getting back a portion of the points they lost. Rather, the RAP is an exam routine that students understand is a required component of an exam before a score will be made available to them. The fact that students understand that there is no final exam grade made available to them before completion of the RAP communicates that revisiting the work is a mandatory piece of the exam, and they will be held accountable for it without necessarily gaining or losing points.

This exam process honors epistemic heterogeneity by allowing multiple ways to demonstrate mathematical proficiency and additional opportunities to defend them. The RAP also provides an opportunity for students to seize epistemic authority, or agency, over their
learning by explicitly requiring and including their input on the understanding each student demonstrated on their exam.

**Purpose of the Study**

The purpose of this study is to explore student reflection, experience, and perception of the RAP as a useful component in a larger assessment system. Ideally, it will contribute to the practical implementation of reflection-based assignments that contribute to students’ development of mathematical proficiency, increase students’ epistemic authority over their mathematical knowledge and agency over their own learning, and provide alternatives for the inequitable practice of assigning assimilative compliance-based grades.

This study of the RAP focused on the first two of quarterly exams in a College Algebra mathematics course with a developmental mathematics corequisite support course. On these exams, the students received solely descriptive feedback for each problem and its parts; students did not receive any numeric scores or letter grades until after they completed their exam RAP based on the qualitative feedback they received. Students were given about three days to complete the RAP assignment outside of class time (as long as the third day did not fall after a weekend or holiday to ensure students had an opportunity to make use of university services such as tutoring or instructor office hours). Students were allowed to submit the assignment late, and students who did not submit on time were followed up with until they submitted it. The three-day timeline offered ample opportunity for late submissions to be submitted within a week of the exams being passed back, which was an instructor goal.

Of particular interest for this research was the extent to which the reflection process impacted students’ development of mathematical proficiency. Due to constraints, the research focused mostly on the aspects of mathematical proficiency pertaining to students’ mathematical
dispositions. Also of interest was the extent to which the RAP contributed to Steiner’s (2007) call for more learning experiences for college math students that challenge nonavailing beliefs about math (i.e., beliefs that are not conducive to learning mathematics) and promote a positive self-conception.

Specifically, this study seeks to answer the following questions:

1. How do students’ self-assessment scores and instructor-assigned scores compare from Exam 1 and Exam 2?
2. How do students’ rationales for their self-assessment scores compare from Exam 1 to Exam 2? Specifically, in terms of epistemological beliefs, what were the characteristics of any shifts that occurred from Exam 1 to Exam 2?
3. What is the student experience and perception of the RAP?

**Theoretical Bases and Organization**

The research uses a complex mixed methods research approach to gather and analyze both quantitative and qualitative data to produce more robust information about the RAP as a system of assessment and feedback. In the study, data collected included the student self-score from the 4-point rubric for each concept on each exam, students’ corresponding rationales for each of their scores, the instructor-assigned scores, and the student responses from a survey with three open-ended questions. The analysis included a comparison of students’ self-scores to the instructor-assigned scores for each concept on each of the two exams. Then students’ rationales for their self-score from Exam 1 to Exam 2 were analyzed and compared, particularly with respect to whether and what kind of shift in rationales of epistemic beliefs occurred. Finally, students’ survey responses were analyzed with specific attention to any clues about their epistemic beliefs before, during, and after the RAP process.
The study used principles of Design-Based Research in that the method was developed in practice by a practitioner for practical use and then researched by the practitioner. The study also incorporates aspects of Participatory Design Research (PDR) and educational justice as the method specifically seeks out perspectives of the participants and aims to disrupt assimilative academic requirements from evaluation of mathematical understanding and honors epistemic heterogeneity and epistemic authority of the students. The survey elicited student feedback about the RAP so further adjustments can be made to improve the assignment for future implementation.

**Limitations and Delimitations of the Study**

There were several limitations to this study. As this research used a complex mixed-methods model and is a descriptive study without an experimental design, the reader should exercise caution to assume only a correlation and not a cause-and-effect relationship. The quantitative portion of the study was not designed as an experiment with control group and treatment group(s). Future research is needed to determine a cause-and-effect relationship and whether or not the RAP process influenced development of mathematical proficiency. Also, due to research limitations, only the first two exams were analyzed in this study.

Due to the exploratory nature of the study, there was inconsistency with feedback format from one exam to the next. Exam 1 had shorter feedback written by hand on the test itself on each part of the problem. On Exam 2, in part because of the campus-wide switch to distance learning following spring break, students received feedback in paragraph form over email with their Exam 2 scanned and attached. Each paragraph first directed them to the corresponding problem, then followed the SE2R format recommended by Barnes (2015).
It was beyond the scope of this research to consider other aspects of the mathematics classroom that led up to and followed students' participation in the exam feedback and reflection process. Other aspects included lessons and tasks, non-exam based informal and formal feedback that took place in and outside of the classroom, assessments other than the exams for which students also received feedback with delayed grades or scores to provide an opportunity for students to see feedback without any scores present but for which students were not required to submit a RAP, the disruption caused by the mid-semester transition to distance learning, and students’ prior experience with mathematics or previous math classes.

It is assumed that students were honest in their survey responses regarding their perception about the RAP.

There is the potential for bias in the analysis. The goals of the study may have influenced what was observed, especially for qualitative correlations and interpretations.

**Definition of Terms**

**Developmental mathematics** – “any course taught on the college level (2-year, 4-year, or university) below the level of “college algebra” or “precalculus”: arithmetic, pre-algebra, beginning of intermediate algebra, and (high-school level) geometry” (MAA Online, 2005, [http://www.maa.org/t_and_l/developmental/dm.html](http://www.maa.org/t_and_l/developmental/dm.html) as cited in Steiner, 2007).

**Epistemic authority** – having to do with who has the responsibility of developing knowledge or expertise; exercising the right or power to know as a form of agency in processes of mathematics problem solving and learning (Booker & Goldman, 2016).

**Epistemic heterogeneity** – validity of various forms of knowledge free of assimilative requirements.
Epistemology – a study of or having to do with knowledge, the nature of knowing, and justification of belief.

Feedback – The process that bridges the gap between what a learner currently understands and what a teacher desires a learner to understand (Sadler, 1989); a consequence of performance (Hattie & Timperley; 2007).

Nonavailing epistemological beliefs – “an individual’s beliefs about the nature, justification, sources, and acquisition of knowledge that either do not correlate or negatively correlate with better learning outcomes” (Muis, 2004 as cited in Steiner, 2007).

Personal epistemology – an individual’s beliefs around knowing and the nature of knowing (Steiner, 2007).

Rubric Score – references a 4-point rubric adapted from the high school curriculum series, College Preparatory Mathematics (CPM), used to measure the level of understanding demonstrated by a student on an individual problem or concept


Summarize: Provide a one- or two-sentence summary of what students have accomplished on an activity.

Explain: Share detailed observations of what skills or concepts have been mastered based on the specific activity guidelines.

Redirect: Instruct for students the lessons, presentations, or models that they need to review to achieve an understanding of the concepts and mastery of skills.

Resubmit: Encourage students to revisit activities or projects after redirection, rework them, and submit for more feedback.
Defining Mathematical Proficiency

Mathematics educators use feedback to develop students’ mathematical learning, ability, and understanding. Learning, ability, and understanding terms are often used interchangeably without much regard to the specificity of their meaning. For the sake of giving feedback, one term that encapsulates robust mathematics learning, ability, and understanding is mathematical proficiency. Kilpatrick et al. (2009) define mathematical proficiency with the following five components: conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and productive disposition. The authors use the metaphor of a rope with the five components of mathematical proficiency as strands of the rope; the idea is that the strands must strategically weave together and are interdependent with one another. These components, or strands, are defined as follows:

- *Conceptual understanding* – comprehension of mathematical concepts, operations, and relations;
- *Procedural fluency* – skill in carrying out procedures flexibly, accurately, efficiently, and appropriately;
- *Strategic Competence* – ability to formulate, represent, and solve mathematical problems;
- *Adaptive Reasoning* – capacity for logical thought, reflection, explanation, and justification; and
• **Productive Disposition** – habitual inclination to see mathematics as sensible, useful, and worthwhile, coupled with a belief in diligence and one’s own efficacy.

The eight Standards of Mathematical Practice (SMP) outlined in the Common Core State Standards (CCSS) offer additional information as to how mathematically proficient students should engage with mathematics, and hence educators should seek to develop these processes within their students. Students should (1) make sense of problems and persevere in solving them, (2) reason abstractly and quantitatively, (3) construct viable arguments and critique the reasoning of others, (4) model with mathematics, (5) use appropriate tools strategically, (6) attend to precision, (7) look for and make use of structure, and (8) look for and express regularity in repeated reasoning (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010).

In her National Council of Teachers of Mathematics (NCTM) article about the learning theory, constructivism, and how it pertains to mathematics teaching, Confrey (1990) describes typical characteristics of a powerful construction of mathematical knowledge. To summarize, these characteristics include the following: (1) students must have “internal consistency” – that is, they must actually believe their construction (in contrast, students will make statements without any awareness of contradiction that they “know” something to be true but that they do not “believe” it to be true); (2) multiple forms and contexts must agree with each other and tie into various symbolic representations; (3) the student must be able to reflect upon, describe, defend, and justify their constructions; and (4) all constructions should have continuity with previous construction, have use across a variety of concepts, and have the potential to act as a tool and a guide for future constructions. Constructions should also agree with experts in the
field, but Confrey deemphasizes this aspect in favor of the other measures that pertain to student agency and student justification. The most powerful premise of constructivism is that students must exert authority over their own learning and their right to understand. Similarly, students should reject the notion that knowledge is simply passed down from an expert lest they fall victim to the contradiction of “knowing, but not believing” facts.

Clearly, there exists a lot of overlap and agreement among these definitions and descriptors of mathematical proficiency. Mathematics instruction, including feedback and grading practices, must aim to develop well-rounded mathematical proficiency in students.

**Constructivist Learning Theory**

Teachers and researchers who study the role of feedback as an integral part of the instructional process tend to cite constructivist learning theory. As Confrey (1990) illustrates with her list of traits of powerful mathematical constructions summarized above, constructivism as a learning theory best describes how students should learn mathematics and come to powerful constructions of knowledge (Confrey, 1990). Ertmer, Newby, and Medsker (2013) summarize constructivism as the philosophy that learners actively construct knowledge through their experiences within their environment and social interactions. Standing in stark contradiction to the premise of active constructions is direct instruction.

Direct instruction has been closely studied over the last century (Good & Grouws, 1979; Peterson et al., 1984; Rosenshine, 1976 as cited in Confrey, 1990) and is defined as a structure that largely follows the format of “introductory review, a development portion, a controlled transition to seatwork, and a period of individual seatwork” (Confrey, 1990, p. 107). The abundant research about direct instruction has illuminated teachers’ limited opportunities to assess students’ mathematical proficiency. Direct instruction tends to favor immediate
procedural skills over any other aspect of mathematical proficiency. The typical feedback systems are focused on performance measured by scores and grades on quizzes, tests, and homework. These measures are accepted as evidence of understanding of primarily (if not only) procedural fluency. Typically, a score is given by the teacher without any further probing questions of the students to elucidate deeper reasoning, thought processes, and other traits of powerful constructions or mathematical proficiency. A score or grade is often the primary means through which feedback is provided, often through a deficiency model in which points are taken off for errors and mistakes. The deficiency model of grading provides very limited information to the learner and teacher about the learning that has taken place; it essentially tells the learner only what they did not understand at that point in time. These scores and grades provide very little, if any, guidance as to what the learner should do next. Opportunities for the learner to revisit the content and demonstrate a higher level of proficiency are commonly nonexistent in such classrooms anyway. The fact that the teacher’s model of direct instruction condenses opportunities for feedback only to formal measures rather than consistently leveraging opportunities throughout a lesson for informal, formative feedback exacerbates these assessment problems. Thus, direct instruction and its corresponding systems of assessment often result in lasting misconceptions and gaps in students’ mathematical understanding (Confrey, 1990).

Confrey (1990) summarizes the challenges that constructivism poses about direct instruction and its meager opportunities for feedback: (1) students are only expected to demonstrate their knowledge in the short term through homework and tests to assess the quality of instruction rather than to assess students’ process and problem solving; (2) students under-develop or fail to develop cognitive skills that enable independent problem solving; and (3) the epistemic authority primarily lies with the teacher. To elaborate on the last item, the teacher is
the one who decides whether an adequate amount of learning has been achieved; they may even check frequently with the class and simply conclude, whether correctly or incorrectly, that students are understanding the instruction satisfactorily (Peterson & Clark, 1978). Often, the teacher fails to identify when students have reached their maximum cognitive capacity for the instruction in that they can no longer receive and integrate the information into their former framework of understanding. Thus, the teacher often fails to revise instruction once the students’ capacity for the instruction is exceeded and simply continues as planned (Confrey, 1990). Confrey’s research suggests that as students receive more direct instruction, they tend to demonstrate immediate success with completing unambiguous procedural problems (but are typically not successful in the long term) and thus become increasingly more reliant on and crave more direct instruction. This perceived success exacerbates the problem of a reduced opportunity to develop cognitive skills and mathematical proficiency as the students demand more direct guidance in the future.

Constructivists advocate for indirect instruction wherein students engage in rich, often more ambiguous tasks, and receive individualized feedback that enables the development of their understanding. Constructivists argue that teachers, through informal and formal assessment and feedback, should facilitate students’ construction of knowledge through hands-on learning experiences and reflection opportunities. The teacher should act as a facilitator for the students’ processing to ensure they struggle productively and that their frustration does not become too high (Confrey, 1990). Confrey (1987) found in her research that traditional direct instruction, with its corresponding testing and grading practices, tends to measure short-term procedural skills and achievement and leaves students with lingering misconceptions that are quite resistant to correction through traditional instruction. Rather, students need to experiment
and reflect on the strength of their mathematical constructions to revise them in productive ways rather than relying on the teacher or other authority as an arbiter of truth and fact. However, Doyle et al. (1983) acknowledges that forms of indirect instruction are the most difficult to teach in practice, and thus practical forms of indirect instruction, including assessment and feedback, need to be researched.

**Feedback as Grades**

**Grading for Compliance and Motivation**

Feedback from teachers in classroom environments is often inextricably linked to grades. Almost all educational and professional development materials for pre- and in-service teachers argue that teachers should assign grades that reflect the extent to which a student understands the learning objectives (Randall & Engelhard, 2010). However, Randall and Engelhard (2010) studied how teachers assign grades in practice and found that, even though teachers agreed grades should reflect student learning and proficiency, the same teachers assigned diluted final grades that measured a combination of ability, achievement, behavior, and effort rather than mathematical proficiency. Teachers, of course, use grades in their courses to gather and report information about assignments and understanding, but they often focus and prioritize grades as a means to influence student behaviors and offer an extrinsic motivator for students to complete assignments. Assignments may include items such as homework, classwork, projects, and assessments. Examples of student behaviors include attendance, participation, and how well students follow classroom expectations.

Alas, social scientists have produced robust research over the last century that suggests extrinsic motivators narrow focus, incentivize production, and inhibit creativity. In an older study cited by Pink (2009), Duncker (1945) ran an experiment called, “The Candle
Problem” in which he gave each participant a box full of tacks, a matchbook, and a candle. He tasked the participants with fixing the candle to the wall such that the wax did not drop on the table. Duncker told some of the participants that they intended to observe how long the task took and told the other participants that the fastest participants would earn a prize. The participants with the allure of a prize took, on average, three minutes longer to complete the task because of a concept called functional fixedness – their focus was too narrowed to think creatively about the materials, specifically that they could use the box that contained the tacks by dumping out the tacks, literally thinking “outside the box.” Glucksberg (1962) extended the experiment by considering how changing the task slightly would affect the completion times. For this experiment, he separated the participants into similar reward or no-reward groups, but further split those groups in two – one in which the tacks remained in their container, and one in which the tacks and box were separate. This time, with the solution to the task much clearer in the latter group, the incentivized group performed better. In other words, extrinsic motivators work very well for rote and mechanical tasks, the exact opposite of the creative and deep mathematical proficiency and thinking that math educators strive to develop within their students. When students, educators, and other stakeholders treat grades as a reward-and-punishment system, a deeply ingrained cultural practice in America at the time of writing, it dulls this rich and connected student thinking and blocks creativity.

Prevalent grading practices, by design, tend to elicit compliance from students by explicitly calculating student behaviors as a factor in the grade rather than measuring mathematical proficiency. Even when individual grades on assignments are focused on understanding, such as that demonstrated on an assessment, the assessment itself tends to test narrow procedural skills based on what will be easiest to grade rather than measuring all of the
components of mathematical proficiency (Boaler, 2016). As Boaler (2016) phrased it in *Mathematical Mindsets*, “Such crude representations of understanding not only fail to adequately describe [students’] knowledge, in many cases they misrepresent it” (p. 141). She further contends that the knowledge demonstrated on tests has become so irrelevant in the modern world and so far removed from the adaptive, creative thinking that is in demand that leading companies such as Google have abandoned it as a measurement of how successful a potential employee will perform.

Bryant (2013), through his interview with Lazlo Bock, senior vice president of people operations at Google, reported that test results and Grade Point Averages (GPA) do not correspond at all with success in the workplace, in part because academic classrooms are such artificial environments. His point about artificial environments is further illustrated by the fact that educators frequently lament that they struggle to engage students with *any learning at all* – much less the deep and creative sort – without the allure of earning a desired grade or threat of low grades, colloquially referred to as the “carrot” or “stick.” Kohn (2010) states, “…It’s not really possible to motivate anyone, except perhaps yourself. If you have enough power, sure, you can make people, including students, do things. That’s what rewards (e.g., grades) and punishments (e.g., grades) are for… The more you rely on coercion and extrinsic inducements, as a matter of fact, the less interest students are likely to have in whatever they were induced to do.” In fact, connecting back to Google, the company’s workplace structure follows their philosophy in practice. The company focuses on building community, peer-to-peer coaching and collaboration, creativity and innovation, and flexibility in work schedule and format. Google’s workplace structure embodies for their employees what constructivism emulates in the classroom with students.
There does exist some controversy as to how rewards can diminish intrinsic motivation. Cameron and Pierce argued in their 1994 meta-analysis of over 100 studies (a result which they defended in 1996, 1999, and 2001 in response to rebuttals) that any negative effects of rewards only occurred under very restricted conditions and that these conditions can be easily avoided. However, this result was in contrast with dozens of other reviews of the same literature (e.g., Bates, 1979; Deci & Ryan, 1985, 1987; Kohn, 1993; Lepper, 1988). Several researchers rebutted the article directly (e.g., Kohn, 1996; Lepper et al. 1996; Ryan & Deci, 1996). The rebuttals argued that Cameron and Pierce, perhaps deliberately to defend their “behaviorist theoretical turf” as alleged by Ryan and Deci, drew erroneous conclusions by systematically and consistently misusing meta-analytic procedures. The challengers allege that Cameron and Pierce ignored important distinctions and misrepresented results, used faulty methodology in order to camouflage the detrimental effects of extrinsic motivators with a single effect size, overly simplified results with crude representations of student response behaviors, ignored evidence that contradicted their claims, and demonstrated such clear bias that the audience can predict what conclusions Cameron and Pierce will draw at the outset of the analysis. Lepper et al. (1996) acknowledge that while all researchers are susceptible to strong opinions and bias in their research, “Cameron and Pierce are neither less, nor necessarily more, biased than the authors of the many reviews that have come to the opposite conclusion.” Essentially, Cameron and Pierce’s research falls victim to the same reductive issue that grades do: using a single, overall measure to report varied and nuanced information becomes quite meaningless and oversimplified, despite the illusion of information and ease of acceptance by those who receive it. Constructivist researchers and educators argue that a constructive and student-focused classroom climate should replace such behaviorist measures; classroom decisions on curriculum and norms should
promote students’ own autonomy and inclination to become more competent and make sense of the world around them.

**Grading, Epistemological Beliefs, and Mindsets**

In addition to reducing intrinsic motivation to learn, compliance-based grading also affects students’ epistemological beliefs, that is, their belief about the nature of learning and knowing. Epistemological beliefs vary by subject; mathematics, in particular, tends to suffer from a culture of unproductive dispositions and fixed mindsets that disproportionately affect groups that are underrepresented in Science, Technology, Engineering, and Mathematics (STEM) fields (e.g., Aronson et al., 2002; Boaler & Sengupta-Irving, 2006). Teachers’ epistemological beliefs greatly influence students’ epistemological theories (Hofer, 2001). See Figure 1 below for a diagram that shows how teachers’ epistemological theories, that is, their beliefs about the nature of knowledge and knowing, influence students’ epistemological beliefs through their selection of classroom tasks and pedagogical practices (Hofer, 2001).

