EFFECT OF AN ACTIVE NOTE RESTRUCTURING INTERVENTION ON DEVELOPMENTAL MATHEMATICS STUDENTS' TEST PERFORMANCE

____

A Dissertation

Presented to

The Faculty of the Department Developmental Education Administration

Sam Houston State University

____

In Partial Fulfillment

of the Requirements for the Degree of

Doctor of Education

____

by

Andrew Bloom

December, 2020

EFFECT OF AN ACTIVE NOTE RESTRUCTURING INTERVENTION ON DEVELOPMENTAL MATHEMATICS STUDENTS' TEST PERFORMANCE

by

Andrew Bloom

____

APPROVED:

Susan Troncoso Skidmore, PhD

Committee Director

Linda R. Zientek, PhD

Committee Member

Dustin L. Jones, PhD

Committee Member

Stacey Edmonson, EdD

Dean, College of Education

DEDICATION

I dedicate this dissertation to my family, without whose support I could not have undertaken such a large project. It is through the unwavering strength of my wife, children, and parents that I was able to have the time to complete the coursework and research to see this through to the end. For my wife, Terri, who endured my constant work in the early morning, late at night, and while we were on vacation, without your encouragement and support I would still just be talking about going back to school. For my children, Cailin and Nathan, thank you for putting up with my finishing one more sentence, sending one more email, and reading one more page. For my mother, Carol, who has stood by me for so many years and has offered support through every challenge, I appreciate that you are always there to listen and always seem to know when to step back and when I need a push.

To those I lost during the completion of this dissertation, I offer a special thank you. To my daughter, Samantha, who made me promise to finish. I only wish you had had the opportunity to realize your own academic dreams. Thank you for editing my early papers and pointing out my writing crutches. To my father, Alan, who always asked how things were and listened patiently when I prattled on about this study. Four decades of undying support and encouragement seems like a long time, and not nearly long enough.

ABSTRACT

Bloom, Andrew, *Effect of an active note restructuring intervention on developmental mathematics students' test performance*. Doctor of Education (Developmental Education Administration), December, 2020, Sam Houston State University, Huntsville, Texas.

The purpose of this study was to determine to what extent the completion of a note taking and review intervention impacted developmental mathematics students' performance on course assessments. Note taking is a challenging skill that requires time and effort to improve (Boch & Piolat, 2005). Many developmental mathematics students struggle to capture a complete record of course content in their class notes and then struggle to use their notes to help them prepare for exams (Cafarella, 2014). To complete the intervention, students were asked to create a resource they could use to improve their understanding of the material, their preparation for course exams, and their performance in future classes by reminding them of prerequisite content. Students were assigned to turn in the resource, which was called the journal, on the day of exams. To complete the journal, students were required to define the key terminology of the content being covered on the exam, write out general step-by-step methods for how to complete the problems, and illustrate their methodologies by completing representative problems. The goal of assigning the journal was to provide students with a list of key concepts for the class, which has been shown to improve the quality of class notes, and to improve understanding of their knowledge of class content.

Keywords: Developmental mathematics, Note taking, Note review, Note revision, Guided notes, Metacognition, ICAP theoretical framework, Active learning.

ACKNOWLEDGEMENTS

This dissertation could not have been completed without the help of the outstanding faculty of Sam Houston State University. I am especially thankful to my committee chair, Dr. Susan Skidmore, who guided me through this process. Thank you for patiently responding to constant questions and for your guidance in developing this project from an idea to completion. I am particularly thankful of the hours you spent with me over Zoom in the past few months as I tried to navigate an interesting data set and develop a strategy for analysis.

To Dr. Linda Zientek and Dr. Dusty Jones, who served on my dissertation committee, thank you for your feedback and questions. I appreciate the time and energy required to read the early versions and provide timely suggestions for fine tuning the dissertation. The final product is better because of your thoughts.

The decision to return to school was not an easy one. Since beginning in the program I have never second-guessed the decision because of the amazing faculty I have had the privilege to work with. Each of my instructors provided insights, writing help, and suggestions for refining the ideas underlying this project. I have grown as a writer and scholar as a result of your guidance.

To the members of Cohort 6 I offer a special thank you. With your support I was able to push through difficult personal circumstances. I learned a great deal from each of you and have missed the collaborations that were such a constant over the last four years. To the members of Cohort 7, who never questioned why I was with them for one semester, thank you for welcoming me with open arms and making me feel like a part of your academic family.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

CHAPTER I

Introduction

Lecture remains the predominant teaching method used in higher education today (Badger et al., 2001; Bonner & Holliday, 2006; Bui et al., 2012; Mesa et al., 2013). With the prevalence of lecture, the ability to take effective notes during class becomes an essential skill for college students to possess (Al-Musalli, 2015). Researchers have routinely shown that students struggle to take complete notes, commonly capturing less than 40% of key lecture concepts during class (Bonner & Holliday, 2006; Chen, 2013; Luo et al., 2016; Williams & Eggert, 2002). The failure to capture a majority of key content means that students have a less effective tool to review prior to exams, which impacts student performance (Williams & Eggert, 2002).

There are ways instructors can help students to better utilize their class notes to improve their performance. Note taking is a skill requiring time and focused effort to improve (Boch & Piolat, 2005; Williams & Eggert, 2002). Most instructors do not teach note taking because they lack both the time and the background to teach students how to take more effective class notes (Al-Musalli, 2015). Instructors can help students improve the quality of their class notes by providing cues during lecture when concepts are important and should be included in notes. These cues can be verbal (Titsworth & Kiewra, 2004) or visual (Austin et al., 2004) and have been demonstrated to positively impact the quality of student class notes. Another intervention shown to improve class notes is to provide students with breaks during which they can collaborate with other students to compare the content of their class notes (Luo et al., 2016).

In addition to helping students improve the quality of their class notes, instructors can generate class notes for students to improve the proportion of key lecture content captured (Boch & Piolat, 2005; Cardetti et al., 2010; Kiewra et al., 2018; Williams & Eggert, 2002). The notes created and provided to students by their instructors can be of several forms. Two of the most common are partial guided notes, in which instructors eliminate key words or ideas for students to fill in during class (Austin et al., 2004; Cardetti et al., 2010; Cornelius & Owen-DeSchryver, 2008; Williams et al., 2012), and complete notes (Grabe, 2004; Raver & Maydosz, 2010; Vandehey et al., 2005), in which instructors provide students with a complete copy of the lecture notes. Both forms of notes have strengths and weaknesses, but one primary concern with the provision of complete notes for students is the passive learning this encourages (Boch & Piolat, 2005). When students are given a complete record of course content that was generated by someone else they are not required to actively engage with the course content and learning is stifled.

Because many students do not recognize their class notes are incomplete and may contain incorrect information, students are unaware of the need to review and revise their class notes after class ends (Boch & Piolat, 2005; Cafarella, 2014). For those students who use their class notes to review for exams, most choose to reread (King, 1992) or to recopy their incomplete class notes (Bjork et al., 2013). Many researchers have demonstrated that more active review strategies lead to improved performance on assessments (Bohay et al., 2011; Cohen et al., 2013; King, 1992; Kiewra, 2016; Luo et al., 2016).

Developmental mathematics students often struggle because of a lack of awareness of the concepts they struggle to understand (Bol et al., 2015; Eades & Moore, 2007; Nordell, 2009). Metacognition, which is the ability to recognize what is known and what is unknown, is an essential element of higher education (Nordell, 2009; Schneider & Artelt, 2010). Students need to know which concepts they understand and which they need to spend more time learning (Schneider & Artelt, 2010). Many of the concepts students learn in developmental mathematics depend on the proper application of a methodology to answer a question (Eades & Moore, 2007; Schneider & Artelt, 2010). Underprepared students cannot properly prepare for exams or retain concepts needed in subsequent courses, when they struggle to identify which concepts they do not understand (Eades & Moore, 2007). Improving metacognition for students is a valuable component in improving their academic success.

Theoretical Framework

Active engagement has been shown to improve student learning, but learning activities have historically been classified as either passive or active (Chi, 2009). In an attempt to differentiate between different types of active learning, Chi (2009) categorized active learning engagements as active, constructive, or interactive based on the overt actions of the learner. The ICAP theoretical framework was used to rank the effectiveness of learning engagements based on the classification of the learning and established that as learning moves from passive to active to constructive to interactive, the effectiveness of the learning increases and, with it, student performance (Chi & Wylie, 2014). Several researchers were able to confirm the conclusions of the ICAP framework in both laboratory and classroom settings (Menekse et al., 2013; Wiggins et al., 2017).

Note taking, which is the act of recording a summary of an activity, is an essential skill in higher education. Beyond the act of taking notes, the process of reviewing and revising notes to ensure the record of class content is complete and correct is often overlooked by students, but is an important component of retention (Al-Musalli, 2015; Boch & Piolat, 2005, Bohay et al., 2011). The teaching of note taking is not common in college, as many instructors find the skill to be outside their area of expertise, but there are several teaching strategies instructors can use to improve the quality of their students' class notes. Active interventions that encourage students to review and revise their notes in preparation of exams may provide a mechanism for improvements in performance and a framework for studying students can use in the future.

Purpose

The purpose of this quasi-experimental study was to investigate the relationship between the completion of a note restructuring intervention and performance on course exams. The study was conducted with students in various levels of developmental mathematics courses taught at a northern California community college. The intervention utilized aspects of active and constructive learning engagement, as defined by Chi and Wylie (2014) in the ICAP framework. The independent variable was defined as scores on a note review and restructuring intervention completed by students and turned in for grading on the day of the exam. The dependent variable was defined as performance on course exams covering the course content students examined in the intervention. Participation in the intervention was incentivized through the assignment of points which contributed to students' grades in the course. Students were free to complete the note restructuring activity to the degree they chose, including choosing not to complete it.

Educational Significance of the Study

This study will contribute to the body of developmental mathematics scholarship through the exploration of an intervention that improves the quality of students' class notes and encourages students to review and edit their notes prior to exams. It is the hope that the thorough explanation and examination of the intervention will provide a tool for developmental mathematics instructors to teach students how to better utilize their notes when studying. Because of the sequential nature of most developmental mathematics courses, the intervention will have both immediate and future value for students in that they are creating a resource for improving learning in the current class, as well as a resource containing information needed to successfully prepare for future classes. This resource can provide a framework for students to use in future classes. Because it is written in their own words and contains their own work, it is the hope that students will refer back to the information when questions arise in future classes.

Research Questions

The following research questions were addressed in this study:

1. To what extent does a student's performance on each journal intervention impact their performance on the corresponding exam?
2. To what extent is the pattern of student performance on the journal intervention associated with exam performance?

Definition of Terms

The following terms are defined as they were used in this study:

Active learning. This term is defined "classroom practices that engage students in activities, such as reading, writing, discussion, or problem solving, that promote higher-order thinking" (Conference Board of the Mathematical Sciences, 2016, para. 1).

ICAP theoretical framework. This theoretical framework was proposed by Chi (2009) as a methodology for differentiating active learning into three subgroups, (a) active, (b) constructive, and (c) interactive.

Active engagements. These learning activities are identified by focused movement or behavior undertaken in the learning process.

Constructive engagements. This term is used to define learning activities in which the learner generates a new output extending the lesson.

Interactive engagements. This term refers to learning engagements in which two or more students are engaged in collaborative learning, and all participants are engaged at the constructive level.

Guided (partial) notes. This term is used to describe instructor-generated notes with information intentionally omitted so students can fill in the missing information during class (Austin et al., 2004).

Metacognition. This term is defined as "people's knowledge of their own information-processing skills, as well as knowledge about the nature of cognitive tasks, and of strategies for coping with such tasks" (Schneider & Artelt, 2010, p. 149).

Limitations and Assumptions

There were limitations and assumptions that should be considered when reading this study. Students were eligible to register in one of the three courses as a result of passing the prerequisite course or based on their performance on a standardized placement test. Researchers have indicated that many students placed into the developmental mathematics sequence based on their performance on a standardized placement test may have been under placed, and could have been successful in college level courses (Rutschow & Mayer, 2018). It would be reasonable to assume that some of the participants in this study, particularly those placed one level below college transfer-level classes, were under placed and were successful because they were in a lower-class level than was necessary.

Student participants in this study were selected because they enrolled in sections taught by the author. The choice of a quasi-experimental design using a convenience sample creates some potential problems with both internal and external validity. There are several threats to internal validity, which refers to the confidence with which a researcher can state that changes to the dependent variable were due to the independent variable (Shadish et al., 2002). One potential threat is selection bias (Fraenkel et al., 2015). The author was known for assigning the intervention, which was touted by some members of the counseling department as helpful for struggling students. It is possible some students specifically registered for the instructor’s courses because of their perceptions of the value of the intervention. Also of concern is the potential mortality threat, which refers to the loss of participants during the study (Fraenkel et al., 2015; Wilkinson et al., 1999). Several students in each course withdrew from the course prior to the end of the semester. Although some students may have been succeeding in the course at the time of their withdrawal, the majority of students who withdrew from any course were less successful. Grades were recorded for students who took at least one exam. Students who withdrew from the course could be identified in the data by the lack of a journal and exam score for at least one assessment opportunity at the end of the semester. Finally, there is a potential implementation threat (Fraenkel et al., 2015). An implementation threat occurs when a person's positive experiences with an intervention account for the improvements in performance rather than the intervention itself. Several students registered for multiple sequential courses with the instructor and may have done so because of their comfort with the instructor or their previous experiences with completing the journal.

In addition to threats to internal validity, care needs to be taken to protect against threats to external validity, which refers to the ability of the research to be generalized to appropriate populations (Shadish et al., 2002). Using a convenience sample presents multiple potential problems. There is a possible increase in bias and there exists a possibility that results may not generalize to a larger population (Fraenkel et al., 2015). The generalizability problem can be mitigated through a thorough investigation of the characteristics of the convenience sample, which can help to better illustrate how the sample reflects the characteristics of the population (Jager et al., 2017; Wilkinson et al., 1999). In addition to concerns about the generalizability of the study based on the characteristics of the participants, there are potential generalizability issues because of the author (Fraenkel et al., 2015). The author had taught developmental mathematics students for 10 years prior to the start of data collection for this study. As a result, the author had experimented with different ways of gaining student participation and acceptance of the intervention. Additionally, lessons were taught with a focus on how students would use the information from class in constructing their journals and reference to the journals were made frequently during class.

There are situations in which a random sample cannot be obtained, which was the case in this study. In these cases, the threat to external validity can be addressed through the use of replication, which is a process in which a study is repeated using different participants in different conditions (Fraenkel et al., 2015). If the results are consistent across the repeated attempts, the case for generalizability is strengthened. This study was conducted over multiple semesters and with developmental mathematics students taking classes that were one, two, three, or four levels below transfer-level courses. The study was repeated at least twice for each level of class.

In addition to the limitations in the study, there were some assumptions made. It was assumed by the author that all enrolled students would benefit from a note taking and note review intervention. The journal intervention had historically been assigned in transfer-level classes as well. Further, it was assumed that because the journal was being graded and the grade contributed to overall course grades, students would be incentivized to complete the journal. Finally, because there was only one instructor for all courses and the assessment rubrics were consistent across all sections of the same class, it was assumed that grading practices were reliable and that equal scores given at different times and in different classes represented similar quality of the assessments.

Organization of the Dissertation

This dissertation is organized in five chapters. Chapter 1 represents an introduction to the study, and includes an introduction, theoretical framework, a statement of the purpose, educational significance of the study, research questions, a definition of terms, and a summary of the limitations and assumptions made in the study. Chapter 2 is an extensive review of the relevant research pertaining to the problem. The method of the study is detailed in Chapter 3, which includes an introduction, a statement of the purpose of the study, the research questions, the research design, an explanation of the participants, the data collection, and the analytical strategy. Chapter 4 contains the results of the data analysis for each of the two research questions. The fifth chapter contains a discussion of the findings as well as the limitations and conclusions for the study.

CHAPTER II:

Review of Literature

Although many students understand the need to take notes in class, most do not do so effectively and do not understand how to properly use their class notes to help themselves better learn and understand the material (Boch & Piolat, 2005; Bonner & Holliday, 2006; Castelló & Monereo, 2005; Williams & Eggert, 2002). For developmental students, who often lack the study skills necessary for college success (Cafarella, 2014), the creation of a useful tool for recalling the details of a lesson and preparing for exams is particularly important (Eades & Moore, 2007). This is especially true for developmental mathematics students, who are learning concepts that will build in complexity within one class and in sequential classes in the developmental sequence (Cafarella, 2014; Cardetti et al., 2010; Eades & Moore, 2007).

Background of the Study

Much of the research addressing note review focuses on students taking notes in response to a controlled lecture, which is typically videotaped or recorded, and then being instructed to review for an immediate or a delayed test (Bohay et al., 2011; Bonner & Holiday, 2006; Luo et al., 2016; Williams & Eggert, 2002). There is a shortage of research examining how students change the manner in which they review their notes during the semester to adapt to the requirements of a course and what effect active engagement with notes after class has on student performance. Research is needed to investigate potential interventions that provide students with more effective methods of note review and note restructuring.

A study of the benefits of after-class interaction with class notes is important for several reasons. First, because class notes are an incomplete record of course content, there is a need for students to recognize the importance of revising their class notes (Boch & Piolat, 2005; Williams & Eggert, 2002). Second, through the addition of active interventions, students can learn to review, revise, and restructure their class notes to improve the quality of their notes and to provide a better resource for their preparation for exams (Bjork et al., 2013). By demonstrating the importance of note revisions and restructurings students can learn to create a valuable tool they can use to further their development as students and can utilize in future courses.

Purpose

Developmental students are often lacking in essential college success skills, like note taking (Eades & Moore, 2007). The purpose of this review of literature is to examine the benefit of note taking during class and the review and revision of notes after class. Because college students struggle to take quality notes, interventions that have been shown to improve the quality of class notes and to improve academic performance will be investigated. In order to address the deficiencies in the class notes taken by students, many professors provide their students with skeleton guided notes and others choose to provide students with complete lecture notes (Cornelius & Owen-DeSchryver, 2008). The benefits of providing faculty-produced notes to help students take more complete and correct class notes will be examined.

The note taking process should not end when class ends (Bjork et al., 2013; Boch & Piolat, 2005). Students must understand the benefits of review and the benefits of actively engaging with their class notes in order to create a more complete product suitable for preparing for exams (Cohen et al., 2013; King, 1992). The strategy of passively rereading verbatim notes will not lead to greater learning (Bjork et al., 2013). The review of literature will examine active engagement and the influence on student success when providing students with active and constructive interventions designed to help students review, edit, and restructure their notes after class. Providing developmental students with an understanding of the difficulties inherent in the note taking process, along with interventions shown to improve performance, may provide them with a framework on which they can build future academic success.

Method

This review of literature was conducted through an extensive search of multiple databases and websites. With the exception of the seminal note review study of underprepared college students conducted by King (1992), the timeframe of articles was restricted to those that were published after the year 2000. The search for articles was conducted in three stages. In each stage, the educational databases ERIC, Education Source, and JSTOR were searched to find articles related to the general process and benefits of note taking in class. In addition to the databases, a search of Google Scholar, a search of the website for the American Mathematical Association of Two Year Colleges (AMATYC), and a general internet search was conducted. In the first stage of finding articles, the search criteria were restricted using the terms note taking benefits, note taking developmental mathematics, note taking study skills, note taking process, guided notes, and benefits of guided notes. After collecting articles about note taking, a search was conducted to find articles about active learning. The same databases and internet resources were searched, using the terms active learning, active learning developmental mathematics, benefits of active learning, active learning interventions, active learning improves performance, and criticism of active learning. In the third phase of searching for articles the focus was on note taking interventions and the review and editing of class notes. The search terms used were note taking review, note taking interventions, after class note taking, editing lecture notes, note review developmental mathematics, revising lecture notes, and teaching effective note taking. Further investigation was conducted using the same databases and the terms metacognition, student perspectives of guided notes, study skills of developmental mathematics students, and self-regulated learning.

Review of Related Research

Students take notes in class in order to have a record of material they believe they will need to know in the future (Boch & Piolat, 2005; Bohay et al., 2011; Bonner & Holliday, 2006; Castelló & Monereo, 2005; Kobayashi, 2006; Van Meter et al., 1994). Taking and reviewing class notes has been shown to improve recall of information and performance on tests for students at all levels of higher education (Bohay et al., 2011; Cohen et al., 2013; King, 1992; Luo et al., 2016). Unfortunately, class notes are often an incomplete record of lecture content. Many researchers have attempted to quantify the quality of student class notes, determining that students capture, on average, 30% to 40% of key lecture concepts (Bonner & Holliday, 2006; Chen, 2013; King, 1992; Kobayashi, 2006; Luo et al., 2016; Williams & Eggert, 2002).

To be an effective study tool, class notes need to be as complete and correct as possible. It is essential that students revise their notes soon after class in order to ensure they possess a complete record of course material (Boch & Piolat, 2005; Williams & Eggert; 2002). This revision process can provide students with an opportunity to fill in missing information in their notes, correct any errors, or identify questions that need to be answered or details that need to be embellished (Cohen et al., 2013; Luo et al., 2016). Luo et al. (2016) differentiated between revision and review as “the former process strives to add information to existing notes, whereas the latter process strives to commit noted information to memory” (p. 47). Taking the time after class to assess the quality of class notes and make necessary revisions is an important aspect of effective note taking and provides students a better tool with which to prepare for exams.

Prior to exams, many students never review their class notes (Cafarella, 2014; King, 1992), others choose to review by rereading their notes (Bjork et al, 2013; Boch & Piolat, 2005; King, 1992; Williams & Eggert, 2002), and some choose to recopy their class notes (Bjork et al., 2013; Chen, 2013; Luo et al., 2016). These passive note review activities have demonstrated limited impact on performance (Bohay et al., 2011; Chen, 2013; King, 1992; Luo et al., 2016). Review activities that require more active engagement from students have demonstrated improved performance on assessments (Bohay et al., 2011; Cohen et al., 2013; King, 1992; Kiewra, 2016; Luo et al., 2016). The benefits of review are increased when the process involves more focus and intent (Boch & Piolat, 2005; Eades & Moore, 2007; Williams & Eggert, 2002).

Much of the research addressing the benefit of note taking and review focuses on students taking notes in response to a controlled lecture, which is typically videotaped or recorded, and then reviewing for an immediate test or a delayed test (Bohay et al., 2011; Luo et al., 2016; Williams & Eggert, 2002). There is a shortage of research examining the extent to which active engagement with notes after class has on student performance. A study of the benefits of after-class interaction with class notes is important for several reasons. First, because class notes are an incomplete record of course content, there is a need for students to recognize the importance of revising their class notes (Boch & Piolat, 2005; Williams & Eggert, 2002). Second, through the addition of active interventions, students can see the benefits of engaging with their class notes. By demonstrating the importance of reviewing, revising, and restructuring class notes, students can learn to create a valuable tool they can use to be more successful and a technique they can use in future courses.

Note Taking in the College Classroom

Almost all college students believe taking notes is an important part of being a successful student. Williams and Eggert (2002) reported that 94% of students believe taking notes in class is an essential element of learning. In interviews with genetics students, Bonner and Holliday (2006) found 91% said they always or usually took notes in class. Van Meter et al. (1994) conducted a study of 252 undergraduate students at a major public university in the United States, in which they found that every participant took notes during class.

Quality of Class Notes

Despite the desire to take notes during class, students come into college without much experience with taking notes from lecture, which is the most common delivery method for information in college classes (Badger et al., 2001; Bonner & Holliday, 2006; Bui et al., 2012). As a result of their lack of practice, the quality of student notes suffers. Like any skill, note taking develops over time and, as a student progresses academically, the quality of their note taking improves (Boch & Piolat, 2005; Kobayashi, 2006; Van Meter et al., 1994; Williams & Eggert, 2002). The quality of students' note taking needs to improve as quickly as possible, and Boch and Piolat (2005) contended "learning to take notes well undoubtedly takes as much time as learning to write in a relatively experienced way" (p. 110).

Note taking is a complex skill requiring students to undertake several complicated processes at the same time. Being able to effectively listen to the lecture, identify key points, and correctly and completely record the ideas is difficult and requires a great deal of focus (Al-Musalli, 2015). There is often a difference between what students believe constitutes good note taking and the product they are able to produce (Bonner & Holliday, 2006; Castelló & Monereo, 2005). When comparing students' beliefs of effective note taking to the notes they took, Castelló and Monereo (2005), determined "the students' declarative knowledge is *better?* [*sic*] than their procedural knowledge; in other words, they 'know' what they have to do, but they do not usually do it" (pp. 274-275).

Many studies (Bonner & Holliday, 2006; Chen, 2013; King, 1992; Luo et al., 2016) have considered the quality of student class notes by comparing the number of ideas or concepts recorded in class notes to a master list created by the instructor. Williams and Eggert (2002) reported that the average student records 30% to 40% of key concepts in a lecture. Chen (2013) evaluated the class notes of 38 first-year students and found they recorded, on average, 43% of key lecture points. Luo et al. (2016) found 59 undergraduate students enrolled in an educational psychology course captured, on average, 38% of key ideas in a 14-minute taped lecture.

In a qualitative examination of the note taking habits of 23 undergraduate students enrolled in an upper-level genetics course at a women's college in the United States, Bonner and Holliday (2006) conducted interviews with each student at five different times during the semester. Although students were not provided with guided notes, they were given copies of the instructor's transparencies which were used to guide each day's lesson. After class each day, class notes were collected and photocopied returned to the students. At the end of the semester all note material was collected again. This collection included notes taken outside of class and any notes recorded in textbook margins, including highlighted passages. The two sets of notes were evaluated for their completeness and correctness against a master list created by the instructor. The authors were able to determine which additions students made outside of class to their in-class notes through a process of comparison.

For the first unit, students' after-class notes were found to contain 36% of key lecture concepts, which was an improvement over the 31% of concepts found in their class notes (Bonner & Holliday, 2006). After their experience with the first exam, students understood the importance of supplementing their notes outside of class. The in-class notes for the second unit were found to contain 45% of key concepts and the third unit notes contained 25% of key concepts (Bonner & Holliday, 2006). The authors determined that during the second and third unit, students supplemented their notes outside of class time and raised the quality to 61% and 60%, respectively (Bonner & Holliday, 2006). Despite being aware of the ineffectiveness of their notes, students struggled to capture, on average, even half of the key concepts during lecture. Supplementing their notes after class only led to students capturing less than two-thirds of the main ideas.

Developmental students may lack many of the skills necessary for success in college, including taking effective class notes (Eades & Moore, 2007). For underprepared developmental students, poor class notes may have a greater impact on their academic performance. Cafarella (2014) conducted a study of 20 community college instructors who taught developmental mathematics courses at a community college in Dayton, Ohio. The participants identified attendance and lack of quality note taking as primary obstacles to student success. One reason why developmental mathematics students may not take notes is because they are focused on following material written on the board and do not want to miss anything while taking notes (Eades & Moore, 2007). Students who transitioned from unsuccessful to successful in their developmental mathematics classes reported that a primary difference between their classroom efforts as successful students as opposed to when they were unsuccessful students was their ability and desire to take extensive notes in class (Howard & Whitaker, 2011).

Impact of Note Taking on Recall

Notes taken during class serve the purpose of providing students with a record of material covered during class which might be needed in the future (Boch & Piolat, 2005; Bohay et al., 2011; Bonner & Holliday, 2006; Castelló & Monereo, 2005; Kobayashi, 2006). Students are more likely to record concepts when they believe the information recorded in class notes may be necessary for homework, essays, or exams (Badger et al., 2001; Bonner & Holliday, 2006; Van Meter et al., 1994). Several researchers have attempted to determine the benefits of note taking (Bohay et al., 2011; Titsworth & Kiewra, 2004). Titsworth and Kiewra (2004) found that students who took notes during class demonstrated improved recall of facts immediately after a lecture and after some delay.

Bohay et al. (2011) conducted experiments with two different groups of undergraduate psychology students at Notre Dame University to determine whether there was a benefit to taking notes during a lecture. In the first experiment the researchers had 97 students read two of three different texts and in the second experiment the researchers presented 77 students with video lectures. In the first experiment the participants were divided into two groups, one who took notes and one who did not. In the second experiment the participants were divided into a group who took no notes, a group who took notes by hand, and a group who took notes using a computer. In each experiment the researchers found that students who took notes outperformed those who did not on tests of recall and demonstrated improved performance on tests of deeper comprehension (Bohay et al., 2011). Interestingly, the researchers found that the process of taking notes either by hand or on the computer, which they associated with paying increased attention, was more valuable than the quantity of notes taken.

Titsworth and Kiewra (2004) investigated the impact of providing explicit spoken cues during lecture on student recall. The researchers divided both the cue and non-cue groups of participants into subgroups, those taking notes and those not taking notes. After hearing a lecture for which students were instructed to listen and either take notes or not, they were given a 5-minute review period and then were given a 10-minute organization test and a 15-minute detail test. The researchers found a statistically significant difference in test scores for the students who took notes and the students who did not, although the practical significance was moderate. Students who took notes were found to recall approximately 13% more content than students who did not take notes.

Improving the Quality of Class Notes

Because students recognize the importance of taking class notes but may not be aware of the lack of quality in their class notes, it may be valuable for faculty to find ways to improve the quality of class notes (Williams & Eggert, 2002). There are a variety of techniques available to instructors for helping students construct more valuable class notes. Two strategies include incorporating verbal aids for students into lectures and providing students with instructor developed complete or partial lecture notes (Boch & Piolat, 2005; Kiewra et al., 2018; Williams & Eggert, 2002).

Incorporating Verbal Strategies

Integrating some relatively simple verbal techniques into a lecture can benefit the quality of students' class notes without placing an undue burden on the instructor (Titsworth & Kiewra, 2004). Note taking is a complex skill and teaching students to be better note takers requires time and may not be something instructors are prepared to teach (Boch & Piolat, 2005). Instructors can make the note taking process easier for students by incorporating deliberate verbal strategies during lecture (Boch & Piolat, 2005; Williams & Eggert, 2002).

Titsworth and Kiewra (2004) investigated the impact of providing 60 undergraduate students at a large midwestern university with verbal cues during lecture. The students were enrolled in a basic communication course and were divided into one of four groups: (a) cued, taking notes, (b) cued, not taking notes, (c) not cued, taking notes, or (d) not cued, not taking notes. Students were presented with one of two audiotaped lectures and were given a test to determine the amount they had learned from the lecture. In one of the audiotaped lectures the instructor provided verbal cues identifying concepts as important whereas no verbal cues were provided in the other lecture. Each cue provided students with the topic and subcategory as well as a numeric indicator identifying the order in which concepts were introduced. The content in the two lectures was identical but there was no provision of any topic or subcategory information. After listening to a lecture for which students were instructed to listen and either take notes or not, they were given five minutes to review and then were given a 10-minute organization test and a 15-minute detail test.

During the lecture, students were presented with four communication theories and five subcategories for each theory (Titsworth & Kiewra, 2004). The quality of participant notes was scored by two outside observers. Each observer gave the class notes two scores, one based on the number of theories and subcategories included in the notes and the other based on the number of details included. The average of the two observer scores was used as each students' notes score. Following the lecture, all students were given 5 minutes to review before they were given the two tests. Students in the notes group used the time to review their notes and those in the non-note taking group were told to review the lecture in their heads. The same observers scored each test and the average of the two became each students' score. Each test was scored using the same rubric as that used to assess note quality.

The use of verbal cues during lecture to signal which concepts were important proved valuable for improving student recall (Titsworth & Kiewra, 2004). The researchers estimated that 48% of the variance in student recall of organizational information was due to the provided cues. Students who were provided verbal cues recalled 45% more organization points and 15% more detail points than those students who were not provided verbal cues. For the organizational test there was a statistically significant difference in test scores for the students who were cued during lecture, $F(1, 56) = 52.17$, $p < .001$, $\eta^2 = .48$, when compared to those who were not cued. The overall effect of the cued lecture on the detail test was also statistically significant, $F(1, 56) = 20.95$, $p < .001$, $\eta^2 = .27$. Interestingly, there was not a statistically significant result on the organizational test for those students who were cued and took notes but there was a statistically significant interaction for those students who were cued and took notes on the detail test, $F(1, 56) = 15.70$, $p < .001$, $\eta^2 = .22$. The combination of hearing a cued lecture and taking notes resulted in the greatest amount of recall of details.

In addition to improving student recall, the use of verbal cues in lecture improved the quality of student class notes (Titsworth & Kiewra, 2004). Students who were given verbal cues recorded 60% of key concepts from the lecture, including four times more organizational information and twice as much detail as students who were not cued. The percentage of information captured in class notes by the cued group was far higher than that found in other note taking research, which may indicate that cueing students results in higher quality notes. The multivariate impact of verbal cues on the quality of class notes was determined to be statistically significant, $F(2, 30) = 23.42$, $p < .001$, $\eta^2 = .61$. Univariate testing determined that the use of lecture cues had statistically significant impact on both organizational points, $F(1, 31) = 40.40$, $p < .001$, $\eta^2 = .57$, and details, $F(1, 31) = 39.74$, $p < .001$, $\eta^2 = .56$.

The act of taking notes helped students with recall and in performance on the tests (Titsworth & Kiewra, 2004). Students who took notes, whether prompted through cueing or not, recalled 13% more information than those who did not take notes. Those participants who took notes saw a small but statistically significant improvement on the organizational test, $F(1, 31) = 5.06$, $p < .05$, $\eta^2 = .08$, and a statistically significant impact on the detail, $F(1, 56) = 12.03$, $p < .001$, $\eta^2 = .18$. The researchers examined the correlation between the quality of note taking and the recall of both organizational and detail concepts. Not surprisingly, there was a strong correlation between student notes containing high numbers of organizational points and the recall of organizational points ($r = .66$) and details ($r = .76$). Additionally, those students whose notes had greater numbers of details were highly correlated with recall of organizational points ($r = .61$) and details ($r = .70$).

Providing Lecture Notes

Because many faculty recognize the incomplete nature of student class notes and because notes are necessary for students to have as a resource for studying in the future, many faculty members choose to provide students with some form of lecture notes (Cardetti et al., 2010). Notes provided to students can take on many forms, including providing complete lecture notes, outlines of the lecture, and partial, or guided notes (Kiewra et al., 2018; Williams & Eggert, 2002). One benefit of giving students instructor-generated notes is the improved quality of their class notes and the subsequent improved quality of the resource they have for studying (Austin et al., 2004). Several researchers have investigated the value of providing complete lecture notes versus guided notes with mixed results (Cornelius & Owen-DeSchryver, 2008; Stutts et al., 2013).

In an investigation of faculty and student perceptions of the provision of instructor-created lecture notes, Landrum (2010) surveyed 76 psychology students and 53 faculty members. The author created different questionnaires for students and for instructors, although there were some common questions asked of both groups. Students were asked to complete a Likert-type questionnaire using paper and pencil and received credit for partially fulfilling a research experiment requirement in their general psychology course. Emails were sent to 200 faculty members from departments across campus with a link to an online survey.

Landrum (2010) considered the perceptions of students and faculty about the general process of note taking in class as well as their perceptions of the provision of instructor-generated class notes. Faculty overwhelmingly agreed (83.0%) with the statement that they expected students to take notes in class. There were divergent responses between students and faculty in regards to the preferred method for relating course content. In questions asked only of students, 68.1% of students identified that they agreed or strongly agreed with the statement that they prefer when their instructors use PowerPoint for presenting lecture notes. Using PowerPoint to present lecture notes in class was far more preferred by students than using PowerPoint for presenting information other than lecture notes (21.0%), using a chalkboard or whiteboard (19.7%), or using overhead transparencies (14.9%). Faculty responses were not mutually exclusive and they were permitted to identify all methods they use for providing lecture notes to students. Faculty identified a preference for the use of the chalkboard or whiteboard (59.7%) over the use of PowerPoint to present lecture notes (40.3%). Instructor responses indicated a far lower preference for the use of overhead transparencies (21.1%) or the use of PowerPoint to present content but not lecture notes (9.6%). The researcher asked faculty participants what types of notes they provide to their students, to which 25% of respondents agreed or strongly agreed to the statement that they provide no notes and 23% agreed or strongly agreed to the statement that they provide guided notes with blanks for students to complete during class. There is a clear difference between the manner in which students prefer to get lecture notes from instructors and the manner in which faculty are most comfortable delivering lecture notes.

In addition to questions that were written exclusively for students and those that were written only for faculty, the researcher included identical questions on both surveys in order to compare the responses. Of the eight identical questions, there were statistically significant differences between student and faculty responses on four (Landrum, 2010). The author used a Bonferroni correction and determined statistical significance for the t tests with $p < .006$. Students felt, at a statistically significantly greater rate than faculty, that lecture notes should be provided to students prior to the start of class, $t(127) = -7.28$, $p < .006$. Faculty, though, disagreed statistically significantly more than students that the provision of lecture notes prior to class has no impact on attendance, $t(127) = 3.35$, $p < .006$. Students indicated a greater agreement than faculty that overall course grades would increase if faculty-provided lecture notes prior to class, $t(127) = -5.41$, $p < .006$, and that the provision of guided notes with blanks to be completed in class would lead to improved attendance, $t(127) = -5.92$, $p < .006$.

Results of Landrum’s (2010) investigation seem to indicate that faculty are not presenting information in the manner that is most preferred by students, which may account for some of the inaccuracy of student class notes, and there is a concern on the part of faculty that providing complete lecture notes will lead to increased absences. The presentation of content in a manner that students find most useful is important to ensure that students are capable of taking effective notes. In the case that faculty are not comfortable presenting information in this way, the provision of lecture notes can help to bridge the gap between the comfort of faculty and that of students and can ensure students have an effective resource to use when studying. The faculty concern regarding attendance is valid. When students are not in attendance, they cannot benefit from the interaction that comes from the answering of student questions and cannot see examples that may naturally arise during a class session (Zientek et al., 2013). Landrum (2010) suggested that the provision of guided notes requiring students to fill in blanks or providing a minimal outline of lecture will mitigate some of the attendance issues because students will need to attend in order to have a complete set of notes. Additionally, Landrum (2010) suggested the provision of notes at the beginning of class, rather than providing them prior to class, although he recognized that this would not allow students to review a complete and correct set of notes prior to coming to class.

Full Lecture Notes

Because of the low quality of student-generated class notes, many instructors choose to provide complete lecture notes for students. The provision of complete notes provides students with several benefits. Students are guaranteed to have a complete and correct record of class content to use when studying, which is an important element of review that is missing when students must use their own incomplete class notes (Williams & Eggert, 2002). In addition, the provision of complete lecture notes prior to class meeting gives students an opportunity to familiarize themselves with the content to be covered in class (Vandehey et al., 2005). Faculty have voiced concerns that class attendance will be negatively impacted by the provision of complete lecture notes, as they believe students will be less motivated to attend class when they will receive a complete record of course materials (Cornelius & Owen-DeSchryver, 2008; Landrum, 2010). There are additional concerns among faculty that students will pay less attention and will participate less when they are provided with complete lecture notes (Landrum, 2010).