![Diagram](image)

*Figure 1. Working Model of How Epistemological Theories Influence Classroom Learning (Hofer, 2001)*
Steiner (2007) found in a study of students’ epistemological beliefs that one-third of students felt that mathematical understanding is very important. Unfortunately, Steiner also found that even those students who valued mathematical understanding in this way felt no agency over their own mathematical learning, a concept referred to by Booker and Goldman (2016) as epistemic authority, or the concept of exercising of the right or power to know. Instead, students expressed that their understanding was measured by external measures such as grades on homework and tests (Steiner, 2007). A disastrous effect can occur if these external measures come in the form of low grades as it can cause students to blame themselves yet feel helpless to improve (Selby & Murphy, 1992). In fact, at least one study showed that these students may not be entirely wrong about the challenges of improving low grades: Jensen and Barron (2014) found in their study of grades in a college biology course that students’ final grade in a course was highly correlated with their first exam or midterm grade, a result that has been replicated in similar studies. Alternatively, students may withdraw and dismiss grades entirely as irrelevant and meaningless (Guskey, 2011). If the entire culture and premise of the classroom is to perform and complete tasks to earn a grade, but many students place no importance on their grades, then the student now has no reason to complete tasks at all.

Other research has found that many students hold the unproductive belief that mathematics knowledge is justified mostly by authority and expertise, and an answer becomes true when a teacher verifies it (e.g., Cobb, 1986; Confrey, 1990; Frank 1988; Lampert, 1990; McLeod, 1992; Spangler, 1992; Steiner, 2007). This external concept of knowledge and justification overrides more meaningful measures of mathematical understanding, such as those outlined by the eight SMPs in the CCSS, the five strands of mathematical proficiency outlined by
Kilpatrick et al. (2009), and the descriptors of powerful mathematics knowledge constructions described by Confrey (1990).

See Table 1 below for a list of types of epistemological beliefs and their corresponding levels of belief (Hofer and Pintrich, 1997 as cited in Steiner, 2007).

Table 1

<table>
<thead>
<tr>
<th>Epistemological Beliefs</th>
<th>Perceptions of Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low End</td>
</tr>
<tr>
<td></td>
<td>Least productive for deep learning</td>
</tr>
<tr>
<td><strong>Nature of Knowledge</strong></td>
<td></td>
</tr>
<tr>
<td>Certainty of Knowledge</td>
<td>Knowledge is unchanging.</td>
</tr>
<tr>
<td>Simplicity of Knowledge</td>
<td>Knowledge is an accumulation of discrete pieces of information.</td>
</tr>
<tr>
<td><strong>Nature of Knowing</strong></td>
<td></td>
</tr>
<tr>
<td>Justification of Knowledge</td>
<td>Knowledge is justified by what feels right.</td>
</tr>
<tr>
<td>Source of Knowledge</td>
<td>Knowledge is handed down through teachers.</td>
</tr>
</tbody>
</table>

**Grading: An Equity Issue**

Another major problem with measuring student compliance with grades is that, historically and presently, these grading practices are based on the normative expectations of assimilation – grades are assigned based on the what the teacher values and measures (see
Brookhart et al., 2016; McMillan, 2001; McMillan et al., 2002). In other words, these grading practices act as transactional and artificial extrinsic motivators for students to act in the ways that are acceptable to the dominant culture, and then all the grades are reduced to a single, confounded grade or score that is impossible to interpret (Brookhart & Nitko, 2008; Cross & Frary, 1996; Guskey, 2001). Simultaneously, they impose such an inequitable and narrowly defined target for many students that there is no flexibility for them to meet the behavioral expectations as they attempt to eschew their own culture and norms to meet the values of the teacher. And to reiterate, while grades have no correlation with students’ potential for success in the workplace (Bryant, 2013; Boaler, 2016), grades do continue to act as barriers to students with regard to accessing opportunities in academic and other environments as grades are used as predictors of dropping out of school, applying and being admitted to colleges, and their academic success in higher education (Brookhart et al., 2016). While low grades clearly act as barriers to access, no research exists that low grades motivate students to try harder or learn better. However, it is worth repeating that low grades can certainly prompt students to withdraw from learning, dismiss grades as irrelevant and meaningless, feel helpless about improving them, and/or develop a negative conception about themselves personally as they blame themselves for earning low grades but simultaneously feel disempowered to improve (Guskey, 2011).

These grading practices can be extremely inequitable in that they developed from deficit approaches to teaching and learning based on the assimilation to the dominant group, indoctrination, and colonization (Paris, 2012). Dominant grading practices tend to measure all students against a normative form of knowledge based on the social positioning of the teachers who are disproportionally among the dominant group. While increasingly more educators are questioning and disrupting these practices in their own classrooms as evidenced by the very
recent increase in the availability of the literature on the topic, their presence continues to
dominate in classrooms. This is especially in mathematics classrooms where students are
typically graded down for making mistakes, mistakes which are critical in the learning of
mathematics (Boaler, 2016). Grading using this deficit model is a practice which takes away
valuable opportunities to learn and grow from mistakes and perpetuates a very negative mindset
about the role and value of making them (Boaler, 2016). Furthermore, the prevalent grading
practices of today ignore what Bang and Vossoughi (2016) refer to as epistemic heterogeneity –
that is, the validity of various forms of knowledge free of assimilative requirements – by firmly
placing the authority of knowledge on the teacher rather than the students.

Clearly, the potential of the negative impact of grading and feedback practices cannot
be understated. Feedback and grading are ubiquitous educational practices that therefore relate to
systemic, large-scale learning ecologies. Learning ecologies consider the physical, social, and
cultural context in which learning takes place, and hence, they go beyond the individual students
or classes that we teach (Vakil et al., 2016). Grading has been historically and systemically used
as a deficit model against marginalized groups in both individual classrooms and system-wide by
offering lower-level math courses and content to students based on their grade performance in
previous classes, a process known as tracking (Boaler, 2016). In the lower-level math classes,
these students are offered less “opportunity to learn” (OTL), a key factor in student achievement.
Once students get into a lower-level track, they rarely get out of it (Boaler, 2016).

Feedback over Grades

Top mathematics education researchers, social scientists, and innovative educators in
recent grassroots movements have resolved to address the shortcomings of mathematics
education concerning grades. Boaler (2016), Barnes (2015), and Sackstein (2015a; 2015b), in
particular, are strong advocates for rethinking grading. But other educators, too, have responded with several solutions and initiatives, some even eliminating grades in their classes entirely (Barnes, 2015). Research and proposals for rethinking grading has coincided heavily with the pendulum-swing back to conceptual- and reasoning-based mathematics standards with the implementation of the CCSS following the procedurally-focused No Child Left Behind initiatives of 2001. Educators, such as the ones listed above, contend that traditional grading practices fall short of a well-rounded development of mathematical proficiency and advocate for more holistic evaluation of student learning by providing feedback rather than grades and points. In fact, Hattie (2009, 2012) synthesized over 800 meta-analyses about how people learn and found that feedback has one of the highest effects on learning as compared to other instructional practices. Though the term “feedback” can be difficult to define, Sadler (1989) proposed that educators and researchers define feedback as the process that bridges the gap between what a learner currently understands and what a teacher desires a learner to understand. Hattie and Timperley (2007) added further that “feedback is a ‘consequence’ of performance” (p. 81).

While grades can serve many purposes as outlined in the previous section (e.g., measuring understanding, reporting standardized results to stakeholders, influencing behaviors), eliminating grades can produce several desirable results. Eliminating grades can focus the classroom on real, tangible learning and the intrinsic motivation that arises from autonomy, mastery, and purpose, as Pink described in his 2009 “Puzzle of Motivation” Ted Talk and as Kohn (1992, 1993, 2006, 2010) advocates for in his extensive collection of articles and books. It can enhance the value of mistakes, intellectual risk-taking, a growth mindset, and student agency over their own learning (Boaler, 2016). It can shift the expectation of assimilating to the dominant culture and disrupt inequitable teaching practices.
The challenges of eliminating all grades, however, can seem insurmountable for a full-time educator. At the crux of eliminating grades is ensuring that the culture of the classroom and the assignments pique the interest of the students to activate their intrinsic motivation (see Barnes, 2015; Boaler, 2016; Kohn, 2010; Sackstein, 2015a; Sackstein 2015b). Educators must adapt or recreate lessons entirely to elicit such a response from students as student interest, but if they seek research and resources to do so, they may find through a lack of lessons that suit their needs that engagement with a lesson has not always been a priority over the last centuries until very recently. And Kohn (2010) further prods that these changes cannot be artificial; students should not simply have to pick “this over that,” and the class should be involved in collective decisions as to how they want to investigate concepts. Leaders and activists of the gradeless movement advocate for a lot of student choice in the selection of assignments that meet the learning goals and within those selected assignments (e.g., Barnes, 2015; Boaler, 2016; Krall, 2018; Sackstein. 2015a; Sackstein, 2015b). To summarize, challenges of a gradeless structure include communicating rationale and structure of a gradeless classroom with students, caretakers, and administrators; shifting the deeply ingrained culture and rarely questioned practice of grades and convincing stakeholders of the value of a gradeless structure; the time commitment to restructure curriculum to activate intrinsic motivation within students and establishing productive classroom norms; and meeting external requirements from administration at the site, local, state, or national level.

With all of these challenges at the forefront of any educator considering a gradeless classroom, this structure is not feasible for many teachers given their time constraints in the current educational settings and demands of the current educational climate. The challenges seem prohibitive to many teachers who would even consider it. Hence, many of the educators
and researchers mentioned previously acknowledge and encourage a *deemphasized* grade approach – “grade *less*” rather than “gradeless,” as some educators in the movement refer to it – as a promising way to mitigate some of these obstacles and leverage some of the benefits of grades. While a final grade will be assigned in a deemphasized grading approach, the feedback (and feed up and feed forward, as will be described in the next section) is the primary focus.

As such an integral component of education, one would imagine there would exist substantial research regarding the topic of feedback. Though research regarding feedback does exist, the specificity of feedback in the mathematics classroom is limited. Furthermore, the extent of the effectiveness of feedback varies incredibly and depends on several factors (Hattie, 2007). Hattie identifies the factors influencing the effectiveness of feedback as type, level, timing, to what extent the learner accepts the feedback, and from whom the feedback comes (2007). The fact that so many factors depend not just on the educator's instructional decisions but on the learners for which the feedback is intended (Hattie & Timperley, 2007; Hattie, 2012) highlights the need for educators to understand and consider how students perceive and respond to feedback in the mathematics classroom. Furthermore, since the goal of feedback is to bring the learner from what they understand to what the educator desires the learner to understand (Sadler, 1989), educators need to consider how they can encourage and promote productive student response behaviors to feedback in the mathematics classroom in a way that improves their understanding of mathematics.

The next sections will summarize current research surrounding feedback including framework models, timing, and formats.
Feedback Framework

As feedback is difficult to define, a means through which to measure and build feedback is critical. The concept of feedback is intuitive to layfolk, but the effective implementation of it in a mathematics classroom is complex. According to Hattie (2007), feedback can be powerful for learning, but it is not as simple as something a teacher does or gives to a student. Feedback can be provided or sought by students, teachers, parents, peers, or other people with whom one might interact throughout the day and "can be accepted, modified, or rejected" by the intended recipient (Hattie, 2007, p. 82). Obviously, we would like students to accept feedback or at least modify it. Unfortunately, Hattie (2012) determined that teachers perceived that they provided an abundance and variety of feedback to students, but students received very little of it. Indeed, when Nuthall (2007) studied the dialogue of students by placing microphones on them throughout the school day, he found that students received most of their feedback from other students and further, that most of that feedback was wrong. While teachers conceive feedback as constructive comments, criticisms, corrections, additional or elaboration of math content, students place a higher value on other aspects of feedback (Hattie 2012). Specifically, students want to know precisely where they are headed in their learning to dedicate their attention and focus accordingly (Hattie, 2007).

Thus, given the complexity of feedback theory and practice, it is clear from the research that a framework upon which teachers can plan and implement effective feedback strategies is necessary. The framework overview that follows will begin by describing categorizations or types of feedback, then describe the levels of student learning and student tasks at which these categorizations of feedback can occur. Next, the framework will describe dimensions of feedback; these describe what educators should seek to address when giving any
type of feedback at any level of feedback. Finally, a model for assessment feedback is provided that educators can use to address the aforementioned parts of the framework.

**Categorizations of Feedback**

Hattie and Timperley (2007) proposed a framework model for feedback with the following main categorizations. Each is characterized by a question of the learner and is italicized.

1. **Feed up** – This category concerns itself with the learning goals of the lesson. The learner wants to know, “Where am I going?” The teacher should provide the learning objectives and success criteria that the student needs to demonstrate to show competency.

2. **Feed back** – The teacher should provide information as to what progress the learner is making toward the goal. The learner wants to know, “How am I getting to where I am going?” This information includes examples of prior performance and the next steps to progress learning.

3. **Feed forward** – The student wants to know, “Where to next?” The teacher needs to provide activities or direction for the student to deepen or extend the learning that has taken place. Activities or lessons may take the form of additional challenges, prompts eliciting metacognition about problem-solving strategies and learning process, or prompts that make evident any areas of understanding that is lacking.

Hattie and Timperley (2007) go as far as to say feed forward is the most impactful cause of learning but is also the most neglected aspect of feedback in the classroom.
Levels of Feedback

The above categories of feedback each occur throughout the various levels of learning. Hattie and Timperley (2007) propose the following levels of learning at which feedback takes place:

- **Task-level feedback** is about the specific task at hand. The teacher should provide the learning goal, specificity of what the student accomplished while completing the task, and what the student needs to do next.

- **Process-level feedback** includes general strategies about accomplishing tasks such as determining to what extent an answer makes sense, how to learn from errors, establishing relationships between ideas, and how to seek out the information needed for completing a task.

- **Self-regulation or metacognitive feedback** primarily focuses on the individual’s independent learning, a willingness and capability to self-assess, effectively internalize and make use of external feedback, and attribute learning as a product of effort and not an innate “you have it, or you don’t” fixed ability.

- **Self- or personal-level** feedback is unrelated to the task and may take the form of praise or even extrinsic physical rewards or punishments.

Of the four levels of feedback, Hattie (2009, 2012) claims that personal-level feedback tends to be the least effective by all measures. In fact, personal-level feedback can detract from learning in several cases. Rattan et al. (2012) determined that positive feedback communicates an entity (fixed) theory of intelligence and learning. Praise can simultaneously comfort students and demotivate them in seeking understanding and engaging with mathematics or sciences, the opposite of the intended goal for feedback. Though it is clear that humans take
pleasure in receiving praise (e.g., Cameron & Pierce, 1996), the praise overshadows any constructive feedback in a way that the learner only hears the praise (Hattie, 2012). They may also come to rely on the praise to engage with a task on any level, as behaviorist learning theory would explain.

Most educators today acknowledge the vast limitations of behaviorist learning theory, which is the theory that best describes feedback that occurs at the self or personal level. Kohn, in particular, has dedicated their life’s research to challenging behaviorist classroom practices and advocating for alternatives that support students’ intrinsic desire to learn (see 1992; 1993; 1996; 2004; 2006; 2010; “Do You Think That Just Because You're a New Teacher, You Can't Create Meaningful Change in Your Classroom? Alfie Kohn Begs to Differ,” 2013). As a result, this study focuses only on the first three levels of feedback and strategically avoids the use of external stimulus or motivators (and demotivators), such as grades or praise, since they are well-studied and the limitations well-documented.

**Dimensions of Feedback**

Robinson et al. (2015) considered the dimensions of feedback in the mathematics classroom as process, engagement, and richness. *Process* concerns itself with the communication of mathematics, how students should set up solutions, learning to think mathematically, and the development of mathematical skills. *Engagement* means that the feedback should encourage more and deeper learning, reflection, self-awareness, and independent learning. Finally, *richness* determines whether or not the feedback enhanced the learning experience and if it complemented and supplemented existing feedback.
A Model of Feedback

Barnes (2015) recommends the following feedback model for both summative and formative assessments in his book, Assessment 3.0. He refers to his model as SE2R: Summarize, Explain, Redirect, Resubmit.

- **Summarize** by providing a one- or two-sentence summary of what the student has accomplished on a task;
- **Explain** by sharing detailed observations of what skills or concepts have been mastered based on the specific task guidelines;
- **Redirect** by guiding the student to the lessons, presentations, or models that they need to review to achieve an understanding of the concepts and mastery of skills; and
- **Resubmit** by encouraging (or requiring) the student to revisit the task after redirection, rework it, and submit for more feedback.

Feedback in the Mathematics Classroom – Forms and Effectiveness

Prior Knowledge and Timing of Feedback

Hattie and Timperley (2007) argue that feedback occurs most powerfully when addressing content about which a learner has some prior level of understanding about a topic, either from previous experience or prior instruction, rather than a complete absence of understanding. More recent studies, however, contend that feedback may *detract* from performance if provided when a student has mid-to-higher levels of understanding (Fyfe & Rittle-Johnson, 2017; Fyfe et al., 2012). When one considers the recent movement of guided discovery learning over learning purely through direct instruction or purely student exploration (Fyfe et al., 2012), it becomes especially clear that research needs to develop an understanding of the following:
• what types of feedback most strongly affect learning in the mathematics classroom;
• at what point in a learner's understanding feedback should occur and how much feedback a teacher should provide at these points; and
• how mathematics teachers can feasibly implement effective feedback strategies that are individualized for specific students and yet allow the teacher to attend to the needs of an entire classroom of diverse learners.

The literature is clear: the timing of feedback in relation to a student’s level of understanding and their progress throughout a task correlates highly with student perception and response to the feedback. Hattie (2012) recommends teachers provide task-level feedback for novice learners, process-level feedback for moderately proficient learners, and regulation or metacognitive-level feedback for competent learners. Teachers should determine their students' level of understanding through formative assessments that openly invite questions and the sharing of mistakes and errors to reveal misunderstandings that students need to correct with instruction (Hattie, 2012). According to Hattie (2012), formative assessment provides teachers with information about their teaching, in particular, what they taught well and what they need to reteach and to whom. Since feedback must occur in areas that do not yet understand well (Hattie & Timperley, 2007), the formative assessment piece is essential.

Other examples further illustrate the need for formative assessment. Fyfe et al. (2012) determined that the timing of feedback in their lesson played a massive role in whether or not students benefited from the feedback. In their guided-discovery mathematics lesson with second- and third-grade participants, students with any moderate prior-knowledge of the task performed worse if they received feedback while completing the task. Only students with no to low prior-knowledge benefited from the feedback. The expertise reversal effect explains this surprising
result (Fyfe et al, 2012). The research characterizes the *expertise reversal effect* as an increased cognitive load induced by the new feedback since the learner now has to reconcile their prior knowledge with the new feedback.

In a more recent study, Fyfe and Rittle-Johnson expound upon their previous research; their study showed that *no* feedback resulted in higher retention of learning for some learners (2017). The students who received feedback during the practice test performed better immediately following the task, which seems intuitive. But these students performed worse than their peers on the post-test that took place a week later when compared to their peers that received no feedback during the practice task. According to the research, educators should consider implementation of instructional practice, that is, seeming less effective at first but ultimately results in desirable long-term learning, as *desirable difficulties* (Fyfe & Rittle-Johnson, 2017).

Educators may strive to elicit these desirable difficulties when considering their feedback method. Fyfe and Rittle-Johnson (2017) explained, when the instructor or researcher provides feedback, a decrease in the student's mindfulness about the task students may cause them to over-rely on feedback instead of their own error-correction and problem-solving strategies or engage the *expertise reversal effect* in which students struggle to integrate the new feedback with their prior knowledge or learning occurring throughout the task. It is worth noting that, according to Hattie (2012), the level of feedback provided during these tasks may not have been appropriate for the level of the students which may need to be addressed by a more robust formative assessment.
Effects and Applications of Feedback in Classroom Practice

Almost all of the literature found for this study regarding the effects and applications of feedback in classroom practice focused on computer-based feedback, perhaps as a result of funding or feasibility. One study in particular focused on a non-computer-based application of feedback which studied the effects of feedback tailored to bilingual students’ mathematics needs on verbal problem-solving, attending specifically to English learners (Cardelle-Elawar, 1990). Since research characterizes the pedagogical strategies that attend to the needs of English learners as effective pedagogical strategies for non-English learners (Cline & Necochea, 2003), the results of this study may be generalizable to other groups of learners as well. The researchers in this study individualized student feedback by identifying the key error the student made, the probable reason for the error, what the student did right, and how they could guide the student to avoid the error in the future. Additionally, the students were asked to engage in metacognitive processes by self-identifying from a list of specific statements they felt applied to them. This list ranged from "I don't understand the meaning of the following word(s)… in this problem” to “I know how to calculate the solution and work out the operation(s) needed” (Cardelle-Elawar, 1990, p. 172). This approach of teachers individualizing the feedback for each student and having students self-identify metacognitive and self-regulation statements showed promising learning results for the group of students studied. However, the researchers noted that this research was theoretical in nature due to the sample of students and teachers. Further, this type of feedback may not be feasible for a typical classroom teacher to implement.

The feasibility of effective feedback systems had sparked a research interest in computer-based feedback. Fyfe and Rittle-Johnson (2016), in their study about the benefits of computer-generated feedback on mathematics problem-solving, showed results similar to that of
the non-computer feedback. The researchers compared feedback that followed each problem immediately to a more summative approach to the computer feedback provided once the student finished the entire task. Feedback was most effective for the second- and third-grade students with low prior knowledge. However, unlike their previous study, students who had prior knowledge performed better in the post-test whether provided immediate feedback or summative feedback. Ironically, students with some knowledge in the domain performed lower if they received feedback immediately, again likely explained by the expertise reversal effect. Students with mid-level prior knowledge may struggle to reconcile the immediate feedback with their current knowledge whereas students with low knowledge have very little conflicting information and students with high prior knowledge can integrate the new information more easily with the existing knowledge. However, like their peers with low and high prior knowledge, students with mid-level prior knowledge still performed higher with the summative feedback. The researchers concluded that providing even minimal feedback on the computer to students affected student learning positively. They suggested this may occur because students are not concerned with how they may be perceived by others, indicating a promising result of using technology to provide feedback to students.

Unfortunately, feedback in computer-based assessment can focus too much on lower-level learning such as memorization simply as a result of software programming capabilities at the time of research. For complex learning to take place, complex feedback must be provided that goes beyond simple corrective feedback (Attali & van der Kleij, 2017). Attali and van der Kleij (2017) examined three main types of computer-based feedback:

- Knowledge of Results (KR) – verified the correctness of the student’s response;
• Knowledge of Correct Response (KCR) – provided the correct response to the student; and/or
• Elaborated Feedback (EF) – provided additional information such as strategic hints, explanations, or worked out examples;

Students sometimes received feedback individually and other times in combination with one another. These three categories focused only on task-level and process-level feedback. EF tended to provide the desired higher levels of learning compared to KR and KCR feedback. However, the advantages of EF were limited if a student initially showed a partial understanding of the content prior to the task, consistent with the expertise reversal effect indicated in some of the other literature. The study showed that as the knowledge and understanding of the student increased, the process-level feedback focusing on problem-solving strategies was more effective than studying worked-out examples at the task-level.

In a similar study, Attali (2015) tested how feedback that allowed a student multiple attempts in solving a problem affected their performance in a similar task. This particular study showed the multiple-try feedback fostered increased student performance compared to feedback providing the knowledge of a correct response without multiple tries. Feedback allowing multiple attempts that included a hint in addition to the knowledge of the correct response was even more effective than the multiple-try feedback without hints. The research showed these effects for both multiple-choice and open-ended questions but indicated it was more effective for the open-ended questions when comparing pre- and post-tests. The researchers suggested that this study, combined with other available research, highlights the importance of immediate feedback. The results also illuminate the obstacles of providing feedback since teachers alone would not be able to feasibly implement such feedback strategies
in a classroom without advanced technological solutions. Furthermore, this technology-based feedback will have the added complication of ensuring the feedback is complex enough to address more than just low levels of learning.

Another study used technology in a completely different way than the methods described above. The instructor created screencast videos recording explanations of mathematical concepts and/or worked solutions to mathematical problems, attending to feedback generalized for an entire class of students (Robinson et al., 2015). After written coursework was marked, the instructor recorded feedback related to common errors and themes of written solutions. Once the instructor passed back the marked coursework to the students, he gave the students access to the screencast. Despite the more generalized nature of the screencast feedback, students reported that it was more personal and easier to understand than other forms of feedback. Additionally, the screencast feedback promoted students to pause, revisit, and think about their answers in more depth than they would be able to with other forms of feedback. Students were very engaged in using the feedback to improve their learning.

**Promoting Students’ Proactive Recipience of Feedback**

Research is limited as to how teachers can strategically give feedback to students in a way that will increase their understanding in any academic subject, so educators and researchers will find an even greater paucity of research in the more specific field of mathematics education. Winstone et al. (2017) systematically reviewed what they described as a “highly fragmented and somewhat atheoretical” (p. 31) research base. Since they could not find an existing theoretical framework for categorizing what they termed *proactive recipience processes*, these authors developed a taxonomy and theoretical base (though the authors make it clear that it is not a
framework per se) for future work. They categorize recipience processes into four distinct categories, which they call SAGE:

- self-appraisal – making judgments about oneself;
- assessment literacy – understanding the grading process and using it to assess their own results and learning;
- goal-setting and self-regulation – determining specific desired outcomes for oneself and adopting the behaviors needed to attain them; and
- engagement and motivation – a commitment and desire to develop and make progress.

They posit the following: teachers need explicit training on how to promote students to engage in these recipience processes. Further research needs to be done in this regard to determine first, the pedagogy of feedback, and second, how trainers can effectively teach these strategies to in-service and pre-service teachers. Both theoretical and practical research is needed.

**Conclusion**

The research regarding feedback is abundant, yet general. Current research does not adequately address feedback specific to mathematics learning. And yet, feedback is a critical component of education (Hattie, 1999, 2007; Hattie and Timperley, 2007; Sadler, 1989). The research that is related to the mathematics classroom is mostly limited to technological applications of feedback involving low-level memorization-based lessons, computer-assessment environments, or screencast feedback generalized to an entire class.

Almost all of the literature called for research to determine how teachers can feasibly implement feedback in classroom settings, how to generalize feedback across different domains,
and how teachers can distinguish between prior levels of knowledge to determine when feedback is no longer necessary. Very little of the research addresses how teachers can implement effective feedback strategies that are both personalized for each learner and have the capability to attend to an entire classroom of students.

Winstone et al. (2017) add that the research is still highly limited and inconclusive, despite the recent surge of interest in students' proactive recipience of feedback. Hattie (2012) suggested that students perceive feedback to the whole class as being meant for others in the classroom and not them individually. And yet, the results of screencast feedback generalized to the group demonstrated that students positively perceived and received the provided feedback in that students showed the desired process, richness, and engagement (Robinson et al., 2015). Research still needs to address these discrepancies.

This study aims to address the research questions of how students perceive and respond to a reflection-and-analysis method of feedback on exams in the mathematics classroom, how the proximity of students’ self-assessment scores to their actual scores changes from one exam to the next, and how students’ rationales for their self-assessment scores varied from one exam to the next.
CHAPTER III

METHODOLOGY

Introduction

This research used a complex mixed methods case study research approach through a combination of action research design and participatory design research (PDR) to investigate the student experience and learning in a developmental College Algebra course with respect to a Reflection and Analysis Process (RAP) for exams. For the RAP, students take an assessment, receive their exams back with written feedback but no scores or grades, and then complete a process in which they reflect on the understanding they showed at the time of taking the exam and analyze their work for errors and mistakes to learn from and revise them.

The students complete the RAP assignment outside of class and are given about three days to complete it (students may be allowed one to two additional days if the three days happened to land on a weekend or holiday). Late submissions were accepted and missing submissions followed up on as the submission of the RAP was required for students before any instructor scores were made visible to them. As a note for the reader, in the event that a student did not submit a RAP, the instructor exam scores were not entered in the gradebook until final grades had to be assigned upon completion of the course.

The RAP can be used for other assignments, but for this study focused only on completion of the RAP after exams. Since studies have shown that scores and numbers detract from feedback (e.g., Boaler, 2016; Kohn, 2013), the RAP delays any numeric grades or scores until after students have engaged in the RAP; the primary focus is the non-numeric aspects of feedback. While eliminating scoring was considered, external constraints and reporting
requirements demanded a numeric measurement. Conveniently, the score did provide an additional measure to contribute to the robustness of this study’s findings and compensated for the shortcomings of gradeless models (e.g., communication of gradeless model with stakeholders, external reporting requirements of a single grade or number, setting class culture of no grading).

The numerical rubric scores serve as a quantitative measurement in this study. While scores have the risk of reducing a lot of information into a single crude measurement, a quantitative comparison can provide a very clear cue to minimize the risk of the student intentionally or unintentionally dismissing any contradictions of qualitative feedback as being more-or-less the same as their own when in actuality, it was very different. The instructor can use any significant differences between the student score and instructor score as a cautionary flag for instructors to revisit their own evaluation of the student’s understanding. Since the numerical score is based on categorical descriptors in the 4-point rubric and the scores are provided through a more holistic lens for an individual or group of concepts – as opposed to, say, taking points off for individual mistakes on problems – it may prevent some of the loss of information that a score can cause. Most importantly though, the numeric 4-point rubric score can provide a clear check-in for both the student and instructor. See Table 2 below for specific information about each score.
Table 2

4-Point Rubric for Scoring a Single Problem or Task (adapted from CPM Educational Program, 2013)

4  
*Fully Accomplishes the Purpose of the Task*
Student work shows full grasp and use of the central mathematical idea(s). Recorded work communicates thinking clearly using some combination of written, symbolic, or visual means.

3  
*Substantially Accomplishes the Purpose of the Task*
Student work shows essential grasp of the central mathematical idea(s). Recorded work in large part communicates the thinking.

2  
*Partially Accomplishes the Purpose of the Task*
Student work shows partial but limited grasp of the central mathematical idea(s). Recorded work may be incomplete, somewhat misdirected, or not clearly presented.

1  
*Makes Little Progress Toward Accomplishing the Task*
Shows little grasp of the central mathematical idea(s). Recorded work is barely comprehensible, but there is definitely some logic present.

0  
*Makes Very Little (or No) Progress Toward Accomplishing the Task*
Shows very little or no grasp of the central mathematical idea(s). Recorded work is not at all comprehensible.

Considering the narrative instructor feedback and the students’ analysis of their errors for the RAP process, much of the analysis for this study is qualitative. The qualitative aspects of the study served to investigate the written aspects of the students’ RAPs with respect to their epistemological beliefs and if there was any shift from one exam to the next. The qualitative data also served as a critical component of PDR to investigate the students’ personal experience and opinions about the process as a way to involve them in the design of future iterations of the RAP assignment.

This section will first list the research questions followed by an overview of the research design. Details about the participants, research process, instruments, and data analysis procedures will follow.
Research Questions

This research aims to address the following research questions:

1. How do students’ self-assessment scores and instructor-assigned scores compare from Exam 1 and Exam 2?

2. How do students’ rationales for their self-assessment scores compare from Exam 1 to Exam 2? Specifically, in terms of epistemological beliefs (see Table 1 in the Review of Literature), what were the characteristics of any shifts that occurred from Exam 1 to Exam 2?

3. What is the student experience and perception of the RAP?

In particular, special attention was paid to student scores and rationales from Exam 1 to Exam 2 to see if there was a shift from rationales based on compliance (i.e., “Knowledge is justified by what feels right.”) and authority (i.e., “The teacher is the source of knowledge.”) to more evidence and mathematics (i.e., “Knowledge is justified by evidence, logic, and reason.”) rationales (Steiner, 2007).

Research Design and Worldview

This study used a combination of action research design and Participatory Design Research (PDR). Action research is characterized as an inquiry-based research approach “by the practitioner, for the practitioner” as a means of improving and directing future work based on an observed problem (Sager, 2000). The RAP emerged as a way to address the inequities and serious limitations of traditional compliance-based grading practices while mitigating the challenge of going entirely gradeless. For example, in a colloquial and metaphorical sense, teachers say, “Jump,” but only some students say, “How high?” Other students, however, exhibit apathy, learned helplessness, hopelessness, an often-invisible-to-the-authority lack of resources,
or some other manifestation of disengagement or noncompliance. My experiences in the classroom begged the question, “How do we, as educators, get students to deeply and creatively learn when our only tool is the artificially moving hoop through which we expect students to jump?” Action research did not completely address the iterative aspect of the RAP design based on student input nor the equity aspect of its design, and so PDR was integrated into the research design as well.

PDR is a relatively new design that has emerged in the field of education from Design-Based Research (DBR). In contrast to DBR, PDR strategically rejects and disrupts the distinctness of the categories of researcher, teacher, and participants, treating them as more inclusive and overlapping roles. All stakeholders play a part in PDR. PDR is popular with newer generations of scholars, often in collaboration with innovative and progressive veteran scholars (Bang & Vossoughi, 2016). There is a special issue of the journal, Cognition and Instruction, that is oriented around PDR in educational research. In their introduction to the special issue, Bang and Vossoughi claim, “…PDR is beginning to shape a newer generation of research epistemologies. These epistemologies may be essential for expanding our fundamental knowledge of learning as well as developing theory that can help create sustainable and transformative social change” (2016). The authors of the articles in this issue showcased exemplary PDR in education to highlight its emergent contributions and future direction in the field. Researchers using PDR seek to expand equitable forms of learning and teaching, often through a transformative philosophical worldview. Through the use of the survey following students’ completion of the RAP assignments for Exam 1 and 2, student feedback about the process was elicited, thereby including the participants in the design process. Furthermore, the
RAP was (and still is as of writing) constantly iterated to equitably serve students’ needs, mathematical development, and their general learning and academic development.

The transformative philosophical worldview shaped the design of the RAP and the research as is typical in PDR. Transformative researchers aim to address inequities and confront oppressive systems by taking political and/or social action (Creswell & Creswell, 2018). Of central importance in a transformative approach is the experience and perception of people in traditionally marginalized groups and their strategies to resist, subvert, and challenge constraints within systems of oppression. According to Mertens (2012), the transformative framework “directly engages the complexity encountered by researchers and evaluators in culturally diverse communities when their work is focused on increasing social justice” (p. 10). Transformative research focuses on power imbalances based on the intersections of gender, race, ethnicity, disability, sexual orientation, and socioeconomic class and the strengths that reside in communities that experience discrimination and oppression based on their cultural values and experiences. Since current grading and assessment practices can act as oppressive measures of students, a transformative philosophical approach informed the study.

**Research Approach**

Mixed methods approaches are ideal for combining both quantitative and qualitative data when one method on its own insufficiently addresses the research questions. Many synonymous terms have been used in the literature to describe and advocate for the mixed methods approach. These include *integrating, synthesis, quantitative and qualitative methods, multmethod, mixed research, or mixed method* (Creswell, 2018). The term *mixed methods* has been adopted by recent literature, such as the *SAGE Handbook of Mixed Methods in the Social & Behavioral Sciences* and *SAGE’s Journal of Mixed Methods* (Bryman, 2006; Creswell, 2015;
Tashakkori & Teddlie, 2010). The reader can find detailed information, including a history of mixed methods methodology, within such handbooks.

Traditionally, many studies have used quantitative approaches to measure the effectiveness of a treatment. For this study, purely quantitative data could have addressed the research question regarding how student self-assessment scores and instructor scores compared to each other from Exam 1 to Exam 2. However, by nature of quantitative research, it could not have provided responses and insight to the open-ended questions regarding students’ experiences with the process and their perception of the RAP. Nor could quantitative research have addressed any shifts of rationales as to how they chose which rubric score most closely represented their understanding. For that, qualitative data was solicited from the students both as a part of the process itself and as a separate addition to the process as part of the research. It is also worth noting that while the rubric scores are numerical, they are closely aligned with categorical information about the mathematical proficiency that was demonstrated. The numerical score is a useful means of comparison, but it obscures and collapses valuable information about student understanding that qualitative data can elevate. Hence, both quantitative and qualitative approaches were crucial in developing a more complete understanding. See Table 3 below for a list of the qualitative and quantitative data collected in this study. See Figure 2 below for a chronological timeline for each piece of data.
Instructor feedback on exams
• Student reflections and rationale for self-assessment scores
• Survey data regarding experience with process

Student-assigned scores
• Instructor-assigned (delayed) scores

The core design of the research is a convergent mixed methods design because both qualitative and quantitative data were collected at the same time with neither driving the other. Exam 1 was compared to Exam 2 using both quantitative and qualitative measures. The student survey occurred following Exam 2 and was included in the original planning phase to explain...
and provide insight into the results from investigating the relationships between Exam 1 and Exam 2. An explanatory qualitative data collection from the student survey then followed.

**Population**

The population for this study were thirty-four students enrolled in a College Algebra class with a corequisite support class in Spring 2020 at a large public university in Northern California. Students are designated for this class using a California State University (CSU) multiple-measures placement that considers high school grades, coursework, and standardized tests. Standardized tests that may be used for placement include the California Assessment of Student Performance and Progress (CAASPP), American College Testing (ACT), Scholastic Assessment Test (SAT), Advanced Placement (AP), International Baccalaureate (IB), and/or College-Level Examination Program (CLEP). If students do not meet the criteria in either the “Category I - Fulfilled Requirement” or “Category II - Enroll in GE Math Course” categories, they are considered “Category III - Enroll in a Supported GE Math Course.” The College Algebra with Corequisite Support is one of two courses for students designated as Category III at CSU, Chico (the other option is a statistics with corequisite support). College Algebra is specifically reserved for students with a STEM or STEM-intensive major. Students may also be designated for this course as a STEM or STEM-intensive major if they were designated “Category IV – Early Start Required; Enroll in Supported GE Math Course” upon admission to the university. The population of this course are Category III and IV students. Some students were taking the course for the first time while others were retaking it a second time. See Figure 3 below for a visual diagram of the categories.
Figure 3. Categories of Math Placement for Incoming First-Year CSU Students

In this course, there were 29 freshman and 5 sophomores. There were 14 female students and 20 male students. Seventy-four percent of the students were designated as underrepresented minority (URM) by the university.

Treatment

The true first stages of this research have taken place over five years prior to this research through anecdotal experiences in the classroom and informal research of grassroots practices. Several iterations of crafting and adjusting took place before the final RAP product for this study was borne. Each iteration progressed with the goal of improving the efficacy of leveraging student interest in earning a good (or passing) grade in the class while emphasizing the autonomy, mastery, and purpose required for the intrinsic motivation to learn. The hope is that instructors can leverage assessment as a continued and integrated learning experience for
students rather than a one-off, grade-focused assignment. Through this process, educators and students may maximize the learning potential that they can derive from each exam, which is typically heavily prioritized in a mathematics classroom, potentially taking with them transferable skills for future classes. The RAP assignment, as it looked for this particular population and study, was the result of these iterations and pilots at various stages of its development. Qualitative information through conversation and survey was collected from the students over the two years prior to this spring 2020 research. Mainly, this formative information helped clarify and revise the RAP prompts to emphasize the development of epistemological beliefs conducive to learning mathematics and to elicit the desired responses from students; eliminate parts that seemed contrived or lacked any real merit for the students; and added parts to develop deeper reflection, analysis, and metacognition for their current and future learning endeavors.

The institutional review board approved the research project. The particular class for this study was structured so that students were scheduled to take exams once every four weeks. Only the first two exams are considered in this analysis. In Exam 1, students took the exam in class during a 50-minute class period. Exam 2 was originally scheduled to take place during a 50-minute class period but had to be rescheduled for the 75-minute class period the day prior to its original scheduled date because of the cancellation of classes due to COVID-19. However, all but three students finished within the 50-minute time frame since the exam was designed to be completed in 50 minutes. Each exam, with the exception of the final which would have more problems with a larger time allotment, has two-to-three problems that have multiple conceptual pieces that are linked together. These problems focus on in-depth analysis of various types of functions and the necessary procedural skills required for analysis.
For example, Exam 1 focused on linear functions. The problems range from procedural skill-based prompts to deeper conceptual problems. The problems required students to analyze a tile pattern through the lens of the multiple representations of each iteration of the tile pattern (figure number), graphs, tables, equations, and written descriptions. See the Appendix for reference. As a necessary skill, one of the problems required students to set up and solve a linear equation to represent the number of tiles in the tile pattern in terms of its figure number. They were then tasked with finding the figure number that would result in each pattern having the same number of tiles having the same quantity of tiles in the same figure number (see Appendix). The final problem tested student ability to reason about linear patterns more abstractly than either of the previous problems. Whereas the second problem asked students to solve a system of equations and interpret the result in the context of the problem, the last problem gave a hypothetical situation in which students were asked to interpret what it would mean about the tile patterns if a certain result occurred (e.g., What would it mean about the tile patterns if both of the lines that represented them were parallel?). Both exams are provided in the appendix for reference.