To determine the extent of the impact of providing complete lecture notes on student performance and attendance, Grabe (2004) investigated the performance of introductory psychology students in a moderate-sized mid-western university. The instructor provided online access to complete lecture notes for the 183 students who completed the course. Each student had equal opportunity to access the notes throughout the course. Several measures were used to determine students' perceptions, aptitude, and achievement. In the third week of the semester, students completed a 6-item survey in which they were asked to rate the extent to which statements pertaining to their achievement goals were true. In the twelfth week of the semester, students were asked to gauge their usage of the online lecture notes. Only those 95 students who accessed the notes more than six times during the semester completed the survey. Student achievement was determined based on performance on the three, 50-item, multiple-choice tests given during the semester. In order to ensure the researcher could determine usage, all students accessed the class notes through an online system that required them to enter an identification code in order to see the notes. Each time a student accessed the notes, the specific pages accessed, the date, and the time of access was noted. Total time spent on the page was not used because it was possible for students to access the notes and print them, which made the time spent looking at each page unreliable as a measure of time spent engaging with lecture notes.

In addition to tracking which students were users and which were nonusers for each exam period, an aspect of interest to the researcher was when students accessed the course notes. Grabe (2004) created three variables to track student usage. The variables used were Before, for students who first accessed the notes before the last lecture of a unit, Unique After, for those students who first accessed the notes after the last lecture of a unit but more than two days prior to an exam, and Unique Cram, for those students whose first access of the lecture notes was within two days of the exam. Grabe (2004) also quantified the number of times students accessed the notes using the same time frames around the last lecture of the unit and the unit exam.

Grabe (2004) was also interested in determining whether students would access lecture notes if they were readily available and what impact access to complete lecture notes would have on achievement and attendance. Of the 183 students who completed the course, 73.8% accessed the lecture notes on at least six occasions. Through ANOVA and responses to the usage questionnaire, the researcher was able to determine that statistically significantly more students accessed the lecture notes before class, $F(2, 268) = 51.07$, $p < .05$, and 82% printed their notes and then used the printed notes during class. Students stated on the questionnaire they chose to bring the printed class notes to class to support the class notes they took in class and to aid in their understanding of the lesson content. In addition to accessing and printing the notes prior to class, 48% of the students surveyed indicated that they had reviewed the notes after class between one and five times and an additional 33% indicated they had reviewed the online notes at least six times during the semester.

Grabe (2004) was interested in determining whether accessing complete lecture notes online would impact student performance on examinations. It should be noted that the researcher did not preclude students from taking their own notes during class and using the online lecture notes to supplement their own class notes, nor does this seem to be a question asked of students on any of the survey tools. Students who accessed the course notes six times or more scored statistically significantly higher on exams one, $F(1, 137) = 9.99$, $p < .05$, and two, $F(1,148) = 5.25$, $p < .05$, than students who did not access the course notes at least six times. There was no statistically significant difference in performance on exam three. Interestingly, there was no statistically significant difference found on exam performance based on when students accessed the class notes.

The concern about the provision of complete lecture notes impacting student attendance was also addressed. Students were asked how frequently they had used the course lecture notes as a replacement for course attendance (Grabe, 2004). Of those students who accessed the course notes at least six times, 21% never used the notes as a replacement for attendance, 50% used the notes instead of attending class between one and five times, and 29% used the notes as an alternative for attendance six times or more (Grabe, 2004). The only measure that showed a statistically significant difference between those who never used notes as a replacement for class and those who used the notes as a replacement for class six or more times was performance motivation (Grabe, 2004). The performance goals for those students who never used their notes as a replacement for class were higher than for those who used the notes as a replacement at least six times, which is an indication of a student having greater concern for achievement measures like GPA or higher grades (Grabe, 2004). Grabe (2004) was clear to note that there was no survey question intended to determine the reasons for missing class and no way to definitively associate missing classes with the availability of class notes.

In another study examining the impact on student performance of providing complete lecture notes, Raver and Maydosz (2010) compared performance on assessments for students who received no notes to those who either received instructor-generated notes before or immediately after lecture. The researchers investigated the performance on a posttest for 154 students at an urban, public university in the southeastern United States. Three different conditions were investigated: (a) no notes provided, (b) instructor-written notes were provided during the last five minutes of class, and (c) instructor-written notes were made available to be downloaded prior to class. Three sections of the same introductory special education course, all taught by different instructors, were randomly assigned to one of the three conditions. The three groups consisted of 42, 47, and 65 students respectively. The notes given to students contained the PowerPoint slides presented during the video-taped lecture and provided space for students to add to the notes if they desired. Students were neither encouraged nor discouraged from supplementing the distributed notes.

During the study, students in each section were shown identical video-taped lectures given by an instructor who was not teaching one of the three sections (Raver & Maydosz, 2010). Because of differences in course scheduling, the duration of each section was different but content was identical. Depending on the section students were enrolled in, the study was conducted between 6 and 13 class sessions. Students were given a 50-question multiple choice pretest during the first class session and an identical posttest within seven days of the conclusion of the course content. Both tests were administered via Blackboard LMS and were written by the instructor who delivered the video-taped lessons.

All students in each section were shown the video lessons. Participation in the study was voluntary and neither pretest nor posttest scores counted in their course grade (Raver & Maydosz, 2010). Students were asked to prepare for the posttest as they would any examination and understood they would be tested, in an exam that would count towards their course grade, on the video-taped material following the completion of the study. In addition to taking both tests, all participants were asked to complete an online survey to determine their perceptions of the provision of instructor-generated notes, the timing of when notes were provided, and the impact of the timing of the notes on their performance. Of the students who were enrolled in the three sections, 60% volunteered to take place in the study.

The students who received instructor-generated notes outscored those students who did not have access to the notes (Raver & Maydosz, 2010). Pretest and posttest results were statistically significantly correlated for all participants ($r = .554$, $p < .001$, $r^2 = .307$). Posttest results for each of the groups who received class notes, whether they were received before or after the lesson, were statistically significantly higher than the group who received no notes, $F(2, 154) = 16.902$, $p < .001$. On a 50-item test, students in the notes before group outscored the no notes group by an average of 6.8 points and students in the notes after group outscored the no notes group by an average of 5.4 points, which was noteworthy. There was no statistically significant difference in the performance of students who received the lecture notes before versus those who received the notes after the lecture.

Students' responses to the survey given after the posttest provided some differentiation in the students' perceptions of the provision of notes before versus notes after groups. Students were asked in the survey to identify the most important strategy used by the instructor to help students learn the material. Whereas students in the no notes (38.9%) and notes after (24.5%) groups identified reading the assigned chapter as the most helpful aspect of the course, students in the notes before group (72.6%) identified the instructor-provided notes as the element that most helped them learn (Raver & Maydosz, 2010). Most students in this group printed the notes before class and used them during class to better follow the lecture. Some students in the no notes group commented that the speed of the lecture was too fast for them to take effective notes, and eventually they stopped trying. The lecture speed was calculated to be between 110 and 127 words per minute in three randomly selected lectures, which was determined to be a typical lecture pace.

Guided Notes

In order to encourage students to take notes, and to improve the quality of notes, many instructors provide students with guided notes, which are also referred to as partial notes (Austin et al., 2004; Cardetti et al., 2010). Typically, guided notes are an incomplete version of the professor's lecture notes and students are tasked with filling in information throughout the lecture (Cornelius & Owen-DeSchryver, 2008; Williams et al., 2012). Guided notes provide an outline for students to follow that emphasizes key concepts, provide organization to their class notes, and can help reduce the number of errors present in students' class notes (Cornelius & Owen-DeSchryver, 2008). An added benefit of providing guided notes rather than complete lecture notes is that students are forced to attend class and to remain attentive throughout (Cardetti et al., 2010).

To examine differences in the quality of notes taken when students were required to take their own notes compared to when students were provided guided notes, Austin et al. (2004) conducted an experiment with 23 undergraduate applied psychology students at a large southeastern university. Students were unaware they were a part of the experiment until the end of the semester. During the semester, each class day was taught using one of three methods: (a) students were not given any guided notes and no visual aid was employed, (b) students were not given any guided notes and overhead slides were used to guide the lesson, and (c) students were given guided notes and overhead slides were used to guide the lesson. Students had the opportunity to learn in each class format multiple times throughout the semester. Each class format was evaluated based on the number of key points addressed in the lecture and the difficulty of the concepts. Days with similar numbers of key points and difficulty levels were chosen as experiment days.

In order to determine the quality of notes on the selected experiment days, the researchers collected student notebooks (Austin et al., 2004). The number of key concepts and examples were identified relative to a list created by the instructor. When students were required to take their own notes and no visuals were provided, students recorded 62% of key concepts and 13% of examples presented by the instructor. When provided overhead slides, but no guided notes, students captured 97% of key concepts and 26% of examples in their notes. The authors found that, when the lesson included overhead slides and students were provided guided notes, students recorded 100% of key concepts and 60% of examples from the lesson. Using a one-way ANOVA, the authors determined there was a statistically significant difference in both the percentage of key concepts identified, $F(2, 21) = 232.75$, $p < .0001$, and the percentage of examples recorded for the lesson, $F(2, 21) = 49.36$, $p < .0001$, when students were provided overhead slides and guided notes. Post-hoc *t* tests were conducted to conduct pair-wise comparisons, with the guided notes lecture statistically significantly outperforming the other formats in the number of critical points ($p < .006$), examples ($p < .004$), and extra points ($p < .032$).

In another study examining the impact of providing students with guided notes, Williams et al. (2012) examined the effect of providing guided notes for 71 students in two undergraduate psychology courses. Students were provided guided notes in some weeks and had to take their own notes in other weeks. The goal of the study was to compare quiz results under two different conditions. One examination considered the performance of students in weeks when guided notes were provided compared to weeks when students were given no guided notes. Additionally, the authors wished to compare the performance of students in Class 1, who were given a 5-minute review and then tested during the last 20 minutes of class, to the performance of students in Class 2, who were tested at the start of the following class session. In both classes students had higher quiz grades during the weeks when guided notes were provided and the improvement was approximately equal to a one grade increase (Williams et al., 2012). Overall, quiz grades in Class 2 were lower than those in Class 1. The authors hypothesized the reason was due to the 2-day or 4-day break between classes and a lack of review of the guided notes.

Cardetti et al. (2010) investigated the benefit associated with providing mathematics students with partially completed guided notes. Mathematics courses present students with challenges not found in other courses because concepts are often presented in a manner that builds on prior knowledge and concepts can be complex enough to require students to review multiple times in order to understand (Cardetti et al., 2010). To investigate the impact of providing partial notes to calculus students at a large northeastern university during the Fall 2007 semester, students were provided partial notes at the beginning of each class day. In addition to determining the extent to which partial notes impacted student performance, the researchers surveyed the students to determine their perceptions of receiving daily guided notes. Student performance during the Fall 2007 semester, in which the students received guided notes constructed from the professor's lecture notes, was compared to performance during the Fall 2006 semester, in which students received no notes but lessons were guided by the same set of professor lecture notes. The two courses met for the same number of days and at the same time of day during each semester, and because both were guided from the same set of instructor notes, the timing of each class was identical.

The notes were developed with large blank spaces so students could write in the solutions to problems and take notes on the content (Cardetti et al., 2010). The instructor of the course noted that providing guided notes saved time during each class because it was no longer necessary to record examples or problems on the board, which was filled with interactive in-class engagements with students. Course assessments consisted of two exams, given each semester during the sixth and eleventh week of the course, and a cumulative final exam. Student perceptions of the provision of guided notes were determined through the administration of a mid-semester and an end-of-semester questionnaire. The first questionnaire was given two weeks after the first test and gathered anonymous feedback from students with regards to their learning experience and with suggestions for how the instructor could improve their teaching effectiveness. The end-of-semester questionnaire was given at the completion of the semester and consisted of several 10-point scale quantitative questions to evaluate teaching effectiveness and open-ended questions allowing students to evaluate the instructor's performance.

The results of the qualitative surveys indicated that students appreciated and felt they benefited from the provision of daily guided notes (Cardetti et al., 2010). In the qualitative survey students stated that they found they could follow the lecture more easily when they were not frantically writing out notes and could spend their time listening to the instructor. Students commented that the guided notes provided them a learning resource they were able to use when reviewing after class and preparing for exams. Students were also appreciative that the provided notes aided them in organizing the lecture and their notes.

The quantitative data also demonstrated that students with access to guided notes outperformed those who did not have access. In an attempt to determine whether one class was better prepared than the other for the calculus course, the researchers conducted a *t* test to assess the difference in the mean scores on the mathematics section of the SAT exam for each class (Cardetti et al., 2010). This measure demonstrated that the guided notes class scored statistically significantly lower on the mathematics portion of the SAT exam. Based on mathematics SAT scores, students in the section that were provided guided notes were less prepared prior to the class. On each of the midterm exams and on the cumulative final exam the students who had access to the instructor-generated guided notes outscored the students who did not have the guided notes. The scores on Test 1 for students with access to the guided notes were statistically significantly better, and the standardized mean difference effect size was determined to be very large, $t(52) = 3.50$, $p < .001$, $ES_{sm} = 0.90$. On Test 2, the researchers found a marginal statistically significant result and the effect size was determined to be marginal, $t(54) = 1.61$, $p = .056$, $ES_{sm} = 0.41$. On the final exam the guided notes class scored statistically significantly higher than the non-guided notes class and the effect size was determined to be moderately large, $t(55) = 2.62$, $p = .014$, $ES_{sm} = 0.67$. The combined impact of the guided notes class being less prepared for the course yet outperforming the non-guided notes class on each exam suggested that the provision of guided notes impacted student performance.

The notes developed by students during class create a record of content from the lesson. In order to be an effective record of course content, it is important that class notes be as complete as possible and accurately capture content. Despite the fact many college students know the importance of taking quality notes, many struggle to take notes that capture all of the important content. Taking effective notes is a skill which must be emphasized, developed, and taught (Austin et al., 2004; Boch & Piolat, 2005; Cafarella, 2014; Eades & Moore, 2007). While students are developing the skill, the provision of guided notes can help them identify key concepts (Austin et al., 2004). Additionally, guided notes can highlight relationships between ideas and cue students to the presence of the relationships. Care should be taken to ensure guided notes are incomplete. Even though complete lecture notes will contain 100% of key ideas, students lose their first opportunity to engage with the material. Cohen et al. (2013) stated, “in an attempt to give students the good ‘product,’ we may sometimes forget that it is the process, the engagement with the material—the cognitive exercise involved in recollecting, summarizing, reorganizing and restructuring—that actually matters most” (p. 98).

Comparison of Providing Full and Partial Notes

Several researchers have compared the impact of providing partial guided notes with providing complete lecture notes to students (Barnett, 2003; Cornelius & Owens-DeSchryver, 2008; Stutts et al., 2013). Various studies examined the difference in performance and attendance for students who received instructor-generated partial notes and those who received instructor-generated full notes. The provision of partial notes provides students with an opportunity to engage with material while, at the same time, providing them with an outline to make them aware of important concepts (Barnett, 2003). Full notes allow students the opportunity to focus on the instructor rather than having to divide their attention between taking notes and listening to the lecture (Kiewra et al., 2018).

Cornelius and Owen-DeSchryver (2008) compared the performance of students on examinations when provided with either instructor-produced full notes or instructor-produced guided notes. The participants in this study were 307 students enrolled in one of four introductory psychology sections taught by the authors. One section taught by each author was provided with full notes via Blackboard LMS and one section taught by each author was provided with guided notes via Blackboard LMS. A total of 153 students were assigned to the full notes group and 154 students were assigned to the partial notes group. Students in each group were instructed to print out the notes provided and to bring them to class each day. Exit interviews indicated that 78.75% of the participants in the full note group and 75.45% of the students in the partial note group printed and used the notes during most class sessions.

The authors used pretest information to determine whether the two groups were similar. Based on the pretest, the two groups had no statistically significant difference in demographics or in the understanding of psychological concepts, but the full notes group had statistically significantly higher average ACT scores (Cornelius & Owen-DeSchryver, 2008). No statistically significant difference existed between the group scores on Exam 1 and Exam 2, but the students assigned to the partial notes group scored statistically significantly higher on Exam 3, $F(1, 304) = 5.45$, $p = .02$, $\eta^2 = .018$, Cohen's $d = .20$, and Exam 4, $F(1, 304) = 5.05$, $p = .025$, $\eta^2 = .016$, Cohen's $d = .22$, than those assigned to the full notes group. Final course grades for the partial notes group were statistically significantly higher than for those in the full notes group, $F(1, 304) = 6.21$, $p = .013$, $\eta^2 = .020$, Cohen's $d = .19$. The authors hypothesized that students in both groups engaged in positive academic strategies but, as the course progressed, those in the full notes group relied more heavily on the provided notes while those in the partial notes group continued to engage with the material and continued to use good study strategies.

In addition to considering the impact of the different note styles, Cornelius and Owen-DeSchryver (2008) were interested in determining whether providing students with full notes would impact attendance. The attendance data indicated that students in the partial notes group had better attendance, but the difference was not statistically significant. In response to a question regarding attendance on the exit interview, statistically significantly more students in the partial note group agreed with the statement that provision of notes did not impact their attendance, $\chi^2(1, N = 297) = 11.56$, $p = .001$, $\varphi = .20$. Attendance was charted on only 20% of the class meetings, and the authors hypothesized that had the number of classes checked been increased the difference between the two groups would have been statistically significant.

In a similar study, Stutts et al. (2013) considered the impact of providing full notes versus partial notes to students enrolled in two consecutive fall semesters of an undergraduate animal science course. The authors investigated the impact of different note taking methods on academic performance and attendance. Data regarding academic performance and the method of taking notes were gathered for 160 students enrolled in one of four upper level animal science courses. The study was conducted over two consecutive fall semesters, with students in the first semester being provided an outline of course material, and students in the second semester being given full instructor-generated notes. The outline provided students with an overview of material to be presented during each lesson and the full notes contained all written material used by the instructor. In each year, course sections used for the study were taught by the same instructor and used similar assessment materials.

The performance of students in the four animal science courses were compared to determine whether the manner of notes provided to students impacted overall course grades. The comparison of all animal science students between Year 1 and Year 2 showed no statistically significant difference in final average grade or absences, which led the authors to assume there were no differences in the academic ability or attendance practices of students over the two-year period of the study (Stutts et al., 2013). The authors noted that although there was no statistically significant difference in attendance between the two groups, the partial notes group participants had, on average, more absences than the full notes group. The subgroup of students who received partial notes had statistically significantly higher overall course grades ($p < .05$) than the subgroup who received full notes.

In an experiment designed to test the effectiveness of providing guided notes, Barnett (2003) examined 74 undergraduate psychology students and tested them about a lecture they watched on videotape. Students were placed in one of four groups and provided: (a) instructor-created skeletal guided notes, (b) no notes and were required to take their own, (c) complete instructor notes, and (d) complete instructor notes but they were not allowed to watch the video. With the exception of students in the control group for whom complete notes were provided but the video was not watched, students watched a 35-minute video lecture about the brain. The topic was chosen because it was not a part of the curriculum of the course, which ensured students would have limited prior knowledge on the topic. After the video, notes were collected and students were tested on a 14-question short answer and fill-in-the-blank exam intended to test memory of facts.