After each exam, the instructor provided written feedback on the exam and scores it according to the 4-point rubric on a recording document separate from the exam. The scores were withheld from the student at this point, and instead, students were provided with only their exam with written feedback. For Exam 1, feedback was provided one week after the exam to allow for adequate time to write the feedback and to consider and analyze both individual and class-wide patterns of understanding. While feedback for Exam 2 was planned to be given within a week, the feedback was delayed until three weeks following the date of the exam due to the necessary and unexpected transition to online courses as a result of the COVID-19 pandemic.
The exams were designed to be short enough such that the feedback could be given in a reasonable amount of time and in consideration of Hattie and Timperley’s framework (2007).

Once students received their written feedback on each exam, they were given the Exam RAP assignment. Students completed the RAP for both Exam 1 and Exam 2. Following the completion of both exams and the RAP process, students considered brief survey questions sent through a Google Form about the RAP assignment in general. The questions on the survey were: (1) *What did you think about the Exam Reflection and Analysis Process?* Does anything about the experience stand out? (2) *For you personally, what were the successes of the Exam Reflection and Analysis Process in this Math 116/51 class?* What are your take-aways from that process that you may use in the future? (3) *From your perspective, how would you improve the Reflection and Analysis Process for exams to maximize the benefits?* (4) *What other comments or thoughts do you have that were not addressed in previous questions?*

**Data Analysis Procedures**

The overarching questions this study aimed to answer were: how students’ self-assessment scores compare to the instructor-assigned scores; how students’ rationales for their self-assessment from their first compared to their rationales for Exam 2, and what specifically were the characteristics of any shifts in students’ epistemological beliefs; and what students’ experience and response was to the process.

For the first research question, scores were entered into an Excel spreadsheet for Exam 1 from the paper submissions of the RAP. For the second exam, students entered their responses into a Google Form which automatically populated a spreadsheet with the students’ self-scores, a preferable data entry system. All data was then transferred to a single Excel spreadsheet. R Statistical Programming Software was then used to gather the summary statistics,
scatterplots, and paired line graphs for each concept on Exam 1 and Exam 2. The scatterplots were designed with the instructor scores on the x-axis and the student scores on the y-axis. Then the $x = y$ line was graphed to show that the scores above the line represented when the student scores were higher than the instructor’s and below it when the student scores were lower than the instructor scores.

A command called “jitter” was used on each type of graph to slightly shake up the data such that overlapping points were visible. Additionally, some transparency of the points was introduced such that any points still overlapping each other after the “jitter” will appear darker. Clusters around the same coordinate generally indicate a data point with the same instructor and student scores as each other. An example of the scatterplot for Exam 1 Concept 1 without the “jitter” command is shown at left side-by-side below with one that does use the “jitter” command at right. For the sake of visibility of the points, all scatterplots use the “jitter” command in the remainder of the analysis. A similar feature, “dodge,” was used for the paired line graphs for better visibility of the data that overlapped each other. An example of a graph with “jitter” and one without is shown in Figure 4 below.
For the paired line graphs, the x-axis shows the categorical information of the instructor score and the student score. The y-axis represents the 4-point rubric score. Along the vertical line representing the instructor’s and student’s scores, points were plotted to represent each score, and then a line segment connected the instructor and student score for each student. Two bars in the background of the instructor and student score columns represent their respective mean scores. The correlation coefficient was also used to determine how closely correlated the student scores were to the instructor scores.

As an explanatory qualitative piece, I then investigated the rationales of the students to provide insight to their scores as part of Research Questions 2 to determine the relationship between student rationales for their self-assessment scores from Exam 1 to Exam 2. I also looked for any evidence of shifts of epistemological beliefs. I started with students who had the largest difference between the instructor score and their score in the first exam. I checked to see if the difference between scores was smaller on the second exam and to what extent their rationale changed. I then looked for themes stemming from students who had large discrepancies in their
scores when compared to the instructor score to those with smaller discrepancies to determine whether or not students were using evidence-based rationales from the work they showed on their exam or whether they were using more compliance-based or authority-based rationales. The themes for the rationales were coded using an inductive coding approach based on the exploratory nature of the study, and the epistemological beliefs scale (Steiner, 2007) was used as a reference.

To address the third research question about the student experience and perception of the RAP, I analyzed the students’ responses on their surveys about their experience with the RAP. The first question, in particular, was intended to be open ended without any influence of desired responses so students would share their impression without consideration of whether it was a positive or negative view of the RAP. The second and third questions, respectively, are intended to have students specifically attend to what they felt were the positives and negatives of the experience. And the final question closed with an open-ended prompt for students to address anything that did not fit into the first three questions.

I started by sorting the responses for the first question into general themes of what students thought about the RAP starting three loose groupings of more favorable, less favorable, and neutral. Student responses that mentioned explicitly that they did not like the process or did not find it helpful were sorted into the less favorable category while those who seemed to find it useful in some way were sorted into the more favorable category. Ones that were less clear or decidedly neutral that nothing stood out in one direction or another were placed in the neutral category. The less favored group was then sorted into two groupings: one group with students that clearly stated that they did not like or enjoy the process unequivocally and the others that seemed to allow for the possibility that it could help learning but they personally felt it did not
“really” help or did not help “much.” Neutral groupings were further categorized into two groups, one in which nothing stood out and one in which students stated the process was “unique,” “alright,” or “hit or miss.” The responses in the neutral and less favorable categories were mutually exclusive – there was no overlap of response among the subcategories. Hence, it was possible to count the number of students who responded in each way.

The more favorable category had several themes emerge of what stood out to them the most about the process: the opportunity to deepen understanding, the value of revisiting errors and mistakes and getting the opportunity to grade themselves based on them, the opportunity to review past concepts and recognize growth or continued need for growth, and evidence that the instructor cares about learning. In the more favorable category, there was much more overlap among the responses in that they tended to fit into multiple of the themes. Therefore, an overall count of students who fell into the more favorable category was possible, but the counts corresponding to the subthemes were based on the number of responses that included the theme, and some students were counted multiple times. See Table 4 below for each category and subcategory.
Table 4

Categories and Subcategories of Survey Responses for Question 1

<table>
<thead>
<tr>
<th>More favorable</th>
<th>Neutral</th>
<th>Less favorable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responses not mutually exclusive – responses overlapped among subcategories</td>
<td>Responses mutually exclusive</td>
<td>Responses mutually exclusive</td>
</tr>
<tr>
<td>Created deeper understanding</td>
<td>Nothing stood out to the student</td>
<td>Unequivocally did not like the process</td>
</tr>
<tr>
<td>Value of revisiting errors and mistakes and grading self</td>
<td>Stated it was “unique,” “alright,” or reflections in general are “hit or miss”</td>
<td>Did not think it helped much and indicated a preference toward other learning experiences</td>
</tr>
<tr>
<td>Opportunity to review and keep working on concepts in need of improvement</td>
<td>Evidence to the student that instructor cares</td>
<td></td>
</tr>
</tbody>
</table>

To ensure consistent sorting, a colleague was given randomized responses for Question 1 of the survey to sort. There was a high inter-rater agreement (Saldaña, 2013). The colleague first sorted the responses into the two groups: the students that had nothing positive to say about the RAP and the students that did have positive things to say about the RAP. The sorting corresponded almost identically with the less favorable and more favorable categories, respectively, with the exception of the responses that stated some variation of “No [nothing stood out].” I considered those responses as neutral since the question asked the student, “Does anything about the experience [of the RAP] stand out?” However, the colleague suggested that these were more of a negative response at the “less favorable” end of the spectrum because the student had presumably done a lot of work for the assignment but indicated that nothing stood out to them about it. Upon pointing out the phrasing of the question again, the colleague agreed.
that my interpretation was reasonable but suggested the survey question itself could be improved for clarity.

The colleague then sorted the more favorable responses by a spectrum of categories. The first category was that of students that considered both positive and negative aspects of the RAP in their response about what they thought about the RAP and what stood out to them. The next three categories were ordered from least to greatest in terms of (1) the “narrower” focus on just the exam itself, (2) a slightly broader focus on the concepts that were measured by that exam, and (3) the most holistic perspectives that mentioned that the RAP process as helping them to not only understand the concepts, but also served as guidance for them moving forward to determine “what happens next” in their learning. Incidentally, these categories are reminiscent of Hattie’s framework for feedback (i.e., feed up, feed back, feed forward) but from the student lens rather than the instructor lens, though it may be worth noting that the colleague had not seen this framework prior to sorting. The colleague mentioned that an alternative way they thought of the categories of the lower-ordered thinking to the higher-ordered thinking. There was significant overlap in the themes that emerged from the initial sorting to the colleague’s sorting, and the counts of responses in each theme from the colleague’s agreed with the counts of responses in the subthemes from the more favorable category in the initial sorting.

The responses to the second and third question of the survey fell into an explanatory purpose that provided greater insight into their response for the first. (No one had anything additional to say in the fourth question.)
CHAPTER IV

RESULTS AND DISCUSSION

This chapter reports the findings of the research. The students were assigned a Reflection and Analysis Process (RAP) to be completed after each of two exams. Each exam measured understanding of three main concepts. For each concept, students self-scored their understanding of the concept based on the 4-point rubric and the written feedback, then provided a rationale for their self-score.

To summarize some important and relevant criteria in the generic 4-point rubric, a score of 3 or 4 means the student has demonstrated competency with respect to the concept and correlates to a B or A, respectively. A 4 indicates that the student work showed a “full grasp and use of the central mathematical ideas” with thoughts communicated clearly through written, symbolic, and visual means. A 3 indicates that the student work shows an “essential” grasp of the ideas and in large part communicates the thinking. A score of 2 indicates that a student has a “partial but limited” understanding and has “partially accomplished the purpose of the task.” The recorded work could be “incomplete, somewhat misdirected, or not clearly presented.” A score of 2 is the minimum score that represents a basic mathematical competency with that concept and is the minimum score that a student should earn to demonstrate a basic understanding. A 1, by contrast indicated that the student made “little progress toward accomplishing the task” and showed “minimal grasp of the central mathematical ideas” with their work “barely comprehensible.” Finally, a score of 0 is reserved for cases in which student work demonstrated virtually no grasp of the central mathematical ideas or cases when the recorded work was not at all comprehensible. When students’ work met the of multiple scores, a fractional score was used.
to indicate that the work showed evidence from both of the whole number scores around it. A Google Sheet or Excel spreadsheet with formulas was used to enter the 4-point rubric scores for each concept and convert them to the grade percentage that was entered in the gradebook. Students are not made aware of the score-to-percent conversion, but for the reader’s information, the grade percentage for the gradebook was calculated using the formula $y = 0.15x + 40$ where $x$ is the average of the 4-point rubric scores and $y$ is the final percentage. When concepts were weighted differently from each other, the expression, $0.15x_n + 40$, was used where $x_n$ represented the $n^{th}$ concept’s score and the entire expression was multiplied by the weight desired as a proportion of the total. The results were then summed to get a final grade percent. For example, say on an exam that Concept 1, 2, and 3, were weighted 50%, 30%, and 20%, respectively. The final grade percent was programmed in the spreadsheet to calculate as $0.5(0.15x_1 + 40) + 0.3(0.15x_2 + 40) + 0.2(0.15x_3 + 40)$.

After completion of the second exam, students completed a survey to describe their experience and perception of the RAP. The study leveraged both quantitative and qualitative analysis as part of a complex mixed methods research approach to analyze student interaction with the RAP. The research aimed to address how student self-scores compared to instructor scores from Exam 1 to Exam 2; how student rationales compared from Exam 1 to Exam 2 and what, if any, shifts occurred with regard to epistemological belief; and what the student perception and experience of the RAP was. A comparison of student self-scores and instructor scores are presented and discussed. Following the quantitative analysis, a qualitative comparison of themes observed between student rationales from Exam 1 to Exam 2 is discussed. Finally, common themes from the student survey regarding their perception and experience with the RAP are also reported.
Quantitative Results and Discussion

Exam 1

Exam 1 Concept 1

For Exam 1 Concept 1, the mean instructor score was 2.81 with a standard deviation of 1.21. The mean student score was 3.12 with a standard deviation of 0.81. Hence, the student scores were higher, on average, than the instructor score, and the students deviated much less from their average than the instructor. See Table 5 below for summary statistics.

<table>
<thead>
<tr>
<th>Concept: Determine whether or not a relation is a function using multiple representations</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructor Score</td>
<td>2.81</td>
<td>1.21</td>
</tr>
<tr>
<td>Student Score</td>
<td>3.12</td>
<td>.81</td>
</tr>
<tr>
<td>Difference (Instructor-Student)</td>
<td>-.30</td>
<td>.99</td>
</tr>
</tbody>
</table>

$r = .5697$

<table>
<thead>
<tr>
<th>Percent of Student Scores to Instructor Scores (rounded to nearest percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower by &gt; 0.5</td>
</tr>
<tr>
<td>Within ±0.5</td>
</tr>
<tr>
<td>Higher by &gt; 0.5</td>
</tr>
</tbody>
</table>

In the scatterplot (see Figure 5 below), the $x = y$ line indicates when the student score equaled the instructor score. In general, we can see that most of the students’ scores were higher than the instructor scores as indicated by more points being above the $x = y$ line in the scatterplot. As indicated by only one of the paired lines leading to a student score of 1, students consistently self-scored between 2 and 4 and avoided scores of 0 or 1. Only two students self-scored a 1, with no students self-scoring 0. Interestingly, one of the students who self-scored a 4
but earned an instructor of 0 provided a rationale that, “At first I rushed through this section and did not take my time. After my corrections I realized the difference between a relation or nonrelation.” It seems the student may have misinterpreted the instructions on the RAP assignment that stated that students should assess understanding on the original exam itself based on the feedback that was provided and the 4-point rubric. Alternatively, this student may have underestimated the nature or severity of the errors that were on the exam based on the written feedback. Recall that instructor scores are not made visible to the student until after they have completed the RAP with their self-scores and rationales.

![Graph showing Exam 1 Concept 1 Score Comparison Scatterplot and Line Graph](image)

*Figure 5. Exam 1 Concept 1 Score Comparison Scatterplot and Line Graph*

*Since the “jitter” command was used, there is slight variation from the points shown in the graphs and actual scores. See Chapter 3 for details.*

From these graphs, we can see that students were most likely to align with the instructor-assigned score when they earned between a 3 and 4. We can see this by looking at the
clusters around (3,3) and (4,4). We also see a cluster around (4,3) indicating an instructor score of 4 and a student score of 3. This suggests that these students may have been hesitant to score themselves a perfect score of 4. This was the case even when their student score and rationale suggested they were relatively confident about the proficiency they demonstrated for that concept with one student stating, “I got everything correct, no errors.” Three others allowed for the fact that they “could have been more detailed in [their] response,” and one said, “I could have explained more but for the most part I’m able to determine whether a relation is a function when presented with one.” In fact, in general, when students’ self-scores matched the instructor scores within 0.5 points but that instructor score was not a 4, they seemed to acknowledge that they knew the concept but struggled explaining it with statements such as, “…my justification was incorrect,” and, “It is easy to determine whether it is a function but to explain it, it is a little different.” While these students whose self-scores and instructor scores matched did acknowledge the need for improvement regarding justification, there was only one student who specified what additional justification was needed. The student who provided specifics earned a 3.8, just under a 4 to indicate that their work was almost completely correct and fully justified.

Overall, for Concept 1 of Exam 1, about 26% of the students assigned more than 0.5 points higher than their instructor score, 18% scored 0.5 points lower, and 56% of the students were within 0.5 points of the instructor score. With a correlation coefficient of about $r = 0.57$, there is a moderate positive correlation between the instructor and student score for Exam 1 Concept 1.
Exam 1 Concept 2

![Figure 6. Exam 1 Concept 2 Score Comparison Scatterplot and Line Graph](image)

Table 6

Exam 1 Concept 2 Summary Statistics ($n = 34$)

<table>
<thead>
<tr>
<th>Concept: Linear Functions through Multiple Representations</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructor Score</td>
<td>2.57</td>
<td>1.10</td>
</tr>
<tr>
<td>Student Score</td>
<td>2.84</td>
<td>.56</td>
</tr>
<tr>
<td>Difference (Instructor-Student)</td>
<td>-.27</td>
<td>.96</td>
</tr>
</tbody>
</table>

$r = .4639$

<table>
<thead>
<tr>
<th>Percent of Student Scores to Instructor Scores (rounded to nearest percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower by &gt; 0.5</td>
</tr>
<tr>
<td>Within ±0.5</td>
</tr>
<tr>
<td>Higher by &gt; 0.5</td>
</tr>
</tbody>
</table>
For Exam 1 Concept 2, we can see that almost all students self-scored 2 or 3 as indicated by the concentration of points on the \( y = 2 \) and \( y = 3 \) lines and a relatively low standard deviation of 0.56 (see Figure 6 and Table 6). We can also see this on the paired line graphs by the fact that there is a spread of instructor scores with most of the lines connecting to scores of 2 and 3. The instructor scores were much more evenly distributed as shown by the spread of points along the \( y = 2 \) and \( y = 3 \) lines and the standard deviation of 1.1. The mean scores of 2.57 and 2.84 were relatively close for the instructor and student scores, respectively, but the standard deviations illuminate the difference between how the students self-scored compared to how they were scored by the instructor.

One student, for example, who earned an instructor score of 0.5 but self-scored a 2 stated quite simply yet contrarily, “I understand it, but am still not clearly getting the task.” Another, who earned an instructor score of 1.5 but scored herself at a 3 stated, “After correcting my mistakes, linear functions and problems are easier for me to understand.” This justification is based on their perceived understanding of the concept while completing the RAP. This student is another example of someone who either misinterpreted the instructions or misinterpreted the nature of their errors. Another student, who earned an instructor score of 0, self-scored a 3 and stated, “The first half of 2a-c, I knew somewhat what to do, but I felt like there was an easier way to accomplish the problems.” This student, however, was only able to successfully draw a Figure 0 and Figure 4 (based on Figures 1 through 3) and did not successfully come up with any other representations of the linear relationship (e.g., table of input and output values, graph, equation).

Students whose self-scores matched the instructor for this concept seemed to have similar rationales as in Concept 1 regarding the fact that they needed to elaborate on their
explanation. However, similar to Concept 1, students did not elaborate on what specifically they needed to justify more. Some students for this problem acknowledged that they had the idea of the concept but made careless errors, such as this student who stated, “I have a good grasp of the material but failed to display it due to careless errors.” Another student stated, “I knew what I was doing but I had the wrong numbers. I had the right idea”. Both of these students earned scores of 3.

Overall, about 24% of the students self-scored more than 0.5 point higher than the instructor’s, 12% of the students self-scored more than 0.5 points lower, and 65% were within 0.5 point. With a correlation coefficient of about $r = .46$, there is a moderate positive relationship between the instructor scores and the student scores.

*Exam 1 Concept 3*

*Figure 7. Exam 1 Concept 3 Score Comparison Scatterplot and Line Graph*
For Exam 1 Concept 3, we can see that while the students’ scores are fairly distributed between scores of 1 and 3 with just two scores of 4 (see Figure 7), this concept had a much higher proportion of instructor scores of 0 than the previous two concepts. We can also see that the instructor-assigned scores other than 0 fall mostly on or between scores of 2 and 3 with the bulk of instructor scores falling between the vertical $x = 2$ and $x = 3$ lines. One student, who earned an instructor score of 0 but self-scored a 3 stated, “I think I deserve a 3 because I tried to fix what I did wrong and understand the idea of what was going on.” One student self-scored a 2 but earned an instructor score of 0 and stated as his rationale, “Failed to answer parts of the questions,” with no reference evidence of correctness or understanding.

Students whose self-scores matched with the instructor scores for this concept used the rubric criteria as evidence. One student whose self-score and instructor score were both 2 stated, “Partially accomplished task. Work was not clearly presented. Conceptual error.” Another student stated who self-scored a 2 and earned an instructor score of 2.5 stated, “I did the work, but I missed small details within my work that would have made a huge difference.” Others relied on the feelings they had while taking the exam, such as a student who also self-scored a 2 but earned an instructor score of 2.5, who stated, “Still a little confused, partially accomplished, though I do show little grasp.” A similar rationale from a student whose self-score and instructor score were a 2 and 2.5, respectively, stated, “I know how to solve them, but I get confused during some steps at times.”
Table 7

Exam 1 Concept 3 Summary Statistics \((n = 34)\)

*Concept: Solving linear equations and systems of equations*

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructor Score</td>
<td>1.49</td>
<td>1.22</td>
</tr>
<tr>
<td>Student Score</td>
<td>2.35</td>
<td>.84</td>
</tr>
<tr>
<td>Difference (Instructor-Student)</td>
<td>-.86</td>
<td>1.21</td>
</tr>
</tbody>
</table>

\(r = .3463\)

Percent of Student Scores to Instructor Scores (rounded to nearest percent)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower by &gt; 0.5</td>
<td>12</td>
</tr>
<tr>
<td>Within ±0.5</td>
<td>32</td>
</tr>
<tr>
<td>Higher by &gt; 0.5</td>
<td>56</td>
</tr>
</tbody>
</table>

Notably, about 56% of the students self-scored at least 0.5 point higher than the instructor’s score (see Table 7). About 12% self-scored at least 0.5 point lower, and about 32% self-scored within 0.5 point of the instructor score. With a correlation coefficient of about \(r = .35\) there is a weak positive correlation between the instructor scores and student scores. Very few of the students self-scored scores of 0 or 1, but the instructor-assigned scores had a high concentration of 0s (see the vertical \(x = 0\) line). Considering that students self-scored mostly 2s and 3s while the instructor scores had a high concentration of scores of 0, the weaker correlation for this concept falls in line with the students’ patterns of self-scores for the previous concepts.

**Exam 2**

**Exam 2 Concept 1**

For Exam 2 Concept 1, the mean scores of the instructor and the students were 2.33 and 2.55, respectively, with standard deviations of 0.84 and 0.66. See Table 8 below.
Table 8

Exam 2 Concept 1 Summary Statistics ($n = 27$)

**Concept: Graphing polynomial functions and key features**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructor Score</td>
<td>2.33</td>
<td>.84</td>
</tr>
<tr>
<td>Student Score</td>
<td>2.55</td>
<td>.66</td>
</tr>
<tr>
<td>Difference (Instructor-Student)</td>
<td>-.22</td>
<td>.76</td>
</tr>
</tbody>
</table>

$r = .4870$

<table>
<thead>
<tr>
<th>Percent of Student Scores to Instructor Scores (rounded to nearest percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower by $&gt; 0.5$</td>
</tr>
<tr>
<td>Within $\pm 0.5$</td>
</tr>
<tr>
<td>Higher by $&gt; 0.5$</td>
</tr>
</tbody>
</table>

We can see again by the concentration of student scores along the $y = 2$ and $y = 3$ lines (see Figure 8 below) that students were most likely to self-score between a 2 and a 3 with 89% of the participants that submitted their RAP assigning themselves a 2, 2.5, or 3. We also see the bulk of instructor scores falling on or between scores of 2 and 3. For this concept, all but eight of the data points are contained by the square constrained by the $y = 2$, $y = 3$, $x = 2$, and $x = 3$ lines. The one student who earned an instructor score of 4 and self-scored a 3 stated, “I was on the right track but there were some things that I forgot to include or got wrong.” This student did not refer to the mathematical evidence shown on the exam in his rationale for their self-score. Interestingly, they predicted that their final grade for the entire test would be about a 70% because they “studied but could have studied some things more.” (They earned a 97% on the exam overall once the rubric scores were converted to percentages.) Their rationale indicates a focus on feelings and behavior rather than the mathematical evidence on the exam.
Many students continued to conclude in their rationales that they understood it better at the time of submitting the reflection compared to when they took the test: “I feel I got this score because I understand it way more compared to the day of the test. After some YouTube videos, I understand it enough to give it a decent answer. I saw how I messed up in the beginning, how it was \((x + 3)(x + 0)(x - 2)\), but after some help, I concluded that \(-x(x + 3)(x - 2)\). That was the biggest stump but after that it got so much easier.” Unfortunately, this was not the error; the instructor feedback stated that he did successfully demonstrate that he understood how the factored form related to the zeros of the polynomial, the very issue he stated in his rationale was what the YouTube videos cleared up for him. He did not address any of the other feedback on that problem that stated the items he needed to correct, implying that perhaps he understood the concept much less than on the day of the test. This was common for students on concepts on which their instructor score was 0 or 1. It was explicitly stated in the RAP.
instructions that their scores should be based on the understanding shown on the exam itself. Many students claimed that they understood how to solve those types of problems (operations with polynomials) but simply could not do the problem on the exam or had what they thought was a very simple error.

The students whose self-scores matched the instructor’s score identically seemed to rely much more heavily on the mathematical evidence shown on the exam as compared to the previous exam. For example, a student whose self-scores and instructor scores were both 2 stated, “In my work I noticed mistakes that could have easily been fixed with my exam. I still need to review a bit more since this was a step back for me. I solved each equation over again and all I have to pay attention to every aspect that's happening in the equation. I rushed through the exam without paying attention to what exactly I have to do in order to answer the question.” This student included information about the behaviors that they would need to change in order to avoid the errors in the future, but also referred specifically to solving errors they made in their attempt. A student whose scores were both 1 stated, “I’m giving myself a one because although I found the two points on the x-axis, I began to second guess the graph because of the other point on (−1, −1) which made the roots, standard form, and factored form equations wrong.” This student correctly identified their first error and the consequent errors that stemmed from it.

Students whose scores were both 3 tended to be even more specific about which errors they saw. For example, this student whose scores were both 3 stated, “I believe this score because as stated in your points, a lot of my mistakes were creating confusing analysis of the problem along with poor diction when explaining. On the other hand, I showed understanding of polynomial, but quickly revised my points on multiplicity. Lastly (1f) dictating a constant, I had realized that it was not constant, but increasing an extremely small rate across the x-axis.”
For this concept, overall patterns indicate that students seemed to use the evidence shown on their exam. They used mathematical vocabulary in their rationales, some of which was paralleled in the feedback they received. As a reminder to the reader, Exam 2 feedback was sent by email with feedback in narrative form in the body of the email and exact copy of the work they submitted attached. No feedback was left on the exam itself.

**Exam 2 Concept 2**

For Exam 2 Concept 2, we can see that the instructor scores are bimodal with the most common scores being 0 or 4 (see Figure 9). The students however, continued the patterns indicated in the previous concepts by assigning themselves mostly scores between 2 and 3. One student who self-scored a 3 and earned an instructor score of 0 stated, “I fully understand how to solve these types of problems. When working on them I tend to follow my own pattern on how to solve them such as for 2b; the function on the side had to be constant while the function at the top was the one that changed when needed. But I do make small errors when I don't take my time to solve them.” Another student who also self-scored a 3 but earned an instructor score of 0 stated, “I feel like I did okay with 1b and distributing polynomials although I did struggle with 2b a lot.” Only two students’ self-scores were within 0.5 of the instructor score. One student did not provide a rationale. The other stated, “I did get 1b wrong, but it was due to a careless error I should’ve caught when doing the test and 2b I got right. I understood how to do both problems.”
While 37% of the students were assigned an instructor score of 0 or 0.5, not a single student self-scored a score of 0. Just 7% of students self-scored within 0.5 point of the instructor scores with 41% self-scoring at least 0.5 point lower and 52% scoring 0.5 higher. The correlation coefficient of about $r = .06$ indicates an extremely weak positive correlation between the instructor scores and the student scores, the lowest correlation of all concepts. See Table 9 below for Exam 2 Concept 2 summary statistics.

*Figure 9. Exam 2 Concept 2 Score Comparison Scatterplot and Line Graph*
### Exam 2 Concept 2

**Concept:** Operations with polynomials

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructor Score</td>
<td>2.15</td>
<td>1.77</td>
</tr>
<tr>
<td>Student Score</td>
<td>2.61</td>
<td>.79</td>
</tr>
<tr>
<td>Difference (Instructor-Student)</td>
<td>-.46</td>
<td>1.87</td>
</tr>
</tbody>
</table>

$r = .0565$

<table>
<thead>
<tr>
<th>Percent of Student Scores to Instructor Scores (rounded to nearest percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower by &gt; 0.5</td>
</tr>
<tr>
<td>Within ±0.5</td>
</tr>
<tr>
<td>Higher by &gt; 0.5</td>
</tr>
</tbody>
</table>

### Exam 2 Concept 3

For Exam 2 Concept 3, we see that almost all of the students’ self-assessments are either at or above the instructors’ scores as indicated by almost all values in the scatterplot on or above the $x = y$ line (see Figure 10). There is a concentration of students’ self-assessment scores concentrated on scores of 2 and 3. Upon closer inspection of their rationales, these students did not refer to the evidence shown on their exam when evaluating themselves. One student who self-scored a 3 but earned an instructor score of 0 relied on their corrected work from the RAP and their feelings of confidence. They stated, “I felt I improved the most in parts a-c. I went back and fixed everything that I had before and put the correct answer for each problem. It helped a lot with the time I had to find all roots and being able to put them into factored form. I will say there is room to still improve that is why I put the score of a 3. Overall, I felt confident in what I was writing.” Another student who self-scored a 3 but earned an
instructor score of 1 stated, “[I} understand how to use the rational root theorem and how to factor but the problem was hard to solve.”

Looking at the rationales for students whose self-scores did match the instructor score, the students still seemed to revert back to rationales that were more general or were based on their feelings while taking the exam or doing the RAP. One student stated, “I would give myself a 3 in this section because I was correct on the first two problems, but the last question is where I made a mistake.” They did not elaborate further as to what was correct or understood and what was not. Another comment was, “I gave myself a score of 2 because I still really struggled with this, but I attempted to figure it out and try again on it. I’m not sure if I did it correctly or gave the answer you were looking for but I’m really hoping I did this reflection correctly… I don’t know how I missed that, but I feel foolish.”

Figure 10. Exam 2 Concept 3 Score Comparison Scatterplot and Line Graph
The correlation of about $r = .64$ indicates a moderate positive linear relationship
between the instructor and student scores for Exam 2 Concept 3. See Table 10 for Exam 2
Concept 3 Summary Statistics.

Table 10

Exam 2 Concept 3 Summary Statistics ($n = 27$)

<table>
<thead>
<tr>
<th>Concept: Equations of polynomial functions (Rational Root Theorem, factoring, solving)</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructor Score</td>
<td>1.67</td>
<td>1.2</td>
</tr>
<tr>
<td>Student Score</td>
<td>2.44</td>
<td>.84</td>
</tr>
<tr>
<td>Difference (Instructor-Student)</td>
<td>-1</td>
<td>1</td>
</tr>
</tbody>
</table>

$r = .6431$

<table>
<thead>
<tr>
<th>Percent of Student Scores to Instructor Scores (rounded to nearest percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower by $&gt; 0.5$</td>
</tr>
<tr>
<td>Within $\pm 0.5$</td>
</tr>
<tr>
<td>Higher by $&gt; 0.5$</td>
</tr>
</tbody>
</table>

**Overall Themes of Quantitative Results and Discussion**

When the instructor scores were in the range of 2 to 4, there was a stronger
correlation between student scores and instructor scores. Student scores in general tended to
hover around mostly 2s and 3s. Many students whose self-scores were not well aligned to the
instructor scores relied on their feelings during the exam or while completing the RAP. These
tended to include more generic comments about the problem or parts of the problem being wrong
or right. Often, students referred to their level of confidence rather than specific evidence from
their work. Many of these students based their scores on their confidence about their “new”
understanding after working on the RAP rather than the understanding shown on the exam.
When self-scores did align closely with instructor scores, the student rationales tended to be more specific and they used much more mathematical evidence from the work they showed on their exam. These students often described errors using mathematical vocabulary and included the types of errors they made. Many included some general actions they could take for next steps for learning.

Students were very unlikely, as shown in the data, to assign themselves a score of 0 despite having very little work or having the instructor identify that they had deep misunderstandings and misconceptions. This was the case even when they acknowledged in their own rationale that they had minimal understanding of the concept.

**Qualitative Results and Discussion**

This qualitative results and discussion section will first report the findings and themes observed regarding the student rationales followed by the findings and themes of the student experience and perception survey.

**Shift of Rationales for Students’ Scores from Exam 1 to Exam 2**

Student rationales for their self-scores were used as both an explanatory piece with respect to gaining more insight into the students’ self-scores as well as an exploratory piece to determine what kind, if any, shifts occurred from Exam 1 to Exam 2 with respect to students’ epistemological beliefs. First, we will consider the explanatory piece as it relates to the self-scores. Of special interest are the students who had large discrepancies between their self-score and the instructor score.

One student in particular had large discrepancies on Exam 1 between their self-score and instructor scores. On Concept 1, they earned a score of 0 but self-scored a 4. They earned an instructor score of 1.5 compared to their self-score of 3 on Concept 2 and earned an instructor
score of 0 compared to a self-score of 3 on Concept 3. Their rationales indicated that they believed they deserved their self-scores because after completing the reflection, they understood the concept. Hence, the instructor feedback provided to them upon completion of the RAP was, “Thank you for including what you feel your current understanding is after revisions. Remember for next time that your rubric score should be based on the understanding shown on the exam itself.” While their inflation was not as high for Exam 2, perhaps in consideration of the guidance on Exam 1, this student still systematically self-scored one or more higher than that of the instructor scores. However, their rationales were much more realistic in Exam 2, albeit not specific with respect to describing the mathematical evidence. One statement on Exam 2 Concept 1 was, “I feel like I am somewhat starting to understand this concept, yet I am missing certain steps or getting a bit confused on some problems.” Though they self-scored a 3, their rationale was closely aligned with the criteria for a score of 2 which was the instructor score for that concept. Similarly, they accurately articulated all of the specific items that needed improvement for Concept 2. However, despite their correct articulation of the items that needed improvement, they then concluded that they understood the concepts well and had made minimal errors.

Furthermore, when comparing her Exam 2 to a re-take of Exam 2, this student did completely correct six of the twelve items that were indicated as needing correction, partially corrected two of them, and did not at all address four of the items. Her Exam 2 RAP was very high in quality in that she followed all instructions, and all her corrections and explanations of her errors were, in fact, correct and complete. With regard to how she would mathematically avoid the errors in the future, her responses could definitely have used more specificity, as she tended to respond with statements like, “I will practice the concept more next time.”
The sentiment of minimizing the severity of errors and minimizing a blatant lack of any solution or reasonable progress toward a solution is echoed across many students’ rationales. For example, one student on Exam 2 Concept 3, who earned an instructor score of 0 but self-scored a 3, stated, “I am a bit back and forth when it comes to these concepts. On the second part of the concept, I got more of a grasp on and an understanding compared to the first one, but that might've been the result of a misstep on one of my problems I was redoing. Plus, since the first one is dividing, I do feel like I deserve that score – I showed the correct standard form.” The systematic minimization of errors that were quite severe was a theme that emerged across students with similar rationales, especially when their instructor-assigned scores were lower than 2. These students typically accomplished this by attending very specifically to anything that was correct (e.g., “…I showed the correct standard form.”) and weighing it more heavily than any other aspect of the concept (e.g., “Plus, since the first one is dividing, I do feel like I do deserve that score [of 3]”). Interestingly though, the part of their rationales that described what they needed to work on did become more accurate and detailed when comparing Exam 1 to Exam 2. Students like the examples described above and who earned an instructor score of 0 were especially prone to fixating on minimal correct work and essentially ignoring the larger concept when assigning themselves a rubric score.

While several students were able to articulate their errors and state some general remedies to study or practice, these same students self-scored and had corresponding rationales that indicated they believed they needed (and had completed) minimal studying of the concept. They believed that due to the studying and feedback, they understood it better upon submitting the RAP assignments as compared to when they originally took the exam. Many students came to the conclusions that they truly had a grasp of the concept and had minimal explanation errors
or a simple mix-up of definitions even when the evidence on their exam showed no connection to the concept or when the “simple mix-up” they described was the crux of the problem or concept. This is exemplified by the student who self-scored a 4 on the concept but earned an instructor score of 0 – the student stated, “At first I rushed through this section and did not take my time.” In the RAP assignment, the concept was described, “Functions (#1): Determine whether a relation is or is not a function [through multiple representations]” (Note: the addition of the “through multiple representations” is to represent the norm in the class that was made clear to the students prior to the exam and was a standard for all problems). Hence, this student either did not understand or did not want to communicate that she had no understanding of what makes a relation a function. She recalled her feeling rather than using the evidence that she showed on the exam.

About 24% of students on Exam 1 made statements that they did not fully represent the work on paper that they had attempted or understood in their minds or claimed that they knew the mathematical concept but not what the question was asking. This could indicate that they knew their work was almost completely incorrect or incomprehensible. If that was the case that they knew their work was incorrect or incomprehensible though, they still were extremely hesitant to indicate that with the rubric score that represented that criteria. About 21% of students on Exam 1 and 20% of students on Exam 2 claimed that they “understand it way more [now] compared to the day of the test” and so they deserved their self-score. Students who made this claim on Exam 1 were provided instructor feedback such as, “…Remember [your self-scores] should be based on your work on the exam itself rather than your revisions,” but the proportion of students making such claims were roughly the same from one exam to the next. Their rationales became more detailed and representative of the criteria listed in the rubric by using
mathematical evidence, but when it came to their self-scores, these students did not use the evidence and rubric criteria to support their self-score. This type of rationale was very correlated with students who earned an instructor score of 0 or 1.

There are a few possibilities as to why this occurred. For example, it is possible that students assessed their understanding at the time of completing the RAP more accurately but did not want to communicate a lower rubric score to the instructor, perhaps leery that low self-assessments would be factored into their grade. This is further elucidated by the fact that students were not necessarily consistent about their inflations across their whole exam. This inflation result could have also occurred based simply on the probability of how many of the scores exist above the instructor score. That is, if we consider all the possible outcomes of scores as 0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, and 4.0 and that all student scores more than 0.5 higher than the instructors’ scores are considered inflated, students earning an instructor score of 0 could potentially choose seven out of nine possible choices that would be considered inflated while a student earning a score of 3.0 would only have one possibility out of the nine total. Of course, the previous statement implies an equal probability of each score’s outcome, but in reality, we see that students tend to select scores of or between 2 and 3 much more frequently than other scores.

Examining students’ inflation of their self-scores in Exam 2 Concept 2 for which the instructor scores were bimodal around scores of 0 or 4. The bimodal nature of the instructor scores was based on assessing whether the student could implement the procedure of multiplying polynomials based on the factors they found in Problem 1 or dividing out a polynomial from a polynomial in factored form in Problem 2. While the students’ level of accuracy when implementing the procedure was considered when assigning the instructor scores, students who
scored a 0 or 0.5 did not successfully implement any part of the procedure. A score of 0 does not indicate a simple error, yet the students still self-scored 2s and 3s.

Students did not seem as likely to inflate their self-assessment in such a way if they earned a higher instructor score between 2 and 4 (though it is acknowledged that if a student earned a 4, it is impossible for them to inflate their self-score). These students tended to use more specific evidence from their exam and were able to articulate the evidence using the mathematical vocabulary. For example, “I had a hard time trying to find the factored form of the equation, which I ended up guessing on which then caused me to get the follow up questions wrong. I know how to identify roots and the steps to find the standard form of the equation. Where I get stuck is reading a graph. On some graphs I am able to identify them while others I struggle on. This graph just happened to trick me. I didn't put my full understanding of graphs into play. If I did, I would've been able to see that the shape of where the graph crosses (0,0) is in the shape of an $x^3$ graph.” In contrast, one of this student’s rationales for a concept on the first exam was, “I was confused on some parts. I feel like if I had more time to figure out the problems, I probably would've been able to figure it out.” There did seem to be a shift of rationale from Exam 1 to Exam 2 for this student.