Barnett (2003) compared the quality of students' notes in the two note taking groups against a list of 80 supporting concepts from the instructor-provided notes. Students in the skeletal notes group recorded an average of 56% of concepts and students who took their own notes recorded an average of 64% of concepts. The difference was not determined to be statistically significant. Of the students who took their own notes, 74% recorded all of the 16 key ideas from the lecture. Students in the skeletal notes group and those students who took their own notes outperformed the groups provided complete instructor-generated notes on the exam. There was no statistically significant difference in the performance of the two groups of students who had to take notes. In this experiment, participants who had the opportunity to actively engage with the material outperformed those for whom the interaction was strictly passive.

In a second experiment involving 40 students from the initial population, Barnett (2003) investigated how the four subgroups would perform on a test intended to examine recall of facts and students' ability to draw conclusions. Four questions were removed from the original test and replaced with questions requiring the participants to apply their understanding of the brain to concepts not addressed in the lecture. Again, the skeletal notes group and the group required to take their own notes outperformed the two groups provided with complete instructor notes. The students who took their own notes scored higher than the skeletal notes group, but there was no statistically significant difference in the performance of the two groups.

Whether the provision of partial notes or full notes is more effective may depend on the intended use of the notes (Cornelius & Owen-DeSchryver, 2008). If the goal is for students to actively engage with the material, pay closer attention to lecture, and better encode the content in their minds, then providing partial notes may be more beneficial (Cornelius & Owen-DeSchryver, 2008). If, on the other hand, the goal is to ensure students have a complete and correct resource to use when reviewing their notes and studying for exams, then providing complete lecture notes may help students more (Kiewra et al., 2018).

After Class Review and Revision of Class Notes

Review of Notes After Class

Instructors must teach students that class notes are a beginning to the learning process and not the end of learning. Boch and Piolat (2005) declared “note-takers are assumed to re-read their notes as many times as necessary for them to learn their content” (p. 104). In order to be an effective source for review, class notes need to be as complete and accurate as possible. If a student only captures 50% of the key lecture concepts, they only have half of the material with which to review (Williams & Eggert, 2002). Because students capture more information through the use of guided notes, the review process has added benefit.

The benefits of review are increased when the process involves more focus and intent (Boch & Piolat, 2005; Eades & Moore, 2007; Williams & Eggert, 2002). Reviewing notes by simply rereading them is a passive endeavor employed by many students. Of the 56 students interviewed by King (1992) for a study examining different review techniques, 76% reported that rereading their notes was the only review technique they utilized. In addition to only passively engaging with their notes when reviewing, many students struggle to review their notes in a timely manner. Students often wait until several days after class before revisiting their notes. This delayed time frame makes it impossible for students to identify missing concepts and weakens the effect of note review (Chen, 2013; Cohen et al., 2013; Williams & Eggert, 2002).

To determine the effect of after-class notes on students’ academic performance, Chen (2013) examined the class notes and after-class notes of 38 freshmen students enrolled in a general psychology course in southern Taiwan. Prior to the 4-week unit on Memory, the instructor created a 27-slide presentation. The students were given the presentation slides, printed with 3 slides per page, and encouraged to use the framework to take notes. After each class the researcher collected class notes, photocopied, and returned them. The notes were compared to a list of key concepts from each lecture developed by the instructor. In an attempt to determine how much students had altered their notes after class, the notes were collected and rescored at the end of each unit. Students' after-class notes only contained an average of 4% more key concepts than in-class notes (Chen, 2013).

At the end of the study, students were asked to complete a questionnaire to determine their study methods (Chen, 2013). Students stated that they spent time reviewing and rewriting their class notes after class. Upon further investigation, Chen (2013) determined most students simply copied their class notes in a different location without making noteworthy edits or additions. The goal of the study was to determine whether in-class and after-class note quality could predict academic success. The quality of in-class notes was predictive of academic performance, $F(2, 35) = 7.32$, $p < .01$, $R = .54$, $R^2 = .30$, $R^2_{adj} = .26$. Although the quality of combined in-class and after-class notes could also be used to predict academic performance, the correlation was weakened with the addition of the after-class notes and the predictive capability was lessened, $F(4, 35) = 3.54$, $p < .05$, $R = .55$, $R^2 = .30$, $R^2_{adj} = .22$.

Note Revision Interventions

Reviewing notes with the intention of identifying key concepts and drawing connections between concepts requires students to engage with the material. Luo et al. (2016) proposed that the note taking process involves three steps; students must take class notes, review their notes, and revise their notes in order to get the most benefit out of the process. The authors differentiated revision from review "in that the former process strives to add information to existing notes, whereas the latter process strives to commit noted information to memory” (Luo et al., 2016, p. 47). Rewriting, revising, and restructuring notes elevates the practice of review to a constructive intervention and, based on the work by Chi and Wylie (2014), would provide more benefit to students.

King (1992) considered three different note review strategies and attempted to determine which was most beneficial for students. The author conducted a study in which 56 remedial reading students were randomly assigned to review their notes using either self-questioning, summarizing, or using no intervention. The self-questioning and summarizing groups were trained in proper use of the interventions. After the six training sessions, the students viewed a videotaped lecture and were tested on the content. One week later they were retested to determine their retention of the material. Self-questioners were trained to take a generic list of questions, make them applicable to the specific content of the lecture, and answer them. Summarizers were trained to connect ideas from throughout the lecture and to summarize the concepts using their own words. The control group was given no training. Although all of the students were trained individually, they shared their work in groups and collaborated with other students in their groups to improve their methodology. The initial training taught students in both of the intervention groups how to interact with material on a constructive level and then provided engagement on an interactive level to refine the skills.

In a pretest, King (1992) found no statistically significant difference among the students in the three groups with respect to the strategies students used. After training, an ANCOVA showed a statistically significant difference among the three groups in the quality of their class notes, $F(2, 52) = 10.84$, $p < .001$, and in the performance of students on the posttest, $F(2, 52) = 3.49$, $p < .05$. Tukey post-hoc comparisons showed students in the two treatment groups captured a higher percentage of key ideas from the lecture and scored statistically significantly higher on the posttest than the control group ($p < .05$). There was no statistically significant difference in performance found between students in the self-questioning and the summarizing groups for either measure. In the retention test, conducted one week after the study, an ANCOVA showed a statistically significant difference between the groups, $F(2, 52) = 3.43$, $p < .05$, and Tukey post-hoc testing showed the self-questioning group outperformed the note review group by a statistically significant margin ($p < .05$).

For immediate recall of facts, both of the interventions improved the performance of students (King, 1992). Interestingly, the improvement was not limited to performance on a test but was also apparent in the quality of the notes taken during lecture by students in the self-questioning group. Following training, students in the self-questioning group recorded 14.2% of key ideas in the lecture whereas students in the summarizing and note taking groups recorded 12.4% and 12.2% respectively. The participants in this study were underprepared developmental students and the data reflect their inexperience. In an earlier study, King hypothesized that students trained in self-questioning would begin to employ the technique during lecture and would improve the quality of their notes through their use of the practice (as cited in King, 1992). Observation of the note-review students in the study during the time between the pretest and posttest showed that although all students reread their notes, fewer than 10% of the observed students practiced any more active review strategies, such as underlining or highlighting key concepts. This is indicative of poor study strategies typical of inexperienced and underprepared students (King, 1992).

Active learning techniques and interventions are able to increase student performance more than simple review (Chi, 2009). Cohen et al. (2013) investigated the effects of a note restructuring intervention on the scores of 79 psychology students' first unit exam. Each week one-fifth of the class was assigned to restructure their notes according to directions provided by the instructor. The intervention in this study involved students submitting a typed copy of their restructured notes from class. Students were told their grade was based on the completeness, clarity, and organization of the final product. They were responsible with writing a short summary of the main point of the lecture and a longer explanation of how concepts from that day's lecture related to another concept from the class.

On the unit exam, students correctly answered 72% of questions based on material from the week they restructured their notes, compared to 61% on the material from the weeks in which they did not restructure their notes (Cohen et al., 2013). Because students were required to rewrite and type their notes, the intervention met the criteria of being active (Chi, 2009). The authors used an arcsine transformation to transform the dependent variable because it involved the proportion of correct answers. Cohen et al. (2013) found a statistically significant difference in the percentage of correct answers for students in the week they completed the intervention when compared to the weeks in which they did not, transformed $t(78) = 4.87$, $p < .001$, Cohen's $d = 1.10$, untransformed $t(78) = 3.79$, $p < .001$. The magnitude of Cohen's d indicated a large effect size whether the dependent variable was transformed or untransformed. In order to determine whether the intervention aided stronger students disproportionately to weaker students, the authors considered the correlation of exam scores on the questions based on the lecture and those based on the readings. The belief was that stronger students would outperform their classmates on the reading-based topics. There was no linear or quadratic statistically significant correlation found, which indicated that all students were helped equally by the use of the intervention.

In their study of 59 undergraduate educational psychology students, Luo et al. (2016) conducted two experiments to determine the effects of note revisions on student performance. In the first experiment, students listened to, and took notes from, a tape-recorded lecture. At the end of the lecture they were given the opportunity to study the notes. One group was then asked to revise their notes while the other was asked to recopy their notes. Students completed a multiple-choice test that required them to answer fact-based questions in addition to relationship questions. The authors chose to use independent *t* tests with a Bonferroni correction, which made the critical value for statistical significance equal to .125. There was no statistically significant difference in scores for the revision and recopy groups on the fact-based questions. On the relationship questions the revision group answered 45.9% of questions correctly and the recopy group answered 33.0% correctly. The difference was determined to not be statistically significant but there was a moderate effect size, which indicates there is a practical significance to the difference, $t(57) = 2.06$, $p = .04$, Hedges $g = .53$. An investigation of correlation indicated that revising and adding content to notes was statistically significantly correlated to improved performance on the fact questions ($r = .310$, $p < .05$,

r^2 = .096) and the relationship questions (r = .494, p < .01, r^2 = .244). The authors noted that additions to notes made during revisions are related to achievement.

In an attempt to determine if there was a way to improve the revision effect, the researchers conducted a second experiment wherein they assigned students to revise lecture notes either during three separate 5-minute pauses during the lecture or for 15 minutes at the completion of the lecture (Luo et al., 2016). Students were further divided by randomly assigning half of each group to perform their indicated revisions alone, and the other half to conduct the revisions with a partner. Participants were 72 undergraduate education majors, none of whom were participants in Experiment 1. Students listened to one of two lectures: one that ran uninterrupted from beginning to end, and one that paused at three approximately equivalent time intervals.

The researchers conducted separate two-way MANOVAs in order to determine the impact of pauses and partners on note quality and achievement (Luo et al., 2016). Students in the pause groups captured statistically significantly more content in their original notes than those students in the non-pause group, $F(1, 66) = 10.20$, $p = .002$, $\eta^2 = .13$. Students who worked with partners also captured statistically significantly more content in their original notes over those students who worked alone, $F(1, 66) = 4.33$, $p = .04$, $\eta^2 = .06$. When working with partners, students in the pause group added a statistically significant amount of new content and extended content in their notes ($p < .001$). Students working alone added approximately the same amount of content to their notes whether the lecture was paused or not. Similarly, students who listened to the lecture without pauses added approximately the same amount of content to their notes whether they worked alone or with a partner.

Test scores were improved when students revised during pauses in instruction (Luo et al., 2016). Students in the pause group scored approximately 9% higher on the fact questions than the students in the non-pause groups, which was determined to be statistically significant, $F(1, 68) = 5.1$, $p = .03$, $\eta^2 = .07$. Working with a partner did not produce a statistically significant improvement in achievement, but an investigation of performance in the fact questions for students working with a partner shows an improvement of 12% over those students working alone. Although the difference was not determined to be statistically significant, 12% represents a large practical difference. Students in the pause group had statistically significantly higher scores on the relationship questions, $F(1, 68) = 14.6$, $p < .001$, $\eta^2 = .18$. There was no statistically significant result on the relationship questions for students working with a partner. Revising notes during intermittent pauses during class resulted in students improving the quality of their notes and improving performance on both fact and relationship questions. Although revising notes with a partner did not result in improved performance on the assessment, it did lead to more additional notes being created. The note revisions were determined to be both new notes and extensions of concepts students deemed to be incomplete in their notes. The improvements in the quality of student notes provide solid rationale for encouraging students to work with a partner to review and revise their notes and for instructors to provide regular breaks in the lecture to give students time to revise.

Effective note taking should continue beyond the end of class. Students should be encouraged to review their class notes and ensure the content is complete and correct. This process will make class notes a more valuable tool and can improve performance on assessments (Bjork et al., 2013). The more students interact with their notes and are provided with opportunities to revise and add to the content, the more students will benefit from the intervention (Luo et al., 2016). Effective and valuable reviewing and restructuring of class notes is a skill requiring practice to perfect. In the longitudinal studies that examined students during semester-long classes, very few of the students were willing to continue the prescribed intervention because of the time and energy necessary, which makes it more important that instructors teach students the value in note review and require students to edit their notes as a part of their daily success routine (Bonner & Holliday, 2006; Cohen et al., 2013).

Metacognition

Providing students with a mechanism for improving the quality of their class notes, which will ensure a higher quality product to use for review, addresses one aspect of improving student performance. A second important consideration is helping students recognize what they do and do not understand, which is essential element of effective studying (Howard & Whitaker, 2011). In order to prioritize the time available to study for exams and to learn new concepts, students need to be self-aware enough to recognize those concepts with which they are struggling (Nordell, 2009). This self-awareness is referred to as metacognition and is a key factor to providing students with a realistic perspective of their preparation and abilities (Bol et al., 2015). Several researchers have determined that students struggle to correctly assess their level of preparation or their level of skill development, and both negatively impact their academic performance (Bol et al., 2015; Nordell, 2009).

Schneider and Artelt (2010) defined metacognition as “people’s knowledge of their own information-processing skills, as well as knowledge about the nature of cognitive tasks, and of strategies for coping with such tasks" (p. 149). Researchers consider two aspects of metacognition to be important for academic success: an understanding of what students know and the regulation of that understanding (Langdon et al., 2019). Metacognition is important because it contributes to memory development as well as an awareness of the feelings elicited during complex activities, such as problem solving (Schneider & Artelt, 2010). The learning of mathematics requires attention to procedure, factual understanding, and an ability to determine which of various strategies to implement in order to complete a problem (Langdon et al, 2019; Schneider & Artelt, 2010; Van der Stel et al., 2010). Metacognition is a skill that can be taught to students and can improve student performance, particularly for underprepared students who may lack the self-awareness needed to develop other essential college success skills (Bol et al., 2015; Langdon et al., 2019; Nordell, 2009). Improved metacognitive skills have been demonstrated to increase motivation, improve course performance, and help students determine more appropriate strategies for studying (Langdon et al., 2019; Schneider & Artelt, 2010).

Students rely on their experiences and understanding when solving problems and answering questions. A better understanding of how problems are conceptualized can help students know when their solution strategy is effective, when solutions are reasonable, and when they need to learn a new skill in order to advance their factual understandings (Schneider & Artelt, 2010). Teaching metacognition to students provides an opportunity for them to reflect on past practices and a chance to modify those practices to improve results. Through this reflection students begin to note that the consideration of ineffective practices plays as large a role in learning, as does the discovery of successful learning techniques (Hudesman et al., 2014).

In an investigation of the benefits of metacognition training in teenage mathematics students in the Netherlands, Van der Stel et al. (2010) considered whether there was a relationship between metacognitive skillfulness and intelligence. Students with diverse educational goals were recruited to participate in the study that gauged the level of metacognition in a total of 59 students ranging in age from 13 to 15. Students were recruited from among second-year and third-year students at two diverse secondary educational institutions. Students began by taking a 100-minute intelligence test and then completed a 50-minute problem-solving session designed to measure the quality of their metacognition.

The problems given to students during the problem-solving session were chosen to be representative of concepts taught to each age group (Van der Stel et al., 2010). During the initial 20-minute session, students were given several age-appropriate mathematics questions to consider. Because the researchers were interested in the thought process being used to solve the problems, students were given the answers and step-by-step explanations for each problem. After spending 20 minutes working on the problem students returned all the material and were given 30 minutes to work through a set of parallel problems. Students were asked to think out loud as they answered the problems. The utterances of the students were judged for different levels of metacognition.

The spoken thoughts of students were judged purely on the levels of metacognition being used (Van der Stel et al., 2010). As students expressed their thoughts out loud their utterances were coded based on the type of metacognitive activity it reflected: (a) orientation, (b) planning and systematic orderliness, (c) evaluation, and (d) elaboration. The researchers considered the number of times the utterances for each category occurred to determine the frequency of each subcategory. Each utterance was also given a score on a 5-point scale based on the quality of the utterance and according to a rubric developed prior to the study. Student thoughts were rated on the quality of the utterance and not on the correctness of the process. Incorrect procedures that demonstrated high levels of metacognition would be highly rated.

As students get older, the benefits and impacts of metacognitive skillfulness become more important for their performance in their academic success (Van der Stel et al., 2010). As the participants in this study got older there were increases in both intelligence and metacognitive skills. An investigation using MANOVA demonstrated a statistically significant age effect for students, $F(4, 54) = 3.93$, $p < .01$, $\eta^2 = .23$. Between age 13 and 15 the impact of metacognitive skillfulness changed for the student participants. The researchers conducted MANOVAs on both the quantitative, $F(4, 54) = 13.84$, $p < .001$, $\eta^2 = .51$, and qualitative, $F(4, 54) = 4.90$, $p < .005$, $\eta^2 = .27$, aspects of metacognition and both showed statistically significant effects with regards to age.

As students got older, the relationship between mathematics performance and metacognition changed. Within the younger group of students, academic performance was impacted most by intellectual ability and a students' metacognitive skill made limited contribution to their mathematics success (Van der Stel et al., 2010). There were statistically significant correlations for younger students between intellectual ability and mathematics performance ($r = .79$, $p < .001$, $r^2 = .624$), and 33.6% of the variance in mathematics performance was due to intellectual ability. At the age of 15, though, the impact of the quality of metacognition (r = .78, p < .001, r^2 = .608) outweighed intelligence (r = .46, p < .05, r^2 = .212) in consideration of a students' mathematics performance. As students age, improvements in metacognitive skills play a larger role in academic success. The various metacognitive skills considered by the researchers did not grow in quality at the same rate. The metacognitive skills needed during the completion of a task developed before those skills associated with the time before (orientation) or after the task (elaboration). Participants in the study did not receive instruction on metacognition prior to the study, which indicated that improvements in metacognition were due to maturation and natural development. This finding was not in line with similar studies but may indicate that improvements in metacognition require students to reach a level of intellectual maturity before they are possible.

In an investigation into the impact of the teaching of metacognition strategies to college students, Nordell (2009) examined the impact of a one-hour advanced study skills workshop on exam grades of 348 students in an introductory biology course. The focus of the workshop was on helping students learn new study strategies, providing tools for effective self-assessment, and presenting information about learning styles. The presentation was designed to be interactive and to focus on providing students with new skills, rather than attempting to provide remediation. Workshops were offered immediately following the first exam. Performance on the second exam was then compared to the first exam for each student. Students were surveyed to determine their perceptions of the effectiveness of the study skill strategies presented in the workshop.

The workshops were designed to help students determine the effectiveness of the study strategies they used in preparing for the first exam (Nordell, 2009). The goal of this exercise was to draw students' attention to the most prevalent study technique of looking over their class notes. Students were then shown through several activities how their preferred study technique leads to, at best, a superficial understanding of the concepts. Once students better understood the deficiencies in their study techniques, they were provided with multiple options for more active engagement with course content, techniques for taking better notes, and more proven study strategies. Although the workshops presented students with techniques to improve their success, the primary goal was to help students better identify the weaknesses within their current strategy and to determine which new techniques would improve performance.