Another student showed a similar shift of epistemological beliefs from Exam 1 to Exam 2. In the first exam, an example of the student’s rationale for a concept on Exam 1 was, “Made a small mistake of not labeling my axis, so I got confused,” relying on how they felt during the exam to assess themselves. Another example on Exam 1 was, “Made some simple mistakes, forgot to answer part of the problem.” In the second exam, however, they stated for a concept, “In my original test, I believe I would have gotten a 2, because I needed to make corrections to four of [parts of the] problem. In 1a, I guessed the 'a' value. I would have used
vertex from to find the 'a' value, but for this problem it would have been hard, because you can’t see the exact values of the points in some parts of the graph. In my error analysis I realized that the function was decreasing, so I added a negative sign as my 'a' value…” The student continued with several more sentences of similar descriptions for the remaining parts of the problem.

Similar results occurred for students who earned instructor scores between 2 and 4 and whose self-scores were relatively closely aligned with the instructor scores. The students who were exceptionally detailed in their rationales for Exam 2 tended to be the students to self-score lower than the instructor score.

**Student Experience and Perception of the RAP**

Following their completion of the exam RAPs for both Exam 1 and Exam 2, students took a survey about their experience with the RAP. Twenty-seven of the original thirty-four participants submitted the survey. Student survey responses were critical as a part of the Participatory Design Research (PDR) to ascertain students’ perspectives about the design and involve the participants in determining future directions of the RAP.

In the first question of the survey, students discussed their personal opinions about what stood out most to them about the RAP assignment. Themes that started emerging were students’ personal opinions about liking or disliking the RAP, the extent to which the RAP helped or did not help (in the general sense), time consumption, development of mathematical understanding, or a mention of the uniqueness of the process or a preference for an alternative method of revisiting an assessment once it is completed. The responses were then sorted into categories from most favorable or most useful to least favorable or least useful.

Several themes emerged from the sorting of the responses. Students responses included the extent to which the RAP contributed to developing their understanding of the
content or refreshing their memory of the content, to what extent they found the RAP helpful, the amount of time it took to complete the RAP, or simply that the extent to which they liked the process. Students with more favorable opinions of the RAP provided a lot of detail and tended to fall into multiple categories while the students who felt more neutral or had less favorable feelings about the RAP were able to be coded into just one mutually exclusive category. A summary table (Table 11) can be found below.

Table 11
Percent of Each Type of Student Response from Survey Question 1

\[n = 27\]

<table>
<thead>
<tr>
<th>More favorable Categories below are not mutually exclusive between the total “more favorable” responses</th>
<th>Neutral Categories below were mutually exclusive for the “neutral” responses</th>
<th>Less favorable Categories below were mutually exclusive for the “less favorable” responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Created deeper understanding</td>
<td>Nothing stood out to the student</td>
<td>Unequivocally did not like the process</td>
</tr>
<tr>
<td>Value of revisiting errors and mistakes and grading self</td>
<td>Stated it was “unique,” “alright,” or reflections in general are “hit or miss”</td>
<td>Did not think it helped much and indicated a preference toward other learning experiences</td>
</tr>
<tr>
<td>Opportunity to review and keep working on concepts in need of improvement</td>
<td>15%</td>
<td>18%</td>
</tr>
<tr>
<td>Evidence to the student that instructor cares</td>
<td>4%</td>
<td>56%</td>
</tr>
</tbody>
</table>


Fifty-six percent of the students indicated they viewed the RAP favorably for various reasons. One student simply mentioned that, “It feels like you care about the students more on what we learn on the weekly basis.” On the follow-up questions, this student reflected that working collaboratively with other students was hugely influential in their understanding of the content. Twenty percent of the students indicated specifically that the RAP helped their mathematical understanding with one student commenting, “At first I was frustrated with the process because it was very time consuming to me, and I would get frustrated if I did not know how to correctly complete the analysis. However, I do see how it helps us out, and it does make a difference on our understanding of the concept.” Other students indicated that the attention to mistakes and errors specifically was the helpful part of the RAP. One student commented, “I find out the areas in which I lack in. It helped me understand where I went wrong, and I am able to go back and work on the problem again and actually understand what is going on. Then I know what area exactly I need to work on for any future references.” In total, 32% of the students made comments about the attention to the mistakes being the aspect of the RAP that stood out to them the most.

There were also several quite neutral responses (in terms of favorability) regarding what stood out to students the most about the Exam RAP. One student commented that the process was “unique” and that “no other teacher has put them through a process like this.” Another commented that the thing that stood out to them the most was, “Just it’s mostly a lot of writing, and reflections for math are really hit or miss for me.” That same student did remark in Question 2, however, “With reflections, though, I was able to take out a lot from this class. It forced me to work problems backwards and construct a better understanding,” suggesting that
perhaps this student did find the reflections valuable albeit a bit time consuming for the understanding that was derived for them – perhaps a bit of both “hit” and “miss” for that student.

Sixteen percent of the students brought up the time consumption aspect of the exam RAP as the thing that stood out to them the most. Investigating these students’ responses to Question 3 led to some potential remedies to address the disproportionate time consumption compared to the understanding gained completing the RAP. The student indicating that reflections in math are always hit or miss suggested adding some simpler problems to pinpoint whether or not students understood the basics. One student indicated that, while time consuming, they still would not change the RAP at all. Another student indicated that they simply would like to have more time to complete the RAP. It is worth noting here that students were allowed extensions without having to provide a reason; this fact was mentioned to the students in class but not necessarily advertised. The fourth student indicated that they found the copying of the original problem and work on a separate page for the Exam RAP redundant.

Five of the students (20%) found the RAP was not helpful for them or that they did not like or did not enjoy the process. Three of these five students indicated that the RAP was not helpful with one student stating simply that they “[did] not think it helped much.” Of the other two students, one suggested that “practice problems should be forced on us,” and the other said, “I don’t think it can work because I take a test and do pretty bad on it, then a couple weeks later I am asked to revise the test, and nothing is going to change because I was not told how to do it differently to get a correct answer.” Both of the students who stated they did not like or enjoy the process indicated that they craved more direct instruction. One of the two stated they thought it would be more helpful to get their test back corrected by the instructor and then go over it as a class. The other student was baffled as to how they could be expected to correct their exam or
develop deeper understanding if they already did not understand it when they took the test. This student also disliked the self-grading aspect stating that they did not think that the self-grading contributed to their understanding of the content and that it “added stress to something [they] already didn’t understand.” For these students, their survey responses indicate epistemological beliefs close to the “knowledge is passed down by teachers” end of the scale as opposed to the “knowledge is constructed and developed through learning experiences” end of the scale (see Table 1).

The second question of the survey aimed to determine specifically what, from the students’ perspectives, were the most successful aspects of the RAP and what they would use in the future. From the responses, it appears that some students interpreted this question to pertain to the class overall. This is not necessarily an issue since the RAP is intended to be integrated naturally into the class culture and to contribute to the development of students’ mathematical dispositions and epistemological beliefs. The Exam RAP likely did influence the students’ perception of the class overall as it was a significant component of the course, and significant amount of time was dedicated to it. Even students who seemed to think the question pertained to the class in general brought up the importance of reflecting on past work for a better understanding and using mistakes as a learning opportunity. Fifty-two percent of the students who completed the survey brought up in their responses that they will take from the class that reflecting on their work was one of the biggest contributors to their understanding of the concepts indicating that they may have held epistemological beliefs toward the “knowledge is evolving” end of the epistemological belief scale as opposed to the “knowledge is unchanging” end in the certainty of knowledge category (see Table 1). While epistemological beliefs were not measured prior to the Exam RAPs taking place and it would be impossible to compare whether
or not a shift occurred, one student said, “One of the main things that I am going to take away from this class is going back and reflecting on my work. Before taking this course, I hardly ever looked back at my work and saw where I went wrong but doing this has actually helped me along the way in understanding in where I can improve.” Three students simply stated that they would use a reflection process in their future courses to get a deeper understanding of the content.

There were a few comments that addressed more general aspects of the class. Two students brought up managing their time effectively, one student commented that it was evident to them that their instructor cared a lot about making sure everyone understood the content, and one student mentioned that they realized the importance of organizing their materials since they were required frequently to use the materials from lessons and activities in their reflections. One student said, “I tried way harder than I normally did in any of my other math classes and tried to understand the math topics. My takeaways are that I have to ask questions to try to help myself grasp the topics.” Incidentally, this student was one of the students who did not like the reflection process and wanted more instructor guidance when doing corrections, so it is unclear to what extent the RAP contributed to their stronger efforts in this College Algebra class by comparison to their other math classes.

The third question offered an opportunity for students to share any specific improvements they would suggest for the RAP assignment. Nine of the students (36%) wrote that they would make no changes, commented “not applicable,” or said they did not know what changes they would make. Five of these students explicitly stated that they would not change anything about the process because the instructions for the RAP assignment were very clear, the RAP was very helpful, or the RAP was already “perfect” in their opinion. Seven students stated
that they needed more specific lessons, more practice, or more scaffolding, such as buildup from simpler to more complex questions, on concepts they missed. One student commented that “the instructor should do it for us when we correct our mistakes,” indicating a very decisive epistemological belief that knowledge is handed down through teachers.

There were also some performance, grade, and score-oriented comments. One student stated, “I feel this doesn't help after a student has taken the test because its more work that they normally won’t accomplish on their own,” suggesting a belief that “knowledge is passed down from the teacher.” There is an implication of a finality of the test in the first part of the statement that “this doesn’t help after [emphasis added] a student has taken the test.” One student suggested that students should, “Just do check boxes from 1-4 [for the scores] instead of a complete reflection.” In contrast, another student stated, “I feel like [we should just be] able to reflect on how we feel rather than ranking. I feel like I ranked myself lower than what I probably should have at times because I felt like I wasn't doing my best and that I just didn't deserve the ranking that I actually got. Reflecting on where I went wrong and how I was able to go back and edit it made the bigger impact on me.”

Four students reflected more generally about themselves as students that they realized that they needed to review and study more or spend more time reflecting, that they should have written all of the thoughts that came to them while taking the exam, or that they needed to take the reflection process more seriously because it did help them understand the topics better.

In the space for additional comments, only one student had an additional comment, and it was a reiteration of what they had stated in the previous questions.
CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

This study examined one method for mathematics educators to provide feedback to students in a way that promoted students’ reflection, analysis, and recipience of narrative feedback on their exams and a Reflection and Analysis Process (RAP) assignment. Students in a supported College Algebra course at a large public university in Northern California completed the RAP assignment after receiving only narrative feedback on each of two midterm exams. Based on the feedback provided to them by the instructor, they self-scored their work based on a 4-point rubric by reflecting on and analyzing their correct, incorrect, and incomplete work to improve their understanding. Additionally, students had to correct or complete work when it was applicable. They were required to elaborate on any incorrect or incomplete work by including information about what type of errors they made using a reference guide that was provided to them and specify what they would do to avoid the error in the future.

The study sought to answer three overarching research questions:

1. How do students’ self-assessment scores and instructor-assigned scores compare from Exam 1 and Exam 2?

2. How do students’ rationales for their self-assessment scores compare from Exam 1 to Exam 2? Specifically, in terms of epistemological beliefs, what were the characteristics of any shifts that occurred from Exam 1 to Exam 2?

3. What is the student experience and perception of the RAP?

The study used a mixed methods design to analyze both quantitative and qualitative data using a Design-Based Research (DBR) approach – research by the practitioner for
practitioners’ use – and principles of Participatory Design Research (PDR) to involve the participants in further design improvements. For the quantitative data analysis, students’ self-scores scores were compared to the instructor-assigned scores for each concept on Exam 1 and 2. Qualitatively, student rationales for their self-scores were analyzed in two ways: (1) through the explanatory lens with respect to the student self-score to instructor-score comparison and (2) through an exploratory lens to determine common themes among student rationales with clues to their epistemological beliefs and to determine whether there was a shift in reasoning from Exam 1 to Exam 2. Student responses from a survey following the completion of the processes for both Exam 1 and Exam 2 were analyzed qualitatively to determine student perception and experience of the RAP and determine future potential improvements to the RAP.

**Research Question 1: Comparing Scores**

**Findings**

The quantitative results from comparing the student self-scores to the instructor-assigned scores showed that student scores correlated to the instructor-assigned scores to varying extents. The self-scores were not necessarily correlated to the instructor score based on their accuracy but rather how the instructor scores were statistically distributed for that concept. Students self-scored 2s and 3s most frequently and seemed reluctant to assign the highest scores of 4 or the lowest scores of 0 and 1. This central tendency bias is a well-documented result that occurs when participants of studies are asked to respond with a scalar measurement (Douven, 2018). As a result, the overall predictor of how well students approximated the instructor scores for a given concept seemed to rely mostly on how closely the instructor scores deviated from scores of 2 and 3 rather than how well students understood the concept or how well they understood their understanding of the concept.
This could have occurred for several reasons. The feedback may have been ineffective in getting them to reckon with their true understanding. They may have needed additional instruction or guidance about the learning process. Or, according to some of the literature, it could have had to do with the timing or form of feedback (Hattie, 2007). Lastly, students could have perceived potential benefit that self-scoring higher would positively influence their exam score. Considering that scores of 2 and 3 were often overinflations of the instructor score begs the question of how educators might get students to improve their self-scoring accuracy or honesty, though it was not a focus of analysis this particular study. Speculations include the aforementioned central tendency bias but may also have to do with students’ concern about how their self-scores would influence their final grade on the exam. There was a deliberate lack of transparency as to how students would be assigned a final score since the goal of the RAP is to deemphasize grades and refocus students on learning and revisiting the concepts. For the first RAP, students asked about how their self-scores impacted the instructor scores and if “doing the corrections” would improve their grade. In response, they were told, “While I have a separate scoring sheet for the instructor scores, a final grade is not assigned without the RAP assignment. There is no final grade to improve as of right now [before your RAP is submitted]. I use your thoughtful reflections and compare them with the scores I assigned according to the rubric to find patterns. Focus on creating as authentic of a reflection as you can.”

As a result of the lack of disclosure about how final grades were assigned, students likely had their own speculations as to how their self-scores could impact their exam scores. For example, two students after the first RAP, speculation that self-scores and instructor scores would be averaged or that the instructor would consider “bumping” their original score up
considering the self-scores and reflections. In either case, they believed it benefited them to self-score higher. It is also entirely possible that any overinflation was completely unintentional and had to do more with the cognitive demand of using mathematics as evidence. Assuming the students who earned lower scores had, in fact, lower understanding of the concepts (as opposed to other reasons students may earn a lower score such as testing anxiety), it is possible that they did not understand enough of the mathematics to grasp the severity of the limitations in their mathematical constructions as is well documented by the Dunning-Kruger effect (i.e., that there exists a cognitive bias to overestimate one’s abilities when they have a lower ability) (Kruger & Dunning, 1999; Dunning, 2011). Alternatively, the lack of experience with grading and scoring among students compared to the instructor’s level of experience could explain this discrepancy. These students frequently relied on feelings and behavior to justify their scores, both which require far less cognitive demand than the demand required to understand and make use of the mathematical evidence on the exam.

Considering the central tendency of the self-scores following the completion of the first RAP, instructor feedback was provided on the first RAP in an attempt to guide students to gauge their understanding more accurately and honestly. General feedback for students who measured their “new” understanding directed them to the RAP instructions which stated they needed to base their self-score on the evidence shown on the exam. This feedback was given to about 32% of the students on their Exam 1 RAP. It is unclear whether or not this successfully impacted the accuracy of student self-scores on Exam 2. There were students who made similar overinflations on the second exam RAP, but they tended to be the students that did not receive that feedback on the first RAP because their scores were more accurate (and their feedback on their Exam 1 RAP told them that their scores were fairly closely matched when that was the
case). It is possible that students would become more accurate as they completed more RAPs if data were analyzed for all four of the quarterly exams.

**Recommendations for the Practitioner**

To improve students’ self-appraisal and assessment literacy (Winstone et al., 2017) prior to the exam, instructors could task students to engage with a mock grading assignment before a RAP takes place in which students collaboratively sort examples of student work into each rubric category. One solution for disrupting participants’ central tendency bias is to rank items rather than using a scale for individual items. In the context of the RAP, this could involve students engaging in a ranking activity in which the students sort sample work in order of which evidence shows the most understanding to the least understanding. As part of the activity to promote students’ thinking about the learning process, students could determine what guidance to give each sample student. Alternatively, students could collaboratively create and justify examples of work that belong in each rubric score category prior to an exam for an assigned concept.

A post-exam assignment could take the form of the instructor de-personalizing solutions shown by students in the actual class (as opposed to outside examples of student work). For example, instead of providing individualized written feedback for each student, which proved quite time and labor intensive, the teacher could instead observe themes of misconceptions from the class. Using the themes, the teacher can create sample student work with the goal of showcasing common and valuable mistakes and errors. This is similar to a popular activity, “Favorite kNOw,” which is also known as “Favorite No,” wherein a teacher selects a wrong answer given by a classmate (or themselves) for the class to analyze. The
learning activity for the students could take place in small groups or as a whole class with frequent Think-Pair-Share opportunities.

The *Summarize, Explain, Redirect, Resubmit* (SE2R) method of giving feedback advocates that scores not be assigned at all (Barnes, 2015). However, when grades *must* be assigned, the student should be required to resubmit until mastery is demonstrated. In addition, Barnes (2015) recommends that students use specific evidence from their work in their self-grading process. Yet, as the results of the RAP suggests, students need much more guidance and training as to how they can accurately assess their work based on evidence rather than feelings or behaviors. This is a class structure that would likely require a major overhaul such as that described in Barnes’ *Assessment 3.0* book (2015). For example, in a math classroom, a portfolio assignment would likely take place over the entire course and would be the main, if not the only, graded component of the class; the course assignments and assessments would be informed by the robust development of a portfolio as the main course objective.

However, the goal of the RAP and this study was to work *within* the parameters that are typical of common course structures rather than a complete overhaul and to consider a reasonable adaptation of the principles that inform Barnes’ (2015) approach. Components of these principles were used in the development of the RAP, such as how instructors can still assign grades but make it clear to the students that the RAP is not an optional grade-improving test-correction assignment, but a required component of assessments, hence the delay of the grade until *after* students had reflected on their work. As such, adaptations or improvements of the RAP could consider incorporating Barnes’ principles around assessment to a greater extent without a major overhaul. For example, instructors could add a simple component that tasks students with an additional opportunity to reflect on their understanding and learning process.
before feedback is ever provided by having students score their understanding directly after the exam takes place. This would offer the teacher (and/or researchers) an opportunity to determine whether students continue their tendency to self-score 2 and 3 as well as a means of comparison to their self-score post-feedback.

For those practitioners who are more inclined to explore eliminating self-scoring altogether and integrate more of the gradeless philosophy into the design without quite as much a major course structure overhaul, they could follow Barnes’ recommendation of requiring that students resubmit their work until proficiency is demonstrated at a level 3 or 4 according to the rubric. A feasible method in the literature is the single-point rubric (Fluckiger, 2010), a rubric tool that has only one column of the criteria needed to demonstrate proficiency. On the right side of that column is space to describe the evidence that shows students met the standard as well as a column to be used to describe evidence that the student went beyond the level of proficiency. On the left-hand side is space for the educator to describe areas for the student to improve in order to meet proficiency. Educators will need to find a way to make this process work feasibly within time constraints, perhaps by including fewer but more meaningful learning activities. This adjustment may sufficiently redirect students away from focusing on grades; in particular, this could address the student who suggested that they should be permitted to provide a only a score without a rationale. As a key component of the RAP, eliminating the rationale piece would undermine the entire premise of the assessment system.

On some problems, it seemed students were not provided ample opportunity to demonstrate a full conceptual understanding of the mathematics assessed and, as a result, the instructor scores were bimodal. To mitigate concerns about students not being able to demonstrate full mathematical proficiency because of the design of the exam, it is recommended
that practitioners take care to design assessments with more open-middled tasks. Depending on the needs of the class, one such example prompt could be, "Completely analyze and describe the graphed polynomial in such a way that someone reading your work could understand and draw the polynomial precisely without seeing the graph."

More open-middled or open-ended problems can make the grading and feedback aspect more challenging for the teacher since the responses will not be as streamlined. Teachers may want to consider using some of the other recommendations as well to alleviate any potential added strain from giving these types of problems or consider giving just one problem for an exam if the concept is rich enough. However, more open problems are critical when measuring and developing all of the strands of mathematical proficiency, as opposed to only procedural fluency. In particular, practitioners may want to predict the kinds of student thinking and work that will result from a task prompt. The goal is for the instructor to develop an understanding of where the students are in their learning, and too narrow of a prompt will limit instructor’s ability to develop lessons that address misconceptions and limitations. Furthermore, authentic problems will make it less likely that students simply mimic what they think the teacher wants and therefore they are more likely to exhibit agency over their learning.