Participation in the workshops was optional following the first exam, and students could opt to attend one of four identical workshops (Nordell, 2009). Approximately 20% of the students enrolled in the two sections of the introductory biology course attended the workshops. The participants primarily came from the group of students who earned a B or higher on the first exam. Students who attended the optional workshop scored statistically significantly higher on the second exam than those who chose not to attend ($p < .001$). Students who attended the workshop were among the highest performers on the first exam so the improved performance relative to those students who chose not to attend is not unexpected. The students who attended the workshop earned, on average, one-half letter grade higher on the second exam, which represented both statistical and practical significance.

At the end of the semester, students were asked to complete a survey regarding their perception of studying and how they modified their academic success strategies following the workshop and throughout the semester (Nordell, 2009). Of particular interest was the change in behavior noted by students with regards to their preparation for class. A majority of students changed the way they prepared for class and the note taking strategies they employed (56.2%). As the majority of students who attended the workshop were among the highest performing on the exam, further investigation into how to help those students who fared poorly on the exam but may need to improve their metacognition may be valuable.

The workshops presented by Nordell (2009) provided insights into methods for improving metacognition in students. One strength of the workshops was the demonstration for students of how poorly they explain topics they believe they know very well. For example, Nordell (2009) asked students to describe the Mona Lisa in as much detail as possible. Despite their belief that they could describe the painting in great detail, students struggled to give more than cursory details. The first step to making workshops valuable to students is to assuage their beliefs in the strengths of their customary practices and help them to see they need to make changes. After discussion relating the exercise to their study habits, students recognized that they could frequently identify where in their notes the information was located but were unable to visualize the details needed to answer questions. During the presentations students were shown how a skill they learned through active engagement and practice, such as learning the multiplication tables, became well understood. Finally, students were shown how little they recalled from a recent lesson in the introductory biology course. The object lesson of the shallow nature of their understanding of concepts they believe they know, paired with the reminder of how they learned something they have retained for years, is a strong methodology for providing students with an impetus to change their study strategies.

Once students were prepared to listen and absorb new ideas, they were presented with new study strategies that encouraged metacognition (Nordell, 2009). One of the strategies students were shown was to prepare for class each day by reading the textbook and looking for new terminology. Students were instructed to write out the definitions, which provided the benefit of encouraging engagement with new terms prior to class and prepared them to better follow the lecture. Additionally, students were shown how active reading of the text prior to class could improve their ability to understand the text, help them create an outline of key concepts, and help them create connections between concepts. Students were told about the importance of taking thorough notes during class but were not taught strategies for improving the depth and completeness of their notes. After class, students were encouraged to reengage with the course content by using their textbook to review their class notes and correct any parts that were incomplete or incorrect. Finally, students were shown how to create concept maps in order to improve the depth of their understanding and their ability to connect concepts from within a lesson and between lessons.

In an attempt to determine the extent to which training students to better self-regulate their learning would improve their performance in a developmental mathematics class, Zimmerman et al. (2011) attempted to train students to use self-reflection to consider their errors and to respond to academic feedback. The participants in the study were 496 students from a large public technological college offering both associates and baccalaureate degrees. Students were randomly assigned to one of two groups, an experimental classroom, in which they were taught self-regulation techniques and their self-regulation was regularly assessed, or a control classroom, in which they were taught through traditional methods. Students were enrolled in one of six developmental sections or twelve introductory college-level classes. Three developmental and six transfer-level sections, which were taught by a total of four instructors, were designated as experimental classrooms and equal numbers of each, taught by nine different instructors, were designated as control classrooms.

Teaching in the experimental classrooms required instructors to collaborate throughout the semester (Zimmerman et al., 2011). During pre-semester meetings instructors were taught about self-regulated learning and how to effectively implement self-reflection worksheets to help students learn from in-depth correction of errors. A primary focus of this study was to teach students how to use the feedback they received through identified errors to change their approach to problems and to have students practice discussing their use of mathematical strategies or solution methodologies. Students took a quiz during every second or third class so they could receive consistent feedback and could be more practiced at diagnosing their errors and making corrections. Following each assessment, students were able to complete a self-reflection requiring them to examine why they made mistakes and how they might change their thought process to avoid similar mistakes in the future. Rather than only correcting the problem from the assessment, students were also required to solve a similar problem and were encouraged to seek help if they could not correct their errors.

Instructors were responsible for teaching self-regulation techniques during class and investigators observed each class session to determine the extent to which strategies were being taught in the class (Zimmerman et al., 2011). Students in all sections were given three tests during the semester and were asked prior to answering each question to rate their confidence, on a 5-point scale, about correctly answering the question. After answering each question, students were asked to use the same 5-point scale to rank their confidence that they had answered the question correctly. In addition to taking common periodic tests, students took a common cumulative final exam at the end of the semester. Students were placed in developmental classes because they had not reached a benchmark score on the COMPASS test. After the completion of the developmental course, students had to retake the COMPASS test in an attempt to meet the benchmark score and be permitted to move into a college-level course.

Students in the self-regulation classes saw an increase in their academic performance (Zimmerman et al., 2011). For those students in the developmental course, students in the self-regulation sections scored statistically significantly higher than the control group on the three periodic exams, as determined by a MANCOVA. Post-hoc testing showed no statistically significant difference between the experiment and control groups on the first exam, but that students in the self-regulation sections scored statistically significantly higher on the second exam ($p = .03$, Cohen's $d = .39$), the third exam ($p = .03$, Cohen's $d = .55$), and the final exam ($p = .02$, Cohen's $d = .50$). As with the students in the developmental class, a MANCOVA demonstrated that students in the introductory college-level course experiment group scored statistically significantly better on the exams than those in the traditional sections. Post-hoc univariate F tests showed statistically significantly higher exam scores for those in the self-regulation sections on the first exam ($p = .01$, Cohen's $d = .39$), second exam ($p = .01$, Cohen's $d = .39$), third exam ($p = .01$, Cohen's $d = .39$), and the final exam ($p = .01$, Cohen's $d = .47$).

One difference between the approach to teaching in the experiment and control groups was the implementation of frequent quizzing and subsequent error analysis conducted by students (Zimmerman et al., 2011). The researchers hypothesized that the increase in performance and pass rate may have been due to the regular quizzes and a testing effect. In consideration of this possibility, the researchers examined the self-regulation form completed by students in the experiment group and compared those students who completed more corrections than the median number to those who completed fewer. This examination showed a statistically significant difference in performance, $F(3, 57) = 3.79$, $p < .02$, $\eta^2 = .17$, which indicated that the self-regulation activities were the reason for the difference rather than the use of frequent quizzing.

In addition to students in the self-regulation classes outperforming those in the traditional sections, there were differences seen in the percentage of students who successfully completed each course (Zimmerman et al., 2011). Moving students out of the developmental sequence and into college-level coursework is essential for assuring that they are moving along a path towards degree completion. Students in the self-regulation classes successfully completed the developmental course at a statistically significantly higher rate than students in the traditional section (68% vs 49%), $\chi^2(1, N = 135) = 5.24$, $p < .05$, $\varphi = .19$. Students in the developmental course were required to retake the COMPASS test, and those in the self-regulation class passed at a statistically significantly greater rate than those in the conventional sections (64% vs 39%), $\chi^2(1, N = 135) = 8.13$, $p < .01$, $\varphi = .15$. Similarly, students in the college-level course who were enrolled in the self-regulation sections passed the class at a rate of 76%, which was determined to be a statistically significantly higher rate than the 62% pass rate for students in the traditional section, $\chi^2(1, N = 135) = 7.70$, $p < .01$, $\varphi = .24$. Providing students with a mechanism for improving their metacognition, through a process of self-reflection, definitively impacted pass rates for both courses.

There were results from the experiment that were contrary to the hypotheses of the researchers, who believed that students who participated in the experimental group would have higher scores on the measures of self-regulation and self-efficacy. Surprisingly, there was no statistically significant difference between how students in the experiment and control groups scored on either measure (Zimmerman et al., 2011). The researchers attributed this result to low- performing students consistently overestimating their belief in themselves. In order to determine whether the students consistently overestimated their abilities due to a personal bias, the authors made corrections to self-efficacy and self-regulation scores based on whether answers were correct or incorrect and then tested to determine whether there were statistically significant differences in overconfidence between students in the experiment and control groups. Students in the developmental self-regulation classes demonstrated statistically significantly less overconfidence in their measures of self-efficacy prior to answering questions, $F(1, 134) = 6.65$, $p < .01$, Cohen's $d = -.52$, and less overconfidence in their beliefs after answering questions, $F(1, 134) = 16.08$, $p < .01$, Cohen's $d = -.69$. Similarly, students in the introductory experimental group showed statistically significantly less overestimation of their ability prior to answering questions, $F(1, 340) = 13.71$, $p < .01$, Cohen's $d = -.39$, and in the belief of the correctness of their responses, $F(1, 344) = 16.64$, $p < .01$, Cohen's $d = -.43$.

Although students in the experimental group did not demonstrate statistically significantly higher levels of self-regulation, there were relationships evident between levels of self-regulation and academic performance (Zimmerman et al., 2011). For students in all developmental sections there were statistically significant correlations between self-efficacy mean scores and mean scores on the course exams (r = .50, p < .001, r^2 = .25) and on the final exam (r = .41, p < .001, r^2 = .168). There was also a statistically significant correlation between self-evaluation scores and scores on the periodic exams (r = .49, p < .001, r^2 = .240) and on the final exam (r = .38, p < .001, r^2 = .144). For students in the introductory course, the same statistically significant relationships were evident between self-efficacy scores and mean scores on the periodic exams (r = .39, p < .001, r^2 = .152) and on the final exam (r = .31, p < .001, r^2 = .096) as well as between self-evaluation scores and mean scores on the periodic (r = .39, p < .01, r^2 = .152) and final (r = .34, p < .01, r^2 = .116) exams.

In considering how the explicit teaching of metacognition strategies impacted the academic performance of undergraduate anatomy and physiology students, Langdon et al. (2019) introduced opportunities for reflection and collaborative learning to a large course traditionally taught through lecture. The researchers hypothesized that the introduction of teaching that addressed metacognition would result in improved levels of understanding, use of learning strategies, and academic performance. Participants were selected from three sections of anatomy and physiology, which had a total enrollment of approximately 450 students. Each section was randomly assigned one of three interventions: (a) reflection practice group, (b) passive knowledge reflection group, or (c) collaborative learning group. Only those students who successfully completed the intervention and answered all questions of both surveys were considered in the final evaluation of data, which led to the final group of 129 participants.

The researchers gathered information from the students using a survey during the second week of the semester and a second survey, which was given following the second exam (Langdon et al., 2019). The survey consisted of questions intended to gather demographic data and the Metacognition Awareness Inventory (MAI), which is a 52-item questionnaire intended to assess students' understanding and regulation of cognition. The MAI survey consisted of statements about metacognitive strategies with which students either agreed or disagreed. In addition to the two surveys, grades on the first and second exams and the final exam were collected. Each of the tests were different and covered material from different portions of the course.

Each of the interventions were conducted on the second class meeting after the first exam and each intervention session lasted 50 minutes (Langdon et al., 2019). Students in the section assigned to the reflection practice group were provided with questions asking them to reflect on their preparation for the first exam. The questionnaires were collected by the instructor but were not graded or evaluated for the purpose of this study. Students in the passive knowledge acquisition group watched five videos designed to address concepts of metacognition and to inform students how to get the most out of studying. Participants in the collaborative learning group were divided into groups of three or four and were given an assignment relevant to the topics of the course and were asked to complete the assignment as a group. Following the activity, one member of each group was randomly selected to take a low-stakes quiz, with the grade being assigned to all members of the group.

The questions of the MAI were disaggregated based on the type of metacognition each question addressed: procedural knowledge, declarative knowledge, or conditional knowledge (Langdon et al., 2019). Results of the preliminary survey, given before the intervention, were compared to the results of the survey given after the second exam. With regards to knowledge of cognition, there was a statistically significant interaction between time and group, $F(2, 126) = 5.17$, $p = .0007$, $\eta^2 = .076$, with students in the reflection practice group seeing a statistically significant increase in their mean scores, students in the passive acquisition group seeing a statistically significant decrease in mean scores, and the students in the collaborative learning group seeing no statistically significant change in mean scores. There were no statistically significant changes in grades between exam 1 and exam 2 and no statistically significant difference in the distribution of final grades between the three intervention groups. There was a weak correlation between regulation of cognition at the time of the second test and scores on the second exam ($r = .28$, $p = .01$, $r^2 = .078$).

The researchers hypothesized that the impact of the interventions was not statistically significant for several reasons (Langdon et al., 2019). First, the participants were in highly competitive majors and had high self-reported GPAs. Second, the implementation of the interventions in one class session may not have provided students with enough opportunity to effectively engage with the content. Students in the passive acquisition group watched the five videos sequentially with no time for questions or follow-up discussion. Third, the content addressed in exam 1 covered introductory material, which may have inflated the exam 1 grades, whereas exam 2 covered more complicated content. The authors did not indicate what historical data may be available regarding student performance on these exams and which may provide greater insight into the relative differences in scores. Students who engaged in exam wrapping, which is the intervention that asked students to consider their cognition most overtly, did demonstrate statistically significant increases in all types of cognition. This increase may indicate that students would benefit from more explicit training in cognition and study strategies.

Students in all three groups failed to see any statistically significant changes in the regulation of their cognition (Langdon et al., 2019). The authors hypothesized this may be due to the increasingly high percentage of freshmen taking the course for the first time. As the proportion of young students who have not been introduced to metacognition previously increases, the need for more explicit instruction becomes more important. The subtlety with which metacognition was introduced in this study may have led to the lack of change in regulation strategies. Additionally, introducing the interventions in a single 50-minute timeframe may have been overwhelming for young students who would have benefitted from the same content broken up over several class sessions. The researchers indicated that future research should be conducted to determine the benefits of blending the different interventions, rather than only showing one to each group, and should be introduced in shorter, more frequent time periods. Providing students with an opportunity to build on the active engagement found in the reflection practice intervention and the modeling found in the passive acquisition intervention may increase the impact to student performance.

Active Learning

Definition

Traditional lectures are passive events for students, who sit and listen while the professor provides information (Mesa et al., 2014; Riley & Ward, 2017; Wiggins et al., 2017). Because of the benefits of active learning, there is a desire among those who oversee higher education to increase the amount of active learning in the classroom (Conference Board of the Mathematical Sciences [CBMS], 2016; Freeman et al., 2014). The CBMS (2016) defines active learning as "classroom practices that engage students in activities, such as reading, writing, discussion, or problem solving, that promote higher-order thinking" (para. 1). More simply, active learning is defined by Bonwell and Eisen as "anything that involves students doing things and thinking about the things they are doing" (as cited in Riley & Ward, 2017, p. 1).

In a meta-analysis of 225 studies of active learning in undergraduate science, technology, engineering, and mathematics classrooms, Freeman et al. (2014) determined active learning increased student performance on assessments by approximately 6%. Students taking courses that were taught using lecture were 1.5 times more likely to fail than students in classes using active learning techniques (Freeman et al., 2014). Courses with active learning had failure rates of 21.8% whereas courses using passive learning techniques had failure rates of 33.8%. This is an increase in failure rates of 55% for courses in which the primary instructional method is lecture. The researchers found that the benefits of active learning were apparent for all STEM disciplines and courses, whether the students were STEM majors or non-majors, or whether the courses were introductory or upper division courses.

Active engagement during class is more beneficial to students than sitting passively during lecture and listening. To determine the impact of this learning method on student performance, Riley and Ward (2017) compared performance of students enrolled in an undergraduate accounting information systems course on an examination when they learned the material through either a passive lecture, an individual active engagement, or a cooperative active learning condition. The research study was conducted over three semesters. The five sections of the course were taught by the same instructor to reduce the impact of confounding variables such as teaching style or course policies on student performance. Each semester, the sections taught were assigned one of three learning conditions for the lesson covering four risk and control frameworks that are commonly studied in accounting and information security and technology. The three learning conditions were: (a) passive learning through lecture, (b) active individual learning, and (c) active cooperative learning. The first semester, a total of 67 students in two sections of the course were assigned to learn the material through the active individual learning condition. The second semester, a total of 47 students in one section were assigned to learn the material using the active cooperative learning condition. The third semester, a total of 77 students in two sections learned the material through lecture. The material being taught for the research study was covered two weeks prior to the final exam and was assessed through the answering of nine multiple-choice questions on the final exam.

Students in each learning condition were presented with a table to be completed during class time (Riley & Ward, 2017). Those in the passive lecture sections heard the material during lecture and were provided with the completed table at the end of class. Those students in the active sections were provided with a blank table and were told to spend the remainder of class time researching the frameworks and completing the table, either alone for those in the sections assigned to work as individuals or in groups of no more than three students for those assigned to the collaborative condition. Students in the active learning conditions were presented with the opportunity to complete a voluntary, anonymous questionnaire for which they would earn five bonus points.

Students in the active learning conditions outperformed those in the passive condition (Riley & Ward, 2017). Students in the individual active learning condition performed statistically significantly better than those in the passive lecture condition, $t(76) = -3.77$, $p < .001$. Scores of the students in the individual active learning condition ($M = 74.79\%$) were higher than those in the collaborative active learning condition ($M = 69.0\%$), although the difference was not statistically significant, $t(66) = 1.59$, $p = .11$. Similarly, students in the collaborative active learning condition ($M = 69.0\%$) outscored those in the passive lecture condition ($M = 63.0\%$), but the difference was not statistically significant, $t(46) = -1.53$, $p = .13$. Although the differences between the collaborative active learning and passive learning conditions was not statistically significant, there was a difference in means between the two groups of approximately six points, which is greater than half of one letter grade, and may provide students with a grade increase that improves their academic performance.

Students in the two active learning condition groups were asked to complete a voluntary questionnaire to provide some qualitative data regarding their experience (Riley & Ward, 2017). Students in both active learning condition groups commented that they found the assignment helpful in seeing the interconnected nature of the concepts. Of those students who completed the questionnaire, statistically significantly more students in the individual condition, $\chi^2(2, N = 66) = 28.09$, $p < .001$, and the collaborative condition, $\chi^2(2, N = 46) = 18.57, p < .001$, viewed the assignment in a favorable light compared to those who listened to the traditional lecture. Students in the collaborative active learning condition commented that they felt the group nature of the assignment was a hindrance in that they divided the work and felt most confident with the portion of the assignment they completed for themselves, but not that completed by the other members of their group. A statistically significant number of students in both the individual, $\chi^2(2, N = 65) = 76.92, p < .001$, and collaborative, $\chi^2(2, N = 47) = 32.21, p < .001$, conditions felt the assignment positively impacted their understanding and preparation for the exam. Finally, students in both groups, $\chi^2_{ind}(2, N = 66) = 45.36, p < .001$ and χ^2_{coll} $(2, N = 46) = 34.22, p < .001$, stated the assignment was an effective way to learn the material.

ICAP framework

Classroom activities have historically been classified as either passive or active learning engagements (Menekse et al., 2013). Chi (2009) attempted to divide learning activities that could be considered to be active into three subcategories of active, constructive, or interactive based on the observable manner in which learning occurred. The ICAP theoretical framework (See Figure 1 for a visual representation of the ICAP framework) states that as the characteristics of an activity progresses from passive to active to constructive to interactive, students become more engaged with the content, and learning improves (Chi & Wylie, 2014). Different behaviors require students to interact with content at different levels, which results in differences in the amount and type of learning that occurs (Chi & Wylie, 2014).