**Research Question 2: Comparing Rationale**

**Findings**

When students had lower instructor scores, their rationales tended to focus on how well they understood the concept following the completion of the RAP. This was despite the fact that it was made explicitly clear in the instructions and emphasized in class that their rationales needed to be based on what was shown on the original exam itself. The RAP is likely one of students’ first experiences with understanding and reflecting on their own learning process.
Hence, the learners may have been relying on a more superficial sense of their understanding: “Does the feedback make sense?” as opposed to “Does the mathematics make sense?” Learners are accustomed to having material presented by an instructor, and students are rarely made aware of most of the explicit pedagogical decisions behind the presentation of the material (Star & Strickland, 2007). Students’ inner dialogue is mostly focused on whether they understand what is being said or shown rather than understanding the learning process more deeply (Star & Strickland, 2007). When instructors prompt students to think about whether something makes sense, they are thinking more about if they understood what was just said rather than interpreting their learning process and assessing whether a concept made sense before and after presentation of the material or feedback. In order to effectively communicate understanding and learning as a part of their assessments, students must also be taught explicitly about the learning process as a part of the course (see Recommendations below).

This type of rationale was not as typical when compared to the rationales of students who earned higher instructor scores. There was a shift from Exam 1 to Exam 2 regarding the rationales: students became much more detailed about their errors and mistakes and tended to articulate them much better. Despite the shift in rationale, the findings from Research Question 1 suggest that their self-scores did not shift significantly. It does not appear, based on the results of this study, that students’ scores became more or less accurate based on the level of understanding they demonstrated on the exam. While there was progress, educators need to set their own expectations appropriately that such a cultural and cognitive shift takes time beyond that of a semester to develop, and educators need to provide additional support and learning experiences that develop students’ understanding their learning process (Star & Strickland, 2007).
Evidence of student belief in external epistemic authority manifested with comments in the RAPs such as, “At first I wasn’t sure what you [emphasis added] wanted me to do…” Instead of focusing on demonstrating mathematical proficiency, the student was very focused on trying to do the problem the way the teacher wanted. This is a result supported by Fyfe and Rittle-Johnson’s (2017) study that when an instructor or researcher provides feedback to the student, the student’s mindfulness about the task decrease as students become more reliant on the teacher rather than their own analytical problem-solving skills. This result aligns with the reflection from one student’s statement, “I saw how I messed up in the beginning, how it was \((x + 3)(x + 0)(x - 2)\), but after some help, I concluded that \(-x(x + 3)(x - 2)\).” This rationale addressed one minor error but did not successfully integrate the feedback into their larger construction of mathematical understanding. Hattie’s (2012) study supports this result: teachers perceived they provide an abundance and variety of feedback to students, but students in reality received very little of it. Clearly, the RAP is not necessarily an exception to this result, and additional improvements to improve the extent of the recipience by students need to be considered.

In an effort to improve students’ accuracy and use of evidence from their exams, Moore’s (n.d.) error analysis assignment was a component that was adapted and added to the RAP during one of the pilots. The implementation was motivated by student feedback about the RAP and instructor observations that students were very unsure how to describe their errors and potential remedies; students needed additional guidance. For example, prior to this implementation and the *Error Analysis Reference Guide* (see Appendix) that was adapted from Moore’s assignment for the RAP, students would simply show the correction of the work and state things along the lines of, “I needed to study more,” or, “I did not get it, but now I get it.”
The addition of the reference guide alone seemed insufficient to remedy the issue, though it did provide language that the students used in their RAPs.

Also, the inadvertent switch of feedback format in Exam 2 as an email message sent to the student with the exam attached rather than writing feedback directly on the exam itself seemed to provide some of the vocabulary that students used in their RAPs. This suggests that the format of feedback provided on Exam 2 may have been more effective in guiding students’ analysis alongside of the Error Analysis Reference Guide. Consider, for example, the student who stated on Exam 2 Concept 1, “I will practice the concept more next time.” While the specificity of how they could correct their error could be improved, this student did successfully correct six of the twelve errors in a re-take of Exam 2. This could be a piece of evidence that the RAP may have contributed to more understanding (though a cause-and-effect relationship cannot be determined), but more importantly, it is a strength of the RAP that provides a cue to the instructor that the student needs additional guidance to develop their metacognitive skills and setting specific actionable goals (Winston et al., 2017).

Moore (n.d.) states in a cover page for their error analysis assignment their observation that students frequently conclude that an error is “careless.” That observation was noticed in this study as well. A careless mistake is one that students can automatically correct upon their own review of the exam without having to reference any feedback or revisiting other resources. Upon examining their errors, however, the instructor can often find that the student has still made errors on their corrections, which was the case in this study as well. It is speculated that the similar issues observed in this could have contributed to students’ rubric mismeasurement of their understanding. If a student is making the same errors in their
corrections when they are not under the pressure of exam conditions and have received feedback on it, the meaning of a “careless error” as defined in the reference guide is clearly contradicted.

Moore’s (n.d.) error analysis assignment and cover page recommends that students visit a tutor as a required component of the assignment after they have completed corrections independently. The tutor’s role is as a subject matter expert to guide the students to more deeply consider errors and misunderstandings that are undermining their successful completion of the problems. This could be a component easily added to the RAP that has the potential to address inaccuracy of students’ self-scores. From the survey, 20% of the students indicating they craved more direct guidance (e.g., the student who stated they did not know how they could be expected to correct the errors when they did not know how to do it the first time and were not told how to do it differently), so the implementation of the tutor requirement before submission of the RAP could provide this direct guidance while still putting the onus on the student to seize epistemic authority. It also values epistemic heterogeneity through the individual support. Early on in this study, the possibility of improving students self-scoring calibration to the instructor score was explored but considered a delimitation as it would require too large of a scope of study. Following completion of the study, student self-score calibration and accuracy remains a topic of interest and a potential for future study. The use of a tutor could certainly lessen students’ resistance to assess themselves more accurately and honestly.

Recommendations for Practitioners

The educational tool of The Learning Pit® (Nottingham, 2007, 2010, 2017, 2020) is an excellent metaphor and guide for students to reflect upon the process of learning as one that requires them to step out of their comfort zone as learning can cause many feelings of discomfort. By design, teachers are expected to guide students through the Learning Pit by
challenging, engaging, and questioning them so they can work out how to get out of the pit rather than being “rescued” from the pit. In fact, based on the results of this study, I implemented the Learning Pit in subsequent classes to great success and buy-in from students. Practitioners who are interested in this tool may wish to explore the Challenging Learning website at www.challenginglearning.com/learning-pit for additional details.

Additionally, several of the recommendations that pertain to Research Question 1 will address many of the needs for improvement pertaining to Research Question 2.

**Research Question 3 – Student Perception and Experience**

**Findings**

The results of the student survey suggest that, in general, students perceived the RAP as being helpful in some way. Over half the students mentioned in their surveys that they realized the importance of reflecting and looking back over their work. A quarter of the students stated specifically that they believed it helped develop their mathematical understanding. From their perspective, student recommendations for improvement included: removal of the requirement to copy their previous work on separate pages in order to save time, only having to write the narrative rationale piece rather than self-scoring, or only having to self-score and not providing a rationale. This highlights another example pertaining to Research Question 2 with respect to how some students continued to focus on the scores and grades aspect of the exam. However, it bears repeating that deemphasizing grades and scores is a key component of the RAP and, as such, the rationale for the self-score is critical to combat less productive epistemological beliefs (Steiner, 2007).

One-fifth of the students stated they did not like the RAP process. One of the students who viewed the RAP less favorably in general made a comment that they valued how clear it
was that their instructor cared so much about them and their learning. Students who did not care for the RAP process elaborated that they preferred direct teacher instruction over reflecting and analyzing their own work, even with the individualized teacher guidance through feedback. These types of comments could indicate epistemological beliefs of external epistemic authority in that they expect the teacher to provide the knowledge, a belief that is considered less productive for deep learning (Steiner, 2007). On the other hand, students who mentioned wanting practice problems or “simpler problems first” could be indicating a desire for more experiential learning through scaffolded problems which is on the more-productive-for-learning end of the scale of epistemological beliefs (Steiner, 2007). For these students, it would have been useful to know this desire early on the course which was not possible due to the timing of the survey. This speaks to the importance of incorporating consistent feedback throughout a course as well as the PDR study design. Ideally, as a goal for the development of the most powerful constructions of mathematical understanding, we want to guide students toward being capable of experimenting and creating their own learning experiences (Confrey, 1990). Had a portfolio or journal format been implemented as a means to track progress with meaningful opportunity for student-instructor communication, many more learning opportunities could have manifested. The simple suggestion of resources for students to find their own practice problems (with or without solutions provided) would likely have been very effective for these students.

**Recommendations for the Practitioner**

Teachers should consider discussing the learning process and roles of grades more explicitly with their students near the start of the course and consistently throughout. This discussion should be student-driven, but ultimately, students should question the role of grades
as they are typically used and that grades do not necessarily represent in practice what we think they represent in theory.

Students brought up in their surveys that time consumption was a detractor from the overall benefits of the RAP. An example that students brought up was the requirement to copy their work from the exam onto their error analysis. The purpose of students copying down their original work is so they can annotate and correct the work as a part of their error analysis without changing the original exam, but in practice, students did not make use of it. A potential revision of the RAP assignment could make clearer that the purpose of copying the original work is so students can mark up their work and make annotations to the original work without changing the original copy of the exam. Alternatively, teachers can suggest that students may photocopy the original work to include in their Exam RAP to avoid the redundancy.

On the teacher side, giving individualized narrative student feedback is also very time intensive. There were a few minor inconsistencies observed between students upon closer analysis. It is recommended that practitioners take advantage of the increasingly available technological tools that can afford grader consistency and timely feedback in a more efficient manner. For example, software such as Gradescope allows educators to grade by problem easily and efficiently with one-click applications of previous feedback. If a teacher adjusts feedback or rubric scoring for an item of feedback on a problem, Gradescope provides the option for that change to apply to all of the problems to which the change applies. Such software and ones similar to it could easily resolve this issue.

**Future Directions**

Due to the qualitative emphasis of the research design, more research is needed to determine causation relationships such as whether and to what extent a process like the Exam
RAP affects student understanding of the mathematical concepts or the influence the RAP had on students’ epistemological beliefs about mathematics before and after engaging with the RAP for exams. Originally, a question of interest was how completing the RAP before receiving a score impacted students’ understanding of the content. However, due to research constraints and the data that could feasibly be collected within the parameters, study of this question had to be postponed to future studies when an experiment could be conducted with a control and treatment group. An experimental design could be constructed based on the results from this research with a pre-survey and post-survey to measure changes in epistemological beliefs. Additionally, since this study only analyzed Exams 1 and 2, more research is needed to determine if there would have been more of a shift as students completed more RAPs. It would also be interesting to do a longitudinal study to follow up with students who had completed the RAP throughout an entire course to see if they used skills or epistemological beliefs developed through the RAP in any other classes.

The RAP was designed with the intention that the problems selected for each exam addressed all five strands of mathematical proficiency while also considering the feasibility of giving feedback promptly. However, exam design was beyond the scope of this study. Future studies could consider the design of the exams more heavily in the research design with an emphasis on designing exams that elicit demonstration of robust mathematical proficiency from students.

Furthermore, although the design of the exams heavily considered the feasibility of giving feedback, feedback for Exam 2 was delayed longer than intended because of the mid-semester transition to online learning. Research shows that some delay in feedback is acceptable at the task-level (i.e., testing situations) when the task is likely to be more complex and the result
of a significant amount of learning (Hattie & Timperley, 2007), but it is likely that too much delay is not ideal for learning as the assessment and feedback should drive teaching – the longer the delay, the more diluted the impact of the assessment on the next teaching steps. The delay of feedback for Exam 2 (or timing of feedback, in general) may have affected the results of the study, and future research is needed in which the timing of feedback within the RAP is standardized.

It is unclear what caused a low level of recipience of the feedback for students such as the example of the student who corrected a minor error that was not an overall focus of the feedback. More research and analysis need to be done regarding this aspect of the feedback. Potential reasons this result may have occurred include the level of the feedback being inappropriate for the level of learning of the student at the time they took the exam (Hattie & Timperley, 2007; Hattie, 2012); a failure of the feedback to engage the student with more and deeper learning and reflection or perhaps not enough richness of the feedback to enhance the learning experience was offered (Robinson et al., 2015); or the timing of the feedback in terms of the student’s prior level of understanding was not appropriate (Fyfe et al., 2012; Fyfe & Rittle-Johnson, 2017; Hattie, 2012). It is clear that developing the quality and effectiveness of the feedback with respect to proactive recipience by the students is critical. Categorizations such as SAGE developed by Winstone et al. (2017) could be a useful tool in this regard.

For future studies, the Participatory Design Research (PDR) aspect of this study could be explored and implemented much more thoroughly to involve the students in the design and research to an even greater extent, such as eliciting feedback more frequently through surveys, conversations, and interviews. This would entail frequent and clear communication with students throughout the process, and researchers must make sure they are politically prepared to have the
important conversations surrounding grades and equity with students (Vakil et al, 2016). Booker and Goldman (2009) suggest four principles of PDR: sustained open dialogue about what counts as the phenomenon of interest, simultaneous positioning of each party as learner and authority in ways of knowing, cycles of collaborative data analysis and design that extend the dataset and (re)direct the work, and removal of individual and cultural deficit as an explanation for systemic phenomena. Including students to an even greater extent in future studies may encourage more authenticity and honesty from the students in their reflections as well as improving the RAP assignment or other reflection-based assignments. There are many ways that students could have been involved more deeply with this process such as through journaling throughout the process about their cognition and preparation before and after the exams which could make the final reflection (collected as the survey in this study) more meaningful, portfolios of mathematical work throughout the semester, or interviews to elucidate more detailed student thoughts and experiences about the process.

Summary

The goal of the RAP is to enhance learning by promoting students’ reflection and analysis of their own work. It aims to disrupt the notion of grades based on behavior and compliance and promote students to seize their own epistemic authority, or agency, over their learning. The RAP may serve to answer Steiner’s (2007) call for more academic experiences in a mathematics classroom that challenge students’ nonavailing beliefs about mathematics. It is a hope that the RAP can lower the intimidation-factor of educators who wish to explore a grade-deemphasized (or gradeless) classroom or simply increase the usefulness of exams, potentially resulting in larger scale change in education as more educators experience alternatives to current standard grading. Optimistically, it may even help influence perspectives and provide evidence
that educators assign grades quite arbitrarily despite significant effort to remain objective and in a way that obscures the grade’s true meaning. As a standard component of a class, educators can use tests as their introduction, a cautious foray, into a grade-deemphasized classroom. Students culturally tend to place significance on assessments. Exams are typically weighted more heavily in the final grade, so the option of not doing the assignment can have serious consequences. Yet, grades and scores are not available for the student when they engage in this process, so while they know they need to complete the assignment to earn a final score on the exam, hopefully, the grade is sufficiently removed from being the primary focus.

Educators should not regard the process of revising and adapting the RAP assignment as “complete.” These results provide some information about this specific population of students in the spring 2020 College Algebra class. All educators using such a process should adapt it for their population’s specific needs by soliciting feedback from their students through a combination of informal and formal means. Then they should revise the instructions to elicit the desired responses from students with a focus on eliminating parts that seemed contrived or lacking real learning merit for the students and adding parts that develop deeper reflection, analysis, and metacognition in students for their current and future learning endeavors. Educators can consider adapting the assignment for all of their classes, one class specifically, or for individual students as needed, or they can choose to develop the entire process over time by focusing on the assignment’s parts and building students up to high-quality, complete submissions.

At the crux of the RAP assignment is the flexibility to determine and implement what will work best for their group of students. When implementing the RAP, or variations on it, educators should prioritize student learning over any aspect of behavioral compliance. They
should also feel at liberty to eliminate or adjust any parts that are not serving their students’ needs and add parts as they see fit based on their population’s grade, maturity, mathematical development, and general academic development. The RAP was used for exams in this course, but it is a process that educators could adapt or use directly for any task or assessment.
REFERENCES


Douven, I., & Douven, I. (2018). A Bayesian perspective on Likert scales and central
https://doi.org/10.3758/s13423-017-1344-2

Doyle, W., Sanford, J., & Emmer, E. (1983). *Managing academic tasks in junior high school: 
Background, design and methodology* (Report No. 6185). Austin: University of
Texas, Research and Development Center for Teacher Education.


Ertmer, P., Newby, T., & Medsker, K. (2013). Behaviorism, cognitivism, constructivism:
Comparing critical features from an instructional design perspective. *Performance 
Improvement Quarterly, 26*(2), 43-71.

Education Faculty Publications. 5*. https://digitalcommons.unomaha.edu/tedfacpub/5

exploratory mathematics problem solving: Prior knowledge matters. *Journal of 
Educational Psychology, 104*(4), 1094-1108.

Fyfe, E., & Rittle-Johnson, B. (2016). The benefits of computer-generated feedback for 
mathematics problem solving. *Journal of Experimental Child Psychology, 147*, 140-
151.


125


127
APPENDIX
Exam 1

Instructions: For all problems, show all work and justify reasoning clearly using words, numbers, symbols, diagrams, or a combination. When applicable, please box your final answers to make them clear. Please check the box to show you read these instructions: ☐

1. Determine if the following relations are functions. Provide a justification for each of your responses.

<table>
<thead>
<tr>
<th>Function? (Yes or No)</th>
<th>Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Graph" /></td>
<td></td>
</tr>
<tr>
<td><img src="image2.png" alt="Table" /></td>
<td></td>
</tr>
<tr>
<td><img src="image3.png" alt="Diagram" /></td>
<td></td>
</tr>
</tbody>
</table>
2. Consider the tile pattern below:

a. Draw figure 0 and figure 4. Describe and draw a diagram of what Figure 50 and Figure \( n \) will look like, making it clear how you are grouping tiles and seeing the growth. Consider using color or shading to highlight the growth and different groupings.

b. Write a function, \( f(n) \), that represents the number of tiles in the \( n^{th} \) figure. Consider showcasing your grouping strategies in your function.
Problem 2 continued.

c. If you did not already, please also write your function in slope-intercept form. What does the slope and y-intercept mean in the context of this problem?

d. Graph the function on the grid at right making sure to show up to Figure 10. This is just a grid. Be sure to draw in your axes, and label and scale them. Please use a straight edge for any lines.

e. The function, $g(n) = n + 16$ models another tile pattern. On the same set of axes above, graph $g(n)$. Be sure to clearly label each function with $f(n)$ and $g(n)$, respectively. Consider using color for clarity. [see graph above]

f. Set the expressions that represent $f(n)$ and $g(n)$ equal to each other. (i) Solve the equation for $n$. Interpret what the significance of this value is in the context of this problem? (ii) Describe where you see this value on the graph and label it in the graph above - what does this part of the graph mean in the context of the problem?
g. A classmate, Hector, brings you a function for a third tile pattern, $h(n)$. You want to investigate how it relates to $f(n)$, so you set the expressions equal to each other and try to solve them as you did in part (f). But you keep getting weird results like $2 = 2, 7 = 7, 0 = 0$... Write a sentence or two that describes what this means about the two tile patterns. Feel free to use diagrams to support your explanation.

h. Finally, Jason brings you his function, $j(n)$. When you set the expressions equal to each other and try to solve them, your results are even weirder than before - you get results like $2 = 7, 0 = 4, -1 = 3$... Describe what this means about the two tile patterns. Again, you may use diagrams to support your explanation.
Exam 1 Reflection and Analysis Process (RAP) Assignment

Exam Revisions Assignment: Please read ALL instructions before beginning.

1. You will need an Exam Reflection and Self-Assessment Handout (passed out in class on colored paper) and your exam.

2. Carefully consider the feedback on each problem.

3. Work on your own first to see if you can identify the issues yourself first. Then, with one or multiple partners, revise your work by following the steps outlined below. Do not make ANY changes on your actual midterm.

**ON A SEPARATE SHEET OF PAPER, for each imperfect problem:**

1. Copy or summarize the key parts of the problem’s instructions.

2. Copy your original work, especially the incomplete parts or parts with errors.

3. Correct the error(s) and complete the rest of the problem correctly.

4. Describe completely and specifically what your error was (see error types on back of this page) and your original thought process. Describe your new “corrected” thoughts and how you will avoid this error on similar problems in the future. Be specific about that problem. This is not “I will work harder next time” and similar statements. This is the time to specifically address what you did not know before that now makes sense, and how you can use that to problem-solve other problems with on that topic. Use the table of error types to help you with this process. The goal is to learn from it and make progress.

   d. Based on the evidence shown on your exam, give yourself a rubric score based on the rubric criteria with a brief rationale for that score. The rationale is especially important if you think that your score will differ significantly from my assessment of your understanding or if your thoughts that were shown were, upon your reflection, unclear but present on your exam.

   Important: Focus on the level of understanding you demonstrated on the exam and be honest with yourself. Remember that the score does not correspond directly with a percentage (e.g., 2 out of 4 is not 50% - F). If you want, you may write an additional comment about your new level of understanding in your rationale.

DUE __________________________ at the beginning of class. You will submit in the following order from top to bottom: Your completed rubric, your original midterm, and your revisions.

Please know that this is a required part of the exam. You will not earn a grade for it until this assignment is completed.
### Exam 1 Rubric and Scoring Sheet for Students*

*Note: The instructor rubric and scoring sheet is the same as the students but without the “Rationale” column.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Rubric Score</th>
<th>Rationale for Rubric Score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Functions (#1)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Determine whether a relation is or is not a function</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Linear Functions (#2 a-e)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Model a linear problem situation with a function that showcases ways of thinking or grouping the pattern as well as a function that showcases the growth and starting value.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Create a complete graph with axes, scale, and appropriate labeling. Graph linear functions from a problem situation or from an equation.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Interpret in context of the problem the meaning of parts of the function and graph.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Solving Linear Equations and Systems of Linear Equations (#2 f-h)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Solve linear equations.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Solve systems of linear equations.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Interpret solutions to systems of equations in context.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 4-Point Rubric

**For Scoring a Single Problem or Task**

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Fully Accomplishes the Purpose of the Task</td>
</tr>
<tr>
<td></td>
<td>Student work shows full grasp and use of the central mathematical idea(s).</td>
</tr>
<tr>
<td></td>
<td>Recorded work communicates thinking clearly using some combination of written, symbolic, or visual means.</td>
</tr>
<tr>
<td>2</td>
<td>Substantially Accomplishes the Purpose of the Task</td>
</tr>
<tr>
<td></td>
<td>Student work shows essential grasp of the central mathematical idea(s).</td>
</tr>
<tr>
<td></td>
<td>Recorded work in large part communicates the thinking.</td>
</tr>
<tr>
<td>3</td>
<td>Partially Accomplishes the Purpose of the Task</td>
</tr>
<tr>
<td></td>
<td>Student work shows partial but limited grasp of the central mathematical idea(s).</td>
</tr>
<tr>
<td></td>
<td>Recorded work may be incomplete, somewhat misdirected, or not clearly presented.</td>
</tr>
<tr>
<td>1</td>
<td>Makes Little Progress Toward Accomplishing the Task</td>
</tr>
<tr>
<td></td>
<td>Shows little grasp of the central mathematical idea(s).</td>
</tr>
<tr>
<td></td>
<td>Recorded work is barely comprehensible.</td>
</tr>
<tr>
<td>0</td>
<td>Makes No Progress Toward Accomplishing the Task</td>
</tr>
<tr>
<td></td>
<td>Shows no grasp of the central mathematical idea(s).</td>
</tr>
<tr>
<td></td>
<td>Recorded work is not at all comprehensible.</td>
</tr>
</tbody>
</table>
Exam 2

Instructions: For all problems, show all work and justify reasoning clearly using words, numbers, symbols, diagrams, or a combination. When appropriate, please BOX final answers. Please “smiley face” the box to show you read these instructions: ☑️

1. Consider and analyze the polynomial graphed at right, attending to all items in the list below.

a. Equation in factored form.

b. YOUR equation in standard form.

c. Roots (including multiplicity)

d. Local and global extrema. For each value, indicate whether it is a maximum or minimum.

e. End behavior. Write in arrow notation and then interpret your arrow notation in words.

f. Intervals of increase and decrease. Use interval notation and then interpret each of your intervals in words OR label clearly on the graph.
2. Consider the following polynomial: \( P(x) = x^3 + x^2 - 7x - 7 \)

a. Use the Rational Root Theorem to find all possible rational roots.

b. Completely factor the polynomial and write the function in factored form.

c. List all roots. Leave exact (i.e., do not approximate).

3. Now that you have finished Exam 2, how do you think you did? In terms of a percentage, what is your best guess as to the final grade you think you will earn? Write a brief description of why.

\[ \_\_\_\_\_\_\% \text{ because...} \]
Exam 2 Reflection and Analysis Process (RAP) Assignment

Exam 2 Reflection and Error Analysis Assignment: Please read ALL instructions before beginning.

1. Carefully consider my comments and feedback for each problem.

2. Error Analysis: Work on your own first to see if you can identify the issues yourself first. Then, with one or multiple partners, revise your work by following the steps outlined below. Do not make ANY changes on the actual midterm as you will not have room.

ON A SEPARATE SHEET OF PAPER:

(1) Copy (or summarize the key parts of) the problem’s instructions.

(2) Copy your original work, especially the incomplete parts or parts with errors.

(3) Correct the error(s) and complete the rest of the problem correctly. Make sure to address all my comments and answer all my questions in your corrected work.

(4) Describe completely and specifically what your error was (see error types on back of this page) and your original thought process. Describe your new “corrected” thoughts and how you will avoid this error on similar problems in the future. Be specific about that problem. This is not “I will work harder next time” and similar statements. This is the time to specifically address, what you did not know before that now makes sense, and how you can use that to problem-solve other problems with on that topic. Use the Table of Error Types to help you with this process. The goal is to learn from it and make progress in your understanding.

(5) When you have completed this, you will need to use your phone’s scanner app to scan your analysis into a single, multi-page PDF file and be ready to upload it in the Google Form.

3. Reflection: This part was originally completed on the Exam Reflection and Self-Assessment Handout. Instead, you will be completing this part in the Google Form at THIS LINK (Note to reader: link removed; a screenshot is provided on the following page), which mirrors that handout. I will attach the original handout as well for reference and for the rubric you need.

Based on the evidence shown on your original exam, give yourself a rubric score based on the rubric criteria with a brief rationale for that score. (This part was completed on the handout for Exam 1 but will be completed in the Google Form this time.) The rationale is especially important if you think that your score will differ significantly from my assessment of your understanding. This would be most likely to happen if you notice that your work was unclear and hard to understand, but it was in fact present on your exam and may not have made sense to me.

Important: Focus on the level of understanding you demonstrated on the exam and be honest with yourself. Remember that the score does not correspond directly with a percentage (e.g., 2 out of 4 is not 50% - F). If you want, you may write an additional comment about your new level of understanding you developed from your reflection and error analysis, but the main focus is analyzing what you showed on the original exam.

TO SUBMIT: Go to the Google Form and complete the Reflection prompts. Remember that the form will directly mirror the Exam Reflection and Self-Assessment handout, so just type in the form what you would have written on the handout. Then upload your Error Analysis PDF file in the place provided in the form.

DUE ____________________. I chose this due date because I think you can reasonably complete this assignment in that time frame, and I want you to have the opportunity to ask questions this week. That being said, I know you have other assignments, and this is still a weird transitional time for your classes and lives. If you realize you won’t be able to complete this by the due date, please send me an email right away as soon as you realize you won’t have enough time. Do not wait till the last minute for this assignment. You may have questions that you need to get answered from me or classmates which is why I am giving you multiple days.

Remember that this is a required part of the exam. You will not earn a grade for it until this assignment is completed.
Exam 2 RAP Assignment (Screenshots of Google Form)

Exam 2 Reflection and Error Analysis – Spring 20

This form replaces what you would have submitted in class for your Exam 2 Reflection and Error Analysis. You will need the 'Exam 2 Revision and Error Analysis Assignment' that was sent through a Blackboard announcement. You will also need your Exam 2 and feedback that was sent to you in an email.

The name, username and photo associated with your Google account will be recorded when you upload files and submit this form. Not sure how to switch account? Switch account

* Required

**First Name**

Your answer

**Last Name**

Your answer

1. Rubric Score: Polynomial Functions – Graph Given *

Parts 1a-1f. Topics: Completely analyze the graph of a polynomial function including factored form and standard form of the equation, roots and multiplicity, extrema, end behavior, and intervals of increase and decrease.

Your answer

1. Rationale for Rubric Score *

Please provide a brief rationale for your rubric score for "Polynomial Functions - Graph Given" above.

Your answer
2. Rationale for Rubric Score *
Please provide a brief rationale for your rubric score for "Polynomial Functions - Equation Given" above.

Your answer

3. Rubric Score: Operations with Polynomials *
Parts 1b and 2b. Topics: Multiply polynomials using an area model or algebraically. Work should be clearly and correctly shown using standard conventions (1b). | Divide polynomials (or factor by grouping) (2b). | Add and subtract like terms of polynomials (1b and 2b, as appropriate)

Your answer

3. Rationale for Rubric Score *
Please provide a brief rationale for your rubric score for "Operations with Polynomials" above.

Your answer

Now that you have seen your feedback and completed an error analysis and reflection, how do you think you did? In terms of a percentage, what is your best guess as to the final grade you think you will earn? How does it compare to your response in #3 on the original exam? *

Your answer

I know it was a while back, but try to remember. Briefly describe your preparation for Exam 2 (use complete sentences please) *

(1) About how many hours did you spend preparing for Exam 2? (2) What resources (e.g., Quizzes, Lesson Notes and Handouts, Homework) did you use to prepare? (3) What would you do differently to better prepare for the next exam?

Your answer
Please upload your completed Error Analysis as a single, multi-page PDF file.

ENSURE THAT EVERYTHING LOOKS PROFESSIONAL BEFORE SUBMITTING (neat, in order, oriented correctly, clear contrast, pages are cropped with no background material, easily readable). This is the part where you corrected your errors and completed the error analysis using the Table of Error Types on separate paper. It should make up the bulk of your work.

A copy of your responses will be emailed to [blank].

Submit
### Exam 2 Rubric and Scoring Sheet for Students

<table>
<thead>
<tr>
<th>Topic</th>
<th>Rubric Score</th>
<th>Rationale for Rubric Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polynomial Functions - Graph (1a-1f)</td>
<td></td>
<td>Completely analyze the graph of a polynomial function including factored form and standard form of the equation, roots and multiplicity, extrema, end behavior, and intervals of increase and decrease.</td>
</tr>
<tr>
<td>Polynomial Functions – Equation (2a-2c)</td>
<td></td>
<td>Given the equation of a polynomial in standard form, use the Rational Root Theorem to determine possible rational roots (2a). Use any method to determine one of the roots and its corresponding factor, then use polynomial division (or factor by grouping) to completely factor the polynomial (2b). Finally, list all roots of the polynomial (2c).</td>
</tr>
</tbody>
</table>
| Operations with Polynomials (1b, 2b) |  | • Multiply polynomials using an area model or algebraically. Work should be clearly and correctly shown using standard conventions (1b).  
• Divide polynomials (or factor by grouping) (2b).  
• Add and subtract like terms of polynomials (1b and 2b, as appropriate) |

### 4-Point Rubric

**For Scoring a Single Problem or Task**

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
</table>
| 4     | **Fully Accomplishes the Purpose of the Task**  
Student work shows full grasp and use of the central mathematical idea(s).  
Recorded work communicates thinking clearly using some combination of written, symbolic, or visual means. |
| 3     | **Substantially Accomplishes the Purpose of the Task**  
Student work shows essential grasp of the central mathematical idea(s).  
Recorded work in large part communicates the thinking. |
| 2     | **Partially Accomplishes the Purpose of the Task**  
Student work shows partial but limited grasp of the central mathematical idea(s).  
Recorded work may be incomplete, somewhat misdirected, or not clearly presented. |
| 4     | **Makes Little Progress Toward Accomplishing the Task**  
Shows a little grasp of the central mathematical idea(s).  
Recorded work is barely comprehensible, but there is definitely some logic present. |
| 0     | **Makes Very Little (or No) Progress Toward Accomplishing the Task**  
Shows very little or no grasp of the central mathematical idea(s).  
Recorded work is not at all comprehensible. |
# Error Analysis Reference Guide for Exam 1 and Exam 2

<table>
<thead>
<tr>
<th>Types of Errors</th>
<th>Suggestion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Misread Direction Error</strong></td>
<td>These errors occur when you skip directions or misunderstand directions. To avoid this type of error, read all the directions, underlining key words.</td>
</tr>
<tr>
<td><strong>Careless Error</strong></td>
<td>Mistakes made which can be caught automatically upon reviewing the test. To avoid this type of error, watch carefully for simple mistakes as you work each problem. After finishing the exam, if you have time, review each problem step by step, checking that you have not made careless errors.</td>
</tr>
<tr>
<td><strong>Conceptual Error</strong></td>
<td>A problem-solving error you make when you do not understand the properties or principles required to work the problem. To avoid this type of error in the future, you must go back to your resources and learn why you missed the problems. You also may want to have an arsenal of general problem-solving tools. For example, when getting started with analyzing a function, what are multiple consistent ways to get started with those types of problems. Can you make a table? Is there a parent function that is transformed that allows you to quickly sketch the graph first? Can you draw a picture of a problem situation?</td>
</tr>
<tr>
<td><strong>Application Error</strong></td>
<td>Errors you make when you know the concept but cannot apply it to the problem. To avoid this type of error, you must learn to predict the type of application problems that will be on the test.</td>
</tr>
<tr>
<td><strong>Operational Error</strong></td>
<td>Errors you make when you knew the concepts and perhaps even how to apply, but you could not compute the numbers or do the algebra required for the problem. To reduce this type of error, look over past assessments and work and determine which types of algebra or computing of numbers tend to get in the way of successfully completing problems. For example, do fractions consistently create an obstacle? Build procedural fluency by first making sure you understand the underlying concepts. For example, do you know enough to make a reasonable estimate if needing to compute ( \frac{1}{2} + \frac{1}{3} )? If not, this is a conceptual error. Then, build procedural fluency by finding and completing practice problems.</td>
</tr>
<tr>
<td><strong>Principal Error</strong></td>
<td>Occurs when you are not able to answer the final answer which is caused by a previous error in the problem solving. This occurs in conjunction with another error, so you will need to address those first. Then you can come back to finish the problem.</td>
</tr>
</tbody>
</table>
| **Test-Taking Errors**              | Errors that you make because of the specific way you take tests, such as:  
(a) Missing more questions in the first or last third of the test.  
(b) Not completing a problem to its last step.  
(c) Changing test answers from the correct ones to incorrect ones.  
(d) Getting stuck on one problem and spending too much time.  
(e) Rushing through the easiest part of the test and making careless errors.  
(f) Miscopying an answer from your scratch work to the test  
(g) Leaving answers blank.  
If you find that you miss more questions in a certain part of the test consistently, use your remaining test time to review that part of the test first. To avoid this mistake, review the last step of a test problem first, before doing an in-depth test review. If you have this tendency, then write on your test “Don’t change answers!” Only change answers if you have double-checked and can prove to yourself that the changed answer is correct. Set a time limit for each problem before moving to the next problem. If you do this often, after finishing the test review the easy problems first, then review the harder problems. To avoid this, systematically compare your last problem step on scratch paper with the answer on the test. Work with your scratch paper placed on top of the test paper, not off to the side. Write down some information or try at least to do the first step. |
| **Study Errors**                    | Mistakes that occur when you study the wrong type of material or do not spend enough time studying pertinent material. To avoid these errors in the future, take some time to track down why the errors occurred so that you can study more effectively the next time. |

Adapted from Moore Analyzing Exam Errors and Making Corrections and Priyani and Ekawati (2018) Error Analysis on TIMMS.
Informed Consent Letter

Chico State University Informed Consent

You are being asked to participate in a research study. Before you give your consent, it is important that you read the following information to be sure you understand what you will be asked to do. The research will be conducted by me, Allison McConnell, your instructor and a graduate student at CSU, Chico. Since I am a graduate student completing my thesis, my research will be under the supervision of Dr. Brian Lindaman and Dr. Christine Herrera of the Department of Mathematics and Statistics at Chico State.

Purpose of the Research

This research study is designed to explore student ability to judge their performance on upcoming or recent exams. The study will explore how the process of predicting and reflecting on upcoming or recent exams correlates with the accuracy of judgment and performance on subsequent exams and student mathematical understanding in general. Furthermore, the study will explore student perspectives about the process and whether it influenced exam preparation and mathematical disposition.

Procedures

As a part of the normal exam process in this class, you will be asked to engage in a predictive and reflective process learning. First, you will predict how you think you will do on the exam one-to-three days before the exam takes place. Immediately upon completion of the exam, you will be asked to determine how you think you just did on the exam. After taking the exam, you will receive it back with written feedback and comments and will engage in a revision and reflection process. You will also be asked to briefly explain your preparation for each exam and, when applicable, how your exam preparation changed from the previous exam and what influenced it. You may be asked to respond to questions that explain your responses, perspectives, and preparation behaviors as they pertain to the exam process.

If you agree to participate in the study and you give your consent, your responses will be confidentially used in the data analysis for this study.

Potential Risks or Discomforts

There are no foreseeable risks as a result of this study. However, you may feel some discomfort responding to questions about your exam predictions and reflections. If you feel uncomfortable at any time in the study, you may remove yourself from the study immediately without consequences. You will still be asked to complete the exam prediction and reflection assignments as a part of the class, but your responses will not be used in the analysis of the data for this research. You may also decline to participate in any surveys or interviews pertaining to the study. Withdrawing from the study will not affect your standing as a student, your relationship with me as your instructor, or your grade in the course.

Potential Benefits of the Research

The potential benefits of the research will help me as your instructor as well as other mathematics instructors to determine which parts of the prediction and reflection process were beneficial in
your learning. Additionally, it will help me understand whether and to what extent engaging in the process impacted your mathematical confidence and judgement in progressing in mathematics.

Confidentiality and Data Storage

All information I obtain during the research will be kept confidential. All data will be kept secure on a password-protected computer and will be de-identified from any student names using a master coding list kept by Dr. Christine Herrera. Actual student names will not be used. Only supervising faculty, Dr. Herrera and Dr. Lindaman, and I will have access to the data.

Participation and Withdrawal

You must be at least 18 years old to participate in this study. Your participation in this research study is entirely voluntary. You may refuse to participate or stop participation at any time without penalty. Refusing to participate and/or withdrawing from the study will not affect your course standing as a student, your relationship with me as your instructor, or any of your course grades.

Contacts for Questions about the Research

If you have any questions about the research, you may contact me, Allison McConnell at amcconnell4@csuchico.edu, or you may contact my advisors. You may contact Dr. Brian Lindaman at blindaman@csuchico.edu or 530-898-4107. You may contact Dr. Christine Herrera at caherrera@csuchico.edu or 530-898-5492.

If you have any questions regarding your rights as a research participant, please contact the CSU, Chico Human Subjects in Research Committee at 530-898-3145 or irb@csuchico.edu.

Informed Consent:

I have read the information provided above. I am at least 18 years old, and I understand that by selecting “yes” below and signing this form, I agree to voluntarily take part in the research.

I, ____________________________________________(print first and last name), give my consent to participate in the study:

Yes ______  No ______

Signature: ________________________________ Date: _______________
IRB Approval Letter

February 03, 2020

Allison McConnell
[Personal address redacted]

Dear Allison McConnell:

Protocol # 30351

As the Chair of the Campus Institutional Review Board, I have determined that your research proposal entitled "Grade Delayed: The Impact of Using Only Written Feedback on Mathematics Assessments" is exempt from full committee review. This clearance allows you to proceed with your research.

I do ask that you notify our office should there be any further modifications to, or complications arising from or within, the study. In addition, should this project continue longer than the authorized date, you will need to apply for an extension from our office. When your data collection is complete, you will need to turn in the attached Post Data Collection Report for final approval. Students should be aware that failure to comply with any HSRS requirements will delay graduation. If you should have any questions regarding this clearance, please do not hesitate to contact me.

Sincerely,

Patrick S. Johnson, PhD.
Assistant Professor, Department of Psychology
Chair, HSRC, IACUC & IBC
MODC 106
CSU, Chico 95929-0234
530.898.3098