Chi (2009) defined the three categories of active learning based on the type of overt interaction students undertake when engaging with the material. Active learning is characterized by focused movement or behavior, such as taking verbatim notes, highlighting text in a passage, or copying solution steps. Constructive learning requires the generation of output that extends what was provided in the lesson, which can be demonstrated through explaining how to solve a problem, generating questions that connect new content with old, or developing concept maps. Interactive learning involves the exchange of ideas between multiple people, each of whom are engaging with material at a constructive level. Developing questions about lesson content and then answering them in pairs or in small groups is an example of interactive engagement. One qualifier for the classification of activities is that determinations of the class of activity must be made based on the observable, or overt, elements of the engagement (Chi, 2009). Although it is possible students are thinking about material using a higher level of engagement, classification of activities that cannot be observed is not possible.

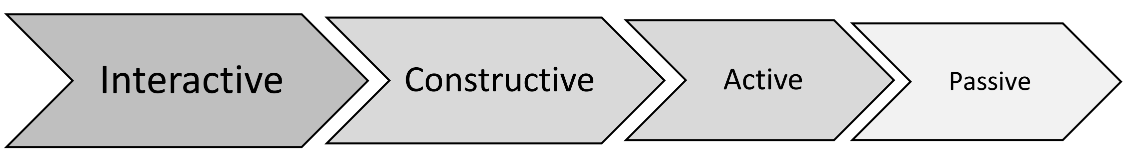

Figure 1. Illustration of the ICAP theoretical framework.

To demonstrate the relevance of the ICAP theoretical framework for classroom learning, Chi and Wylie (2014) examined multiple studies that investigated learning activities designed to help students with note taking, concept mapping, or self-explaining. The authors considered the results of each study and used the ICAP framework to provide a rationale for the differences in learning noted in each. Within the studies examined, the

ICAP framework effectively predicted which activity would lead to higher performance based on the type of engagement undertaken by students. Using the ICAP framework, Chi and Wylie (2014) were able to explain why one activity improved student performance more than another and were able to provide a reason why there was no statistically significant difference in performance in other studies.

Beyond allowing for differentiation of activities and benefits in instructional design that could be an effective use of the ICAP framework, Chi and Wylie (2014) hypothesized that the framework may have further usefulness as a diagnostic tool. When considering studies that compared two learning activities, the researchers were able to use the ICAP framework to describe why the expected results of the researchers were not found, or why results were not as profound as expected. For example, the ICAP framework can be used to explain why the comparison of an active treatment engagement with a passive control did not show as profound of an effect as the comparison of a constructive treatment engagement with a passive control (Chi & Wylie, 2014). The ICAP framework can also be used to explain why no statistically significant difference is found between two different engagements that are measured at the same level of active learning. Thus, care should be taken to consider the category of a treatment activity, the category of the control activity, and whether students were able to carry out both activities according to the instructions.

Chi and Wylie (2014) examined studies that compared learning activities in which engagements were classified across two, three, or four levels of the ICAP theoretical framework. In an attempt to validate the ICAP framework, the authors considered studies that compared activities across all four levels of engagement, two studies that compared three of the levels of engagement, and multiple studies addressing pairwise comparisons of engagements within the topics of note taking, concept mapping, and self-explaining. In the study by Menekse et al. (2013), in which one experiment compared all four levels of engagement and one compared active, constructive, and interactive activities, the ICAP framework was supported and learning improved for each successive level. The investigation of studies comparing two engagements with regards to notetaking also confirmed the ICAP framework (Chi & Wylie, 2014). The hierarchical comparisons confirmed the author's expectations and, interestingly, two studies comparing student performance following the completion of different active engagements demonstrated no statistically significant differences, which further confirmed the diagnostic legitimacy of the ICAP framework.

Menekse et al. (2013) conducted two studies to test the validity of the ICAP framework through the comparison of learning engagements that were passive, active, constructive, and interactive. In the first study, the researchers tested the ICAP hypothesis in an engineering classroom. The authors chose not to include a passive activity because these were students in a real classroom, choosing instead to compare the results of active, constructive, and interactive engagements. In the second study, the authors recruited engineering students to participate in a laboratory setting in which all four levels of engagement were tested.

In testing the ICAP hypothesis with students in a real classroom setting, students in an undergraduate engineering course volunteered to stay after class for 20 minutes on five different days during the semester (Menekse et al., 2013). Because participation was voluntary, students' grades were not impacted by their participation. To determine how learning historically occurred in the introductory materials science course, one of the researchers attended every class meeting in the semester prior to the research study. Nineteen student engagements were observed and were classified as active, constructive, or interactive based on the classification system developed by Chi (2009). The researchers conducted the experiment in two phases. In the first phase, students spent time in one class session completing an active engagement and spent time during a second class session completing an interactive activity. The topics covered in each phase were from the same unit of the course. In the second phase, students spent one class completing activities categorized at each level of engagement: active, constructive, and interactive. To ensure that activities were properly conducted at the interactive level of engagement, the researchers developed prompts for the student group leaders to use.

The researchers expressed concern that learning from the completion of homework would confound the results of assessment (Menekse et al., 2013). To prevent this, student learning was assessed immediately after class on the day lessons were taught. After class each day, students were given a quiz consisting of two multiple choice and two open-ended questions designed to determine how much they had learned and understood the material. The open-ended questions were scored using a rubric and a 4-point scale. Quiz questions were further classified as verbatim, integration, or inference questions depending on the depth of understanding being assessed by each question. In order to compare the results using ANOVA testing, and because each quiz was based on different content, the two days of Phase 1 of the study were compared and the three days of Phase 2 were compared. There was no comparison conducted for content between the two phases.

The results of this study showed statistically significant differences in learning (Menekse et al., 2013). A one-way ANOVA of the Phase 1 results showed statistically significantly greater learning occurred following interactive engagement when compared to active learning, $F(1, 38) = 28.69$, $p < .01$, multivariate $\eta^2 = .43$. The amount of learning that occurred following interactive engagements was statistically significantly improved relative to the learning that occurred following an active engagement as measured by verbatim questions, $F(1, 38) = 11.40$, $p < .01$, multivariate $\eta^2 = .23$, integration questions, $F(1, 38) = 6.02$, $p < .05$, multivariate $\eta^2 = .14$, and inference questions, $F(1, 38) = 23.57$, $p < .01$, multivariate $\eta^2 = .38$.

A one-way ANOVA of the Phase 2 results also showed a statistically significant impact based on the type of learning engagement, $F(2, 34) = 5.40$, $p < .01$, multivariate $\eta^2 = .24$ (Menekse et al., 2013). Post-hoc testing showed statistically significant differences between active and constructive learning ($p < .05$, Cohen's $d = .63$) and between active and interactive learning ($p < .05$, Cohen's $d = .48$). There was no statistically significant difference between constructive and interactive learning. Three separate one-way ANOVAs were conducted and showed statistically significant results for verbatim questions, $F(2, 34) = 8.65$, $p < .01$, multivariate $\eta^2 = .34$, integration questions, $F(2, 34) = 8.55$, $p < .01$, multivariate $\eta^2 = .34$, and inference questions, $F(2, 34) = 5.28$, $p < .01$, multivariate $\eta^2 = .24$. Post-hoc Tukey testing showed the constructive learning produced statistically significantly better results than the active learning on both verbatim and integration questions, and statistically significantly better results than interactive learning for integration questions. For the inference questions, interactive learning produced statistically significantly better results than either active or constructive learning. With the exception of the result from Phase 2 integration questions, the result of this study confirmed the ICAP theoretical hypothesis.

In a second related study, the researchers moved out of the classroom and into the lab, which allowed for better control of confounding variables (Menekse et al., 2013). The participants in the second study were 120 undergraduate engineering students who were randomly placed into one of four groups to test the effectiveness of the four levels of learning engagements. Participants were given both a pretest and posttest, each of which consisted of 15 true-false questions, seven multiple-choice questions, and two open-ended questions. The posttest included one additional true-false question, four additional multiple-choice, and one additional open-ended question. The content for the lesson was the same for each of the four groups. The passive group read an 8-page lesson out loud but were not being allowed to take notes or highlight any content. The active group read the same passage and were instructed to highlight the most critical ideas. The constructive group did not read the passage but studied graphs and figures the summarized the same content, and then provided written answers to questions addressing the content. Students in the interactive group considered the same graphs and figures as those in the constructive group but worked in pairs and worked collaboratively to complete the questions.

Students were randomly assigned to one of the four groups, with 24 students in each of the passive, active, and constructive groups, and 24 pairs of students in the interactive group (Menekse et al., 2013). Each participant was shown the same 2-page introduction written by the researchers and then completed the pretest, the learning activity assigned to their group, and the posttest. The total time commitment for participants was 90 minutes. True-false and multiple-choice questions were marked as either correct or incorrect. The open-ended questions were scored using a rubric and were marked as fully correct, partially correct, or incorrect.

Students' scores on the pretest and the gain in scores from the pretest to the posttest were compared between the four groups. The researchers conducted a one-way ANOVA that revealed no statistically significant difference between groups on the pretest (Menekse et al., 2013). Because of the change in the maximum score between the pretest and posttest, as a result of the additional questions, the researchers considered the percentage of gain when looking at posttest results. A one-way ANOVA showed a statistically significant result, $F(3, 116) = 25.34$, $p < .001$, $\eta^2 = .40$, across all conditions and post-hoc testing using a Bonferroni correction showed statistically significant results between all of the individual engagements comparing a level of engagement against the levels considered to be a lower level of engagement. Interactive engagements showed statistically significant improvement in gain scores over passive ($p < .001$, Cohen's $d = 1.88$), active ($p < .001$, Cohen's $d = 1.42$), and constructive ($p = .003$, Cohen's $d = 0.64$) engagements. Constructive learning demonstrated statistically significant improvements in gain scores over passive ($p < .001$, Cohen's $d = 1.04$) and active ($p = .035$, Cohen's $d = 0.54$) activities. Active learning showed statistically significant improvements in gain scores over passive ($p = .023$, Cohen's $d = 0.65$) learning. The researchers were able to confirm the hypothesis that student performance would increase more as learning moved from passive to active to constructive to interactive levels of engagement. Interestingly, students in the constructive and interactive engagements, who were never presented with the core content provided to students in the passive and active groups, demonstrated greater learning through the development of their own understanding through answering guided questions.

There are a limited number of studies investigating the extent to which interactive engagements lead to improved performance when compared to constructive activities. In an attempt to confirm the ICAP framework, Wiggins et al. (2017) investigated the difference in student learning associated with interactive and constructive learning in a STEM classroom. Participants in this quasi-experiment were 759 students in an introductory biology class at a public university in the United States. Students self-selected into one of two large-lecture sections of the same course taught by the same instructor, which allowed for a relatively controlled environment in which instructional topics, pedagogical methods, pacing, and assessment were relatively uniform between the sections.

Activities were selected based around topics that are common in introductory biology curricula and could be created to be classified as either interactive or constructive using the ICAP classification criteria (Wiggins et al., 2017). On four different dates during the semester, one section of students completed an interactive engagement and the other section completed a constructive activity. Each section completed two interactive and two constructive activities. The researchers were careful to create activities that provided students with the same content and that differences in the two sections were limited to the script used by the instructor, which presented the instructions for the activity, and the structuring of the group activities. Both engagements used group work to protect against the potential confounds that may have occurred if only the interactive activities utilized group work.

The design of the research activities and the development of assessment tools was conducted over several semesters, which provided the researchers an opportunity to refine activities based on results and student feedback prior to conducting the experiment (Wiggins et al., 2017). Student understanding was measured using a pretest and posttest design with multiple-choice questions intended to determine student's understanding through the asking of questions requiring higher-order cognitive skills as ranked by Bloom's taxonomy. In-class observations were conducted at two times during each experiment day in order to determine the extent to which the protocols for the assignments were followed and data were gathered to quantify the number of student-to-student interactions occurring during each observation. Investigators were particularly concerned that engagements between students were interactive, which requires both students to be interacting at a constructive level of activity. Data from one of the four quasi-experiment dates needed to be removed from the study because errors in the established procedures resulted in student confusion.

The data indicated that the volume of student interaction and student performance were increased by the interactive engagements (Wiggins et al., 2017). During observations of student group work, researchers determined that statistically significantly more student interaction occurred during the interactive engagement (observation 1, $p < .001$; observation 2, $p = .01$). Students were 25% more likely to correctly answer at least one more question on the posttest following the interactive engagement than after the constructive activities ($p = .01$). Further, there was no statistically significant difference in performance for any of the demographic groups investigated, which indicated that benefits can be seen for all students regardless of socioeconomic status, gender, or race.

The authors recognized the subtle impact on student performance evidenced with interactive learning and acknowledged the difficulty in designing and implementing successful interactive engagements. The intermittent use of interactive engagements in this course may have contributed to the small change seen in posttest results. The researchers indicated the benefits of interactive engagement may be cumulative and would result in more definitive improvements in student performance if consistent interactive learning was implemented throughout the semester.

Note Taking as an Active Learning Strategy

Based on the definition of active learning, note taking and the review and editing of class notes should be considered to be active endeavors (Chi & Wylie, 2014). Because students do not transcribe a lesson in their class notes, but must actively listen, filter the lecture for the most important content, and quickly summarize the content in their own words, note taking is an activity that engages students in higher level thought (Al-Musalli, 2015; Boch & Piolat, 2005). In recognition of the incomplete and often incorrect nature of class notes, students need to review and edit their notes following a lesson and prior to preparing for an exam. Students should be encouraged to engage with their class notes in a manner that is active, constructive, or interactive based on the ICAP framework. If engagement with note taking material can be elicited from students through the creation of more active interventions prior to exams, there should be a benefit to student learning.

Many developmental students do not appreciate the necessity of proper note taking during class and, as a result, have a tendency to not take comprehensive notes (Eades & Moore, 2007). This leaves them with little or no resources to utilize after class for making connections between the ideas making up a course. As was demonstrated in the literature, most college students capture less than half of key ideas during a lecture. Faculty-provided skeletal guided notes can increase the quality of class notes but students still miss concepts and produce notes with inaccuracies. The provision of complete instructor-created notes can give students a resource for studying, but the passive nature of reading notes created by someone else may not be effective. Through the addition of active interventions, particularly those classified as constructive or interactive using the ICAP framework, students can learn to review, revise, and restructure their class notes to improve both the quality of their notes and the impact of their notes on performance (Chi & Wylie, 2014). By requiring students to revise and restructure their notes, students can be taught to create a valuable tool capable of furthering their development as students.

Interventions considered in this review of literature are time consuming and require a great deal of effort from instructors and students. Interventions that have been shown to improve student performance are more likely to be adopted by both instructors and students, as the value of the time and effort can be seen in practical terms (Hudesman et al., 2014). In developmental classes, one goal of instructors should be to teach students the skills necessary to be successful in future courses. Providing students with a strategy for improving the quality of their class notes, which allows them to have a higher quality resource for studying and preparing for exams, is an important aspect of ensuring that developmental students are prepared for future classes and equips them with strategies they can use in other courses (Eades & Moore, 2007). Those interventions that have demonstrable impact on student performance, and those that students are able to replicate in future courses, have the best likelihood of being adopted by students (Hudesman et al., 2014). Teaching students the value of taking, reviewing, and interacting with notes daily is an invaluable skill which will serve students as they progress through their college careers.

Chapter III:

Method

Developmental college mathematics students often struggle to take notes in class, and then are unsure how to utilize their notes in a manner that effectively prepares them for tests (Bjork et al., 2013; Eades & Moore, 2007; Van der Stel et al., 2010). Developmental students may not understand the importance of reviewing their notes for completion and correctness promptly after class to ensure they have an effective resource to use when studying (Bjork et al., 2013; Cafarella, 2014). Many students choose to review their notes by reading them which is, at best, a passive engagement based on the ICAP framework (Bjork et al., 2013; King, 1992). This passive approach to review does not help students to learn concepts with which they struggle. Developmental college students need to be provided with a technique for improving the quality of their class notes and eliciting a more active and constructive tool for studying, which will improve their academic performance.

Purpose

The purpose of this quasi-experimental study was to investigate the relationship between the completion of a note restructuring intervention and performance on course exams. The study was conducted with students in various levels of developmental mathematics courses taught at a northern California community college. The intervention utilized aspects of active and constructive learning engagement, as defined by Chi and Wylie (2014) in the ICAP framework. The independent variable was defined as scores on a note review and restructuring intervention completed by students and turned in for grading on the day of the exam. The dependent variable was defined as performance on course exams covering the course content students examined in the intervention. Participation in the intervention was incentivized through the assignment of points which contributed to students' grades in the course. Students were free to complete the note restructuring activity to the degree they chose, including choosing not to complete it.

Research Questions

The following research questions were addressed in this study:

1. To what extent does a student's performance on each journal intervention impact their performance on the corresponding exam?
2. To what extent is the pattern of student performance on the journal intervention associated with exam performance?

Research Design

This study used a one group, repeated treatment, posttest only quasi-experimental design. This design is characterized by: (a) the lack of randomly assigned participants, (b) the lack of a pretest, and (c) the introduction, removal, and reintroduction of a treatment over time (Shadish et al., 2002). This design allowed the researcher to investigate the extent to which a student's performance on the note review and restructuring intervention impacted their performance on the associated exam. Additionally, the design allowed the researcher to determine to what extent the pattern of student participation in the intervention impacted exam performance. The intervention was referred to in class and in the syllabus as "The Journal" and required students to define course terminology, write out problem solution strategies, and complete problems that illustrated their methodologies.

Participants

Between the Fall 2012 and Fall 2015 semesters, the author taught multiple sections of four developmental mathematics courses: (a) Arithmetic, (b) Arithmetic and Pre-Algebra, (c) Algebra I, and (d) Algebra II. Participants were 390 students enrolled at a community college in Northern California who took developmental mathematics courses taught by the researcher. Students were eligible to register in one of the four courses as a result of passing the prerequisite course or having been placed at the level of the course based on their performance on a standardized placement test. Convenience sampling is a nonprobability sampling technique typically used when the subjects of the research are readily accessible and, rather than being randomly assigned, are used because of their availability or willingness to participate (Jager et al., 2017). Although it is not the optimal sampling technique, using a convenience sample allowed the author access to multiple groups of students for whom the instruction was consistent, the intervention could be assigned, and similar exams could be used each semester in each course. There is an assumption when using convenience sampling that the population is homogenous and there will be little difference in the research results between the convenience sample and a random sample (Jager et al., 2017). Because of the nonrandom selection of the sample, there may be limitations in the generalizability of the study, which increases the potential bias of the estimators used in drawing conclusions. The lack of generalizability can be mitigated by a thorough investigation of the characteristics of the convenience sample, which can help to better illustrate how the sample reflects the characteristics of the population (Jager et al., 2017; Wilkinson et al., 1999). Descriptive statistics of the sample participants across multiple variables will be provided.

Data Collection

Data were collected on the day of chapter exams, at which time students turned in the journal assignment corresponding to the chapters covered in the exam. Grading for the journal assignment and the chapter exams was not blind and was performed by the author. Journals were given grades from a minimum score of 0 points to a maximum score of 15 points according to a rubric that was distributed to students at the beginning of the semester (see Appendix A for an example of the journal rubric). Journal rubrics were adjusted, depending on the class, to reflect the number of submissions for the semester but the distribution of points was consistent across courses and semesters. Students were expected to turn in either five or six journals each semester, depending on the course. Because the researcher used a rubric and assigned all grades for the participants, there is an expectation that grading was consistent, particularly for all students within a course and between students in similarly leveled courses.

Exams were graded by the author according to an answer key and using a rubric designed by the author prior to giving the exam. Students were able to earn partial credit on exam problems based on criteria developed by the author for each problem. The content covered on each test was consistent across semesters for each course level (See Table 1 for test content summaries). After the first exam, students were presented with data regarding average exam scores for those students who earned perfect journal scores, those who received partial credit on their journal, and those who failed to turn in a journal.

Tests in each course got progressively more complicated as the semester progressed. In the Arithmetic course, the first three tests addressed basic arithmetic operations, with Test 1 covering operations with whole numbers, Test 2 covering operations with fractions and mixed numbers, and Test 3 addressing operations with decimal numbers. In Test 4 students were asked to solve application problems using ratio and proportions and in Test 5 students were asked to convert units.

Table 1

Test Content by Course Level

Test	Arithmetic	Arithmetic & Prealgebra	Algebra I	Algebra II
1	Operations with whole numbers	Arithmetic and algebra with whole numbers	Solving linear equations	Functions
2	Operations with fractions	Solving linear equations	Linear inequalities and graphing equations	Systems of equations and absolute value equations
3	Operations with decimals	Arithmetic and algebra-fractions and decimals	Systems of linear equations	Radical expressions and equations
4	Ratios and proportions	Ratios and proportions	Rules of exponents and polynomials	Solving quadratic equations
5	Unit conversions	Geometry and unit conversions	Factoring and solving quadratic equations	Exponential and logarithmic functions
6	--	--	Rational expressions and equations	Conic sections

Note. Students in the Arithmetic and Arithmetic & Prealgebra class were only given five tests.

The Arithmetic and Prealgebra course combined semester-long courses in arithmetic and prealgebra into a single-semester accelerated course. On Test 1, students were asked to simplify arithmetic expressions with whole numbers and then to solve simple linear equations with whole number coefficients. Test 2 covered solving more complicated linear equations that involved using both the addition and multiplication properties of equality. On Test 3, students were asked to simplify arithmetic expressions involving fractions, mixed numbers, and decimal numbers and were also asked to apply those concepts to solving equations with fractional or decimal coefficients. Students solved application problems involving proportions and percentages on Test 4 and were asked to solve application problems using geometry and unit conversions on Test 5.

On Test 1 in the Algebra I course students were asked to simplify algebraic expressions and solve linear equations. For the second test, students solved linear inequalities and graphed linear equations. They were also required to calculate slope and write the equations of lines in both slope-intercept and point-slope form. On Test 3, students solved systems of linear equations with two equations and two unknowns. The fourth test required students to simplify expressions with whole number and negative exponents using the rules of exponents and served as an introduction to polynomials and operations with polynomials. Students were asked to factor polynomials and to use factoring to solve quadratic equations for Test 5. On Test 6, students had to simplify rational expressions, perform operations with rational expressions, and solve rational equations.

The Algebra II course was designed to build on the foundation of learning established in the Algebra I course. The first test in the Algebra II course covered functions and function notation, domain and range, and asked students to solve problems involving direct and inverse variation. The second test asked students to solve systems of three linear equations, to solve systems of inequalities, and to solve absolute value equations and inequalities. For the third test, students had to simplify expressions with fractional exponents and had to simplify radical expressions and solve radical equations. For Test 4 students had to solve quadratic equations using the quadratic formula and completing the square. Test 5 asked students to simplify exponential and logarithmic expressions and to solve exponential and logarithmic equations. The final test covered the evaluation and graphing of circles, ellipses, parabolas, and hyperbolas.

Students were provided detailed written instructions that defined the components of the journal and were provided lists of the minimum required elements for each chapter. The three components that made up the journal were the same for each class, but each student received guidelines specific to their class (see Appendix B for an example of journal guidelines). Students were encouraged to consider the required elements list as an outline of key concepts for the chapter (see Appendix C for a sample page from the journal summary).

The introduction of the journal project was done on the first day of class and was emphasized as a key learning strategy. Presenting the project to students in a manner that made sense and stressed the relationship between completing the journal and success on exams was an important part of the first day of class and of getting students to understand the benefits of successful completion of the journal engagement. Students were shown good and bad examples of journals created by students in previous semesters and were invited to ask questions or get feedback during class or office hours once they had started working on the journal. Additionally, students were given several reasons why the journal would benefit them in current and future courses. Students were told the act of writing the journal would solidify concepts in their minds and help them to reflect on the specific methodology necessary to complete problems. Students were encouraged to put definitions and concepts into their own words in order to improve the value the project had for themselves. Students were also reminded that developmental mathematics classes are sequential and concepts in one course are built on understandings form prior courses. The journal would provide a bridge between prior knowledge and future understandings, particularly for those concepts that are direct extensions of prior learned concepts.

Analytical Strategy

To determine to what extent students' performances on the journal intervention impacted their performance on the corresponding exam, two one-way ANOVAs were conducted for each of the six journals and the corresponding exams. The first ANOVA was conducted to determine the extent to which a student's performance on the journal intervention impacted their grade on the exam and the second ANOVA was used to determine the extent to which the class a student was enrolled in impacted their grade on the exam. When conducting a one-way ANOVA, the dependent variable must be continuous (Strunk & Mwavita, 2020). In this study, the dependent variable for each of the 12 ANOVAs was student scores on the course tests, which is a continuous variable.

ANOVAs are a group of tests used to compare two or more independent means (Field, 2018). The one-way ANOVA has assumptions of the normality of the residuals and homogeneity of variance (Field, 2018). The assumption of normality of the residuals was checked through a visual inspection of the Q-Q Plot and through an assessment of the standardized skewness and kurtosis ratios. Standardized skewness and kurtosis ratios less than three are indicative of approximate normality of the distribution (Field, 2018). Nonetheless, ANOVA have been shown to be relatively robust to violations of normality, particularly when sample sizes are large (Schmider et al., 2010). The assumption of homogeneity of variance was tested using Levene's test. If the assumption of homogeneity of variance is violated, the Welch test is more robust than the conventional ANOVA (Delacre et al., 2019; Lix et al., 1996).

To determine the extent to which completion of the journal intervention impacted student performance on exams, a mixed ANOVA was conducted. Mixed ANOVAs are commonly used in repeated measure studies and are characterized by the presence of at least one between-subjects categorical variable and at least one within-subjects continuous variable (Murrar & Brauer, 2018). As such, mixed ANOVAs allow for the exploration of main effects and interaction effects. In the present study, the between-subjects factor was the pattern of journal performance (A), and the within-subject factor was the performance on each of the five or six tests (B). The interaction effect (A x B) tested the extent to which test scores depended on journal performance patterns (e.g., a_1 vs. a_2 under b_1 compared to a_1 vs. a_2 under b_2).

Depending on the course, students were expected to turn in either five or six journals, which corresponded with either five- or six-chapter exams. Because not every class had six tests, only the first five exams were used for the mixed ANOVA. Determining the pattern of journal submission was an important consideration in determining the impact of the journal on exam performance. Although many students chose to complete all of the assigned journals, there were those who chose to turn in fewer journals than were assigned. Of particular interest were those students who turned in some complete journals and either failed to turn in or turned in the remainder of the journals as incomplete documents. To classify the patterns of journal completion, a latent profile analysis (LPA; Lazarsfeld & Henry, 1968) was conducted and served to identify the unseen classes of students whose journal submission pattern were similar. These classes were the between-subjects factor.

Conducting a mixed ANOVA requires that certain assumptions be met. The residuals of the between-subjects model and the within-subjects model must be normally distributed and across the between-subjects groups homogeneity of variance in the dependent variable must be displayed (Murrar & Brauer, 2018). Another assumption is that of sphericity, which applies to models with three or more within-subject levels and refers to the need for the variances of differences between each of the groups to be approximately equal. Sphericity is tested using Mauchly's sphericity test and, in the case the assumption is violated, can be corrected using the Greenhouse-Geisser or Huynh-Feldt correction (Murrar & Brauer, 2018). Both eta-squared and omega-squared were reported to describe practical significance of the test results.

The use of LPA in this study provided a person-centered approach to classifying students based on their general pattern of journal performance. To conduct LPA, data were analyzed using the statistical software package Mplus. LPA is particularly useful in determining relationships within a large, heterogeneous population in which relationships between the variables are not clearly visible or definable (Tein et al., 2013). LPA was used in this study to determine student subgroups based on the degree to which the journals were correctly completed by students over the course of the semester.

Latent profile analysis is a model-based version of cluster analysis that uses statistical software to identify clusters of participants with shared characteristics (Tein et al., 2013). The underlying assumption is that the sample is made up of participants from an unknown number of different latent classes and that participants in each latent class share characteristics with each other (Masyn, 2013). It is assumed that within a latent class the observations of a manifest variable are considered to come from the same probability distribution and that the variable is approximately normally distributed within each class (Tein et al., 2013). In this study, the manifest variable will be student performance on the journal interventions, which are hypothesized to be influenced by the participants' presence in one of the latent classes.

One benefit of LPA is the use of a person-centered approach rather than a variable-centered approach (Masyn, 2013). The person-centered approach focuses on the characteristics of the individual participants and may describe how a participant's presence in a latent class can provide insight into the different groups (Williams & Kibowski, 2016), whereas the variable-centered approach describes relationships among variables (Masyn, 2013). The person-centered approach is dependent on an assumption that the overall population is heterogeneous and is made up of a finite number of homogenous subgroups, which differs from a variable-centered approach, in which the population is assumed to be homogenous (Masyn, 2013). The value in using LPA is the ability to identify underlying relationships that may not have otherwise been apparent and which may explain some of the differences between the latent classes (Williams & Kibowski, 2016). In LPA, the latent variable responses of participants within classes are more similar than the responses between classes (Masyn, 2013).

In general, LPA is appropriate for samples of at least 100 participants and for continuously measured latent variables (Williams & Kibowski, 2016). Care should be taken to ensure the sample is large enough to demonstrate sufficient statistical power or there is the possibility of selecting too few or too many classes (Tein et al., 2013).

Because the goal of conducting LPA is to determine a finite number of distinct latent classes, it is essential that the correct number of classes is chosen (Tein et al., 2013).

Conclusion

Note taking is an important skill for college students to possess and develop. There is recognition within the literature that developmental students lack some of the study skills necessary to be successful college students (Cafarella, 2014; Eades & Moore, 2007; King, 1992). Students need to understand the impact of having incomplete notes as well as the importance of using their class notes to prepare for exams. Providing students with an intervention for improving the quality of their notes and requiring them to actively engage with their notes after class, may lead to improved performance on exams. For developmental mathematics students, who are beginning sequential courses with content that builds on previous courses, the inability to take effective notes is particularly impactful. An added benefit of providing students with an effective note taking intervention is the development of a resource students can use in future classes and a skill they can continue to develop.

Participants in this study were 390 students enrolled in one of four developmental mathematics classes at a community college in Northern California. Student participants were assigned to complete the journal note intervention and took exams covering material corresponding to each journal. Data were collected with regards to the score each student earned on the assigned journals and the score earned on the corresponding test. Depending on the course, students were assigned to complete either five or six journals and took either five or six exams. Scoring on the journal was done based on a rubric designed by the researcher, who was the author in all of the classes in which students were enrolled. Data were evaluated through the completion of a mixed ANOVA, with latent data classes determined by LPA, which is a technique used to identify unseen patterns within the data.

CHAPTER IV:

Results

Two different forms of ANOVA were used in the examination of the research questions. Research Question 1 considered to what extent a students' performance on each journal intervention impacted their performance on the corresponding exam. To answer this question, two one-way ANOVAs were conducted for each test. One ANOVA was conducted to investigate to what extent students' scores on the journal impacted their performance on the corresponding exam and the other was conducted to consider to what extent the class a student was enrolled in impacted their performance on the exam. Research Question 2 examined to what extent a students' pattern of performance on the journal impacted their test performance. Through the use of latent profile analysis (LPA), four different profiles of journal scoring pattern were identified. A mixed model ANOVA was conducted to determine to what extent students' journal profile resulted in differences in exam score.

ANOVAs are a family of statistical tests used to compare two or more independent means (Field, 2018). The one-way ANOVA has two assumptions that were addressed prior to conducting each test. The first assumption is normality of the residuals (Field, 2018), which was checked through a visual inspection of the Q-Q Plot and through calculation of the standardized skewness and kurtosis ratios. If the absolute value of the standardized ratios was less than three, then the distribution was determined to be approximately normal. The final assumption is that of homogeneity of variance (Field, 2018), which was addressed through Levene's test.

The mixed ANOVA is a statistical test incorporating both between-subject and within-subject designs (Strunk & Mwavita, 2020). Like traditional ANOVA that examines between-subject effects, the mixed ANOVA has an assumption of normality of the residuals and homogeneity of variance (Murrar & Brauer, 2018). Normality of the residuals was assessed through visual inspection of the Q-Q Plot and through calculation of the standardized skewness and kurtosis ratios. Homogeneity of variance was examined using Levene’s test, which was conducted to determine whether the variances were equal across profiles at the time each test was taken (Strunk & Mwavita, 2020). ANOVAs that involve within-subject designs additionally have an assumption of sphericity, which is the requirement that the variance between any pairs of levels of the within-subject design are approximately equal (Murrar & Brauer, 2018). Sphericity was tested using Mauchly’s sphericity test ($\alpha = .05$) and violations were addressed by correcting for the degrees of freedom using either the Greenhouse-Geisser or Huynh-Feldt corrections.

In tests comparing the means and variances of multiple samples, the presence of outliers in the data can be problematic. Outliers can alter both the shape of the distribution of the residuals and the variance of the groups, which may in turn impact the decisions made following hypotheses tests (Liao et al., 2016). Outliers can lead to greater skew in the distribution, which has been linked to increased Type I error rates and decreased power in the test (Lix, 1996). Liao et al. (2016) showed that the presence of two or more outliers in the data could result in an increase in Type I error. Although the incidence of Type I error was lessened with large sample sizes, the presence of outliers in the data still impacted the statistical decisions.

Mitigating the impact of outliers can be done through the use of transformations, removal of outliers, the use of a non-parametric rank test, or through an accommodation method such as Winsorizing or trimming. Transformations refer to the application of a mathematical function to each data value in an attempt to keep the outliers and to lessen the variance and skewness in the data (Aguinis et al., 2013). Interpretation can be made more difficult when using transformations because the conclusions drawn are not based on the original data, but are based on the transformed data instead (Lix et al., 1996). Removing outliers is possible but should be reserved for those values resulting from measurement or recording errors (Aguinis et al., 2013; Liao et al., 2016). Removal of outliers can result in the loss of useful information and should be avoided (Liao et al., 2016). When the presence of outliers results in non-normal distribution, it is possible to use a nonparametric test. The Kruskal-Wallis test is considered to be the nonparametric equivalent of ANOVA, but because the null hypothesis is different and the test is sensitive to variance heterogeneity, it may not be the best option (Lix et al., 1996). The Kruskal-Wallis test also does not effectively control for Type I error rate when there are three or more outliers in the data (Liao et al., 2016). Outlier accommodation techniques, such as Winsorizing (Liao et al., 2016) or trimming data (Cribbie et al., 2012) have been shown to successfully correct for increased Type I error rates in ANOVA. The use of robust estimators, such as Winsorized or trimmed means, have been demonstrated by several researchers to best correct for Type I error when normality is violated, regardless of whether the variances are equal or unequal (Cribbie et al., 2012, Delacre et al., 2019; Liao et al., 2016; Keselman et al., 2000).

Trimmed data diminish the impact of outliers because the standard error associated with extreme observations is muted (Keselman et al., 2000). In one-way ANOVAs, the use of trimmed data is more effective at controlling Type I error when skewness or kurtosis is present in the underlying data (Delacre et al., 2019). In mixed-model and repeated measure ANOVAs, Keselman et al. (2000) demonstrated that when normality, homogeneity of variance, and sphericity were violated, the use of trimmed data better controlled Type I error. The use of trimmed data with either form of ANOVA had greater power than the same test conducted with complete data sets (Cribbie et al., 2012; Keselman et al., 2000).

The choice to remove data should never be made without understanding why outliers appear in the data (Aguinis et al., 2013). Just as outliers should not be eliminated without investigation, neither should they be included only because they are a part of the data. Judd et al. (2017), cautioned:

> we would argue that it is unethical to include clearly outlying observations that "grab" a reported analysis, so that the resulting conclusions misrepresent the majority of the observations in a dataset. The task of data analysis is to build a story of what the data have to tell. If that story really derives from only a few overly influential observations, largely ignoring most of the other observations, then that story is a misrepresentation. (p. 326-327)

The goal of this research was to include all data, as each student lent an important piece to the story of the data. Because the outliers skewed the data and potentially impacted the decisions made based on the statistical analysis, each ANOVA was conducted twice, once with the complete data set and once with the trimmed data. When the decision made for both tests was the same, the results of the complete data was reported. If the two tests resulted in contradictory decisions, it was determined the presence of the outliers had negatively impacted the test, and the trimmed data was used. Conducting the analysis with and without outliers can provide an understanding of the impact of the outliers on the analysis (Liao et al., 2016) and can reduce skepticism resulting from analysis derived after the elimination of data (Aguinis et al., 2013).

Research Question 1

Data were collected to identify the class students were enrolled in, their journal scores, and their test scores. Students who skipped a test, but remained enrolled in the course, received a score of zero on the skipped test and were removed from the analysis for that test. During the semester there were students who dropped the course, which led to a decrease in the sample size for each test. These students were identified in the data by the lack of a grade for a test. No additional grades were recorded for students after they dropped the course.

The journal scores, which were determined using a rubric that ranged between zero and 15 points, were grouped into one of three journal categories which reflected how information about the journal was presented to students in class. On the first day of class, and again after each test, students were told the average grades for students whose journal score fell into each category. Scores between zero and five points were placed into Journal Category 1, scores between six and 13 points were placed into Journal Category 2, and scores between 14 and 15 points were placed into Journal Category 3. Students who scored between zero and five points typically chose not to turn in the journal or turned in a completed glossary section and a small portion of the problem set. Students in this group did not usually attempt to express the relevant methodologies. Students in Journal Category 2 typically completed the glossary and most of the problems, as well as some portion of the methodologies. Students in Journal Category 3 turned in completed journals or turned in journals that were missing either a few problems or one methodology. For example, a student may have completed all of the elements of the journal with the exception of detailing the methodology for completing word problems, which would result in the loss of one point.

The classes students were enrolled in were assigned a numerical code in SPSS. The Class category depended on the number of levels below transfer-level the class was. Algebra II, which was one level below transfer, was coded as Class 1. Algebra I, which was two levels below transfer, were coded as Class 2. Class 3 was used for students in Arithmetic and Prealgebra, which was a course three levels below transfer and was an accelerated course combining the full-semester courses of Arithmetic and Prealgebra. Students in Arithmetic, which was four levels below transfer, were assigned to Class 4.

Assumptions

The first assumption of a one-way ANOVA is that of normality of the residuals. Through a Monte Carlo simulation, Schmider et al. (2010) demonstrated that ANOVA is generally robust against violations of normality and that Type I and Type II errors remained relatively consistent across a variety of distribution shapes. Despite the robust nature of ANOVA to violations of normality, the researchers encouraged the use of sample sizes greater than 25 per condition in order to protect against increased error rates (Schmider et al., 2010). Delacre et al., (2019) identified several reasons why collected data may be unlikely to be normally distributed, including the presence of measurement boundaries.

For this study, normality was assessed through a visual inspection of the Q-Q plot and through an evaluation of the standardized skewness and kurtosis ratios. If the absolute value of the standardized ratios was less than three, then the distribution was determined to be approximately normal. The maximum possible exam score for each test served as an upper bound to scoring that provided a valid explanation for the presence of negative skew. Violations of normality in ANOVA may lead to increases in Type I error, the likelihood of which decreases when samples are large (Delacre et al., 2019). The incidence of Type II errors can also be more influenced by violations of normality, with leptokurtosis being more problematic than skewness (Delacre et al., 2019).

The second assumption of ANOVA is homogeneity of variance, which was determined using Levene's test ($\alpha = .05$). When homogeneity of variance is violated, the Type I error rate is impacted (Delacre et al., 2019; Skidmore & Thompson, 2013). The increase in Type I error rate is larger when the groups are of unequal size (Skidmore & Thompson, 2013), which was the case in this study. When the assumption of homogeneity of variance is violated, the selection of an alternative test is the best option (Delacre et al., 2019; Lix et al., 1996; Skidmore & Thompson, 2013). Welch's test is the most appropriate alternative test when the assumption of homogeneity of variance is violated, provided all groups contain more than 10 elements and data is not overly skewed, which was defined as skewness values more extreme than ± 2.0 (Lix et al., 1996). Welch's test, a parametric test, is a version of the one-way ANOVA and has been shown to be more robust to violations of homogeneity of variance (Delacre et al., 2019).

Several researchers have suggested avoidance of the conventional ANOVA in favor of Welch's test regardless of whether or not assumptions are violated (Cribbie et al., 2012; Delacre et al., 2019; Lix et al., 1996). Researchers have shown the Welch test performs equally well to the conventional ANOVA when assumptions are not violated (Cribbie et al., 2012; Delacre et al., 2019). When the assumption of normality is violated, whether variance is homogeneous or heterogeneous, the Welch test outperforms the conventional ANOVA, and the trimmed Welch test outperforms both (Cribbie et al., 2012). When homogeneity of variance is violated and the groups are not of equal size, Welch's test controls for Type I error better than the conventional ANOVA and has greater power as well (Delacre et al., 2019). Unlike other alternatives to the conventional ANOVA, Welch's test is not affected by the degree of variance heterogeneity when groups are unbalanced (Lix et al., 1996). There are several post hoc tests available for ANOVA, the choice of which depends on the assumptions. In cases where the Welch test is statistically significant and there is homogeneity of variance, Tukey's post hoc test is most appropriate (Field, 2018). When homogeneity of variance is violated, the preferred post hoc test is the Games-Howell test (Field, 2018).

To describe the practical significance of the test results, two different measures of effect size were reported, eta-squared (η^2) and omega-squared (ω^2). Eta-squared, which is commonly reported in educational and psychological research, has potential sampling error bias when assumptions of ANOVA are violated, particularly in cases in which there are unbalanced group sizes (Skidmore & Thompson, 2013). Eta-squared is an estimator of the strength of the association between the independent and dependent variable but tends to overestimate the effect size in the population (Grissom & Kim, 2012). Omega-squared is less susceptible to sampling error bias (Skidmore & Thompson, 2013) and, as a result, was reported here as well. In the event that omega-squared results in a negative value it is recommended to report $\omega^2 = .00$ (Grissom & Kim, 2012).

Welch's Test

For Research Question 1, Welch's test was conducted to compare the means of test scores for students based on their Journal Category with the complete data set and a second time with the 5% trimmed data set (See Table 2 for complete Welch test results). Because the statistical decisions for all six course tests were consistent using both data sets, the results of the complete data set were reported. A second Welch test was conducted to compare the means of test scores for students based on students' Class category and results were compared for the complete and trimmed data. Because the statistical decisions were not consistent across all six tests, the trimmed data was used for reporting of results. Confidence intervals for statistically significant results were determined using either Tukey's test, if there was homogeneity of variance, or Games-Howell if the variance was heterogeneous. Additionally, eta-squared and omega-squared measures of practical significance were reported for all Welch tests.

Table 2

Statistical Summary of Welch's Tests

Source	F	df_1	df_2	p	η^2	ω^2
Test 1						
Journal Category	9.33	2	117.08	< .01	.060	.055
Class[a]	3.68	3	153.84	.01	.033	.024
Test 2						
Journal Category	26.99	2	122.79	< .01	.134	.129
Class[a]	2.73	3	135.81	.05	.024	.015
Test 3						
Journal Category	16.63	2	134.91	< .01	.088	.083
Class[a]	1.28	3	132.30	.28	.014	.004
Test 4						
Journal Category	19.59	2	97.30	< .01	.127	.122
Class[a]	14.62	3	130.38	< .01	.097	.088
Test 5						
Journal Category	9.71	2	129.53	< .01	.059	.054
Class[a]	1.72	3	125.57	.17	.017	.008
Test 6[b]						
Journal Category	10.62	2	51.86	.03	.126	.120
Class[a]	3.16	1	154.88	.08	.017	.010

Note. *F* scores for Journal Category were results of Welch's test conducted with the full data set. *F* scores for Class were results of Welch's test conducted with the 5% trimmed data set. An alpha value of .05 was used to determine statistical significance.

[a] Data provided from the 5% trimmed data set.

[b] Only students in Class 1 and Class 2 were given six tests.

Test 1 Results

The one-way ANOVA considered whether there was a statistically significant difference in average score on Test 1 based on Journal Category. Scores on the journal were placed into one of three categories: (a) Journal Category 1 for scores between zero and five, (b) Journal Category 2 for scores between six and 13, or (c) Journal Category 3 for scores between 14 and 15. Three students skipped the first test and were eliminated from the data.

Data for the remaining 387 students' journal categories were evaluated for normality and homogeneity of variance. The residuals were checked for normality through a visual inspection of the Q-Q Plot and through an evaluation of the standardized skewness and kurtosis values. The Q-Q Plot demonstrated deviations from normality in the tails and the skewness of -1.58 confirmed moderately skewed data. The standardized skewness (z_{skew} = -12.77) and standardized kurtosis ratios (z_{kurt} = 14.33) both confirmed the violation of normality in the complete data set. Levene's test for the complete data set ($p < .01$) showed a violation of the homogeneity of variance assumption.

The assumption violations and the presence of multiple outliers seen in the boxplot led to the decision to consider the 5% trimmed data set. Examination of the Q-Q Plot for the 5% trimmed data showed a much more normal appearance and the standardized skewness (z_{skew} = -5.82) and kurtosis ratios (z_{kurt} = -0.25) demonstrated much better normality. Welch's ANOVA is generally robust to normality violations, particularly when the sample is large and categories contain at least 25 data values (Schmider et al., 2010). The lowest frequency category for Test 1 contained 49 values and the kurtosis was below the normality threshold. As a result, the residuals for the 5% trimmed data set were considered approximately normal. Levene's test for the trimmed data (p = .94) demonstrated homogeneity of variance. Because the 5% trimmed data was more normal, with no evidence of leptokurtosis, and met the assumption of homogeneity of variance there was confidence in the ability to control for Type I error through the evaluation of the Welch test.

Welch's test was conducted for the trimmed data and returned a statistically significant result, $F(2, 77.75) = 21.25$, $p < .01$, $\eta^2 = .115$, $\omega^2 = .109$. The Welch test was also run with the complete data set in an attempt to gather insight into the full breadth of the relationship between journal category and test scores. The complete data Welch test returned a statistically significant result, $F(2, 117.08) = 9.33$, $p < .01$, $\eta^2 = .060$, $\omega^2 = .055$. Because the results of both tests were statistically significant, and the desire was to use the complete data to provide the most complete picture, post hoc testing was conducted on the complete data set. The violation of the assumption of homogeneity of variance in the data meant Games-Howell post hoc testing was most appropriate. Students in Journal Category 3 (M = 83.18, SD = 14.19) scored between 1.22 and 9.00 points higher on the test than students in Journal Category 2 (M = 78.07, SD = 14.77), which was a statistically significant result (p = .01). Students in Journal Category 3 scored between 3.57 and 18.79 points higher on Test 1 than students who were in Journal Category 1 (M = 72.00, SD = 21.09), which was a statistically significant difference (p < .01).

A second Welch test was conducted to investigate to whether there were differences in scores on Test 1 for students enrolled in different classes. The classes were coded based on the number of levels below transfer for each of the four courses. Class 1 referred to students in Algebra II, Class 2 was used for students enrolled in Algebra I, Class 3 referred to students in a combined Arithmetic and Prealgebra class, and Class 4 was for students enrolled in Arithmetic.

Data for the Class categories were evaluated for normality and homogeneity of variance. The Q-Q Plot demonstrated deviations from normality in the tails and the skewness of -1.65 confirmed moderately left-skewed data. The standardized skewness (z_{skew} = -13.33) and standardized kurtosis ratios (z_{kurt} = 14.29) both confirmed the violation of normality in the complete data set. Examination of the box plot showed multiple outliers and Levene's test for the complete data (p = .02) showed a violation of the homogeneity of variance assumption.

The violations of the assumption of normality, the presence of outliers, and the violation of the assumption of homogeneity of variance led to the decision to consider the 5% trimmed data set. Examination of the Q-Q Plot showed a more normal appearance and the standardized skewness (z_{skew} = -5.25) and standardized kurtosis ratios (z_{kurt} = -0.23) indicated better normality. The lowest frequency category for the trimmed Test 1 data contained 52 values so, as a result, the residuals for the 5% trimmed data set were considered approximately normal. Levene's test for the trimmed data (p = .71) demonstrated homogeneity of variance. Because the 5% trimmed data was more normal, with no evidence of leptokurtosis, and met the assumption of homogeneity of variance there was confidence in the ability to control for Type I error through the evaluation of the Welch test.

Welch's test for the trimmed data was conducted and returned a statistically significant result, $F(3, 153.84) = 3.68$, $p = .01$, $\eta^2 = .033$, $\omega^2 = .024$. This differed from the results seen from the investigation of the complete data set, which failed to return a statistically significant result, $F(3, 177.21) = 2.48$, $p = .06$, $\eta^2 = .015$, $\omega^2 = .007$. Although the complete data set did not produce a statistically significant result, the results of the Welch test with the trimmed data indicated the presence of a statistically significant difference in Test 1 scores based on the class taken by students. Because the homogeneity of variance assumption was not violated for the trimmed data, Tukey post hoc testing was conducted to determine in which class the statistically significant differences occurred. Students in Class 1 ($M = 83.59$, $SD = 11.51$) scored between 1.29 and 10.18 points higher than those students in Class 4 ($M = 77.86$, $SD = 11.74$), which was a statistically significant difference ($p = .01$). Students in Class 2 ($M = 82.67$, $SD = 11.30$) scored between 0.11 and 9.51 points higher than students in Class 4, which was statistically significant ($p = .04$). No other differences in mean scores based on class were statistically significant.

Test 2 Results

A total of 388 students were enrolled in classes on the day of the second test. Of those students, 385 took the exam. Inspection of the Q-Q Plot demonstrated deviations from normality in the tails. The investigation of the standardized skewness ($z_{skew} = -9.25$) and kurtosis ($z_{kurt} = 6.55$) ratios confirmed the violation of normality. Levene's test returned a statistically significant result ($p < .01$). Inspection of the box plot for the complete data set showed several outliers. The assumptions for Welch's test were reexamined for the 5% trimmed data set. The Q-Q Plot for the trimmed data set demonstrated only a slight deviation from normality and the standardized skewness (z_{skew} = -4.70) and kurtosis (z_{kurt} = 0.60) both confirmed the more normal distribution of the residuals for the trimmed data set. The smallest category in the trimmed data contained 46 values, which was above the minimum of 25 suggested by Schmider et al. (2010). When evaluated, Levene's test returned a statistically significant result (p = .02), which indicated a lack of homogeneity of variance. The violation of the homogeneity of variance assumption for the trimmed data was concerning, but Welch's test has been shown to be robust to variance heterogeneity (Delacre et al., 2019).

Welch's test for the trimmed data returned a statistically significant result, F(2, 117.02) = 13.64, p < .01, η^2 = .075, ω^2 = .069. This confirmed the statistically significant result returned for the complete data set, F(2, 122.79) = 26.994, p < .01, η^2 = .134, ω^2 = .129. Because Welch tests were statistically significant for both data sets, the results of the Games-Howell post hoc test for the complete data was interpreted. Students in Journal Category 3 (M = 83.46, SD = 13.82) scored between 7.08 and 16.42 points higher than those in Journal Category 2 (M = 71.71, SD = 19.47), which was a statistically significant difference (p < .01). Students in Journal Category 3 scored between 8.75 and 23.71 points higher than students in Journal Category 1 (M = 67.23, SD = 21.58), which was also statistically significant (p < .01).

Evaluation of the complete data residuals for Class category demonstrated multiple outliers and deviations from normality in the tails. Additionally, the standardized skewness (z_{skew} = -9.86) and kurtosis (z_{kurt} = 6.79) ratios confirmed the lack of normality. Levene's test for the complete data set returned a statistically significant result (p < .01). As a result of the violations of normality and homogeneity of variance, as well as the presence of outliers in the data, the 5% trimmed data was evaluated. Visual inspection of the Q-Q Plot showed only slight deviation from normality. The standardized skewness (z_{skew} = -5.66) and kurtosis (z_{kurt} = 0.14) showed some skew, but a lack of leptokurtosis. Levene's test for the trimmed data was not statistically significant (p = .43).

Welch's test for Class category for the 5% trimmed data, F(3, 135.81) = 2.73, p = .05, η^2 = .024, ω^2 = .015, failed to return a statistically significant result. This differed from Welch's test for the complete data set, which was statistically significant F(3, 154.27) = 6.78, p < .01, η^2 = .007, ω^2 = .000. Based on the recommendations made by Cribbie et al. (2012), in the case in which both normality and homogeneity of variance assumptions are violated, the trimmed data set controls for the incidence of Type I error better than the complete data set. Thus, the lack of any statistically significant result was reported.

Test 3 Results

At the time of the third test there were 365 students enrolled, two of whom skipped the test. The residuals for the remaining 363 students were calculated and the assumptions were checked. The Q-Q Plot for the complete data set showed minor deviations from normality in the tails and the box plot displayed only two outliers. Based on the study by Liao et al. (2016), the use of Welch's test in the presence of two outliers and a large sample results in minor impact on Type I error rate. The standardized skewness ratio (z_{skew} = -5.47) was slightly greater than the normality threshold, but the standardized kurtosis ratio (z_{kurt} = -0.40) was well below the threshold for normality. The category with the smallest frequency contained 54 values, which was above the minimum of 25 suggested by Schmider et al. (2010). Levene's test resulted in a statistically significant result ($p < .01$). Because of the skew of the data and the violation of the homogeneity of variance assumption, the assumptions of Welch's test were assessed for the 5% trimmed data.

The residuals for the 5% trimmed data for Journal Category were examined. The Q-Q Plot displayed approximately normal data with slight deviations in the tails. The box plot showed no outliers. The standardized skewness ratio ($z_{skew} = -3.76$) was just above the normality threshold and showed a left skew in the data but the standardized kurtosis ratio ($z_{kurt} = -1.76$) was below the normality threshold. The box plot showed no outliers in the trimmed data. Levene's test returned a non-statistically significant result ($p = .74$), which indicated the data displayed homogeneity of variance.

Welch's test with the 5% trimmed data was statistically significant, $F(2, 121.19) = 4.68$, $p = .01$, $\eta^2 = .029$, $\omega^2 = .023$. Welch's test on the complete data returned a statistically significant result with regards to differences in student scores on Test 3, $F(2, 134.91) = 16.63$, $p < .01$, $\eta^2 = .088$, $\omega^2 = .083$. Because the statistically significant result of the trimmed data set confirmed that of the complete data set, the Games-Howell post hoc test was evaluated for the complete data set. Post hoc tests showed students in Journal Category 3 ($M = 81.61$, $SD = 15.32$) scored statistically significantly higher than students in both Journal Category 2 ($M = 72.96$, $SD = 19.29$, $p < .01$) and Journal Category 1 ($M = 66.86$, $SD = 21.27$, $p < .01$). Students in Journal Category 3 scored between 3.90 and 13.39 points higher than those in Journal Category 2. Students in Journal Category 3 scored between 7.27 and 22.23 points higher than those in Journal Category 1.

Evaluation of the complete data residuals for Class category demonstrated some deviations from normality in the upper tail. Although the standardized skewness (z_{skew} = -6.39) was high, the standardized kurtosis ratio (z_{kurt} = 0.58) fell below the normality threshold and demonstrated a lack of leptokurtosis. Inspection of the box plot showed two outliers. Levene's test for the complete data set returned a non-statistically significant result (p = .13). Although there was a violation of the normality of the residuals due to the left skew of the data, the assumption of homogeneity of variance was met.

The assumptions were checked for the trimmed data and more closely approximated normality. The Q-Q Plot showed some deviations from normality but there were no outliers apparent in the box plot. The standardized skewness (z_{skew} = -4.07), although improved, was above the normality threshold. The standardized kurtosis ratio (z_{kurt} = -2.61) demonstrated increased kurtosis for the trimmed data. Levene's test (p = .28) returned a non-statistically significant result, which showed the homogeneity of variance assumption was met.

Welch's test for the trimmed data was not statistically significant, F(3, 132.30) = 1.28, p = .28, η^2 = .014, ω^2 = .004. Similarly, Welch's test for Class category did not return a statistically significant result, F(3, 148.54) = 0.68, p = .57, η^2 = .005, ω^2 = .000. The lack of a statistically significant result for both data sets was noted but, in order to maintain consistency of reporting for all tests based on Class category, the trimmed data summary was reported.

Test 4 Results

At the time of the fourth test there were 360 students enrolled and six students who skipped the test. The residuals for the remaining 354 students were considered.

Examination of the Q-Q Plot showed slight deviations from normality and the box plot displayed two outliers. The standardized skewness ratio (z_{skew} = -6.15) and kurtosis ratio (z_{kurt} = 1.20) indicated the data was skewed left but below the standardized kurtosis normality threshold. The smallest category contained 43 values, which was above the suggestion of 25 of Schmider et al. (2010). Levene's test returned a statistically significant result ($p < .01$), indicating a violation of the assumption of homogeneity of variance. Based on the skew of the data and the violation of the homogeneity of variance, the 5% trimmed data was evaluated. The Q-Q Plot for the trimmed data demonstrated normality. The standardized skewness ratio (z_{skew} = -3.94) and kurtosis ratio (z_{kurt} = -0.43) indicated the data remained left skewed but was below the standardized kurtosis normality threshold. Levene's test for the trimmed data was statistically significant ($p < .01$). In an attempt to control for Type I error, Welch's test was considered for the complete and the trimmed data set.

Welch's test for the trimmed data was statistically significant, $F(2, 80.73) = 11.04$, $p < .01$, $\eta^2 = .069$, $\omega^2 = .063$. The results with the trimmed data confirmed the statistically significant Welch's test for the complete data set, $F(2, 97.30) = 19.59$, $p < .01$, $\eta^2 = .127$, $\omega^2 = .122$. Because the results for the trimmed data and the complete data were in agreement, the Games-Howell post hoc test was reported for the complete data. Students in Journal Category 3 (M = 80.19, SD = 15.36) scored between 5.62 and 16.24 points higher on the test than students in Journal Category 2 (M = 69.85, SD = 20.82), which was statistically significant ($p < .01$). Students in Journal Category 3 scored between 9.59 and 31.69 points higher than students in Journal Category 1 (M = 60.14, SD = 29.06), which was also statistically significant ($p < .01$).

Residuals for Test 4 and Class category were examined and displayed a deviation from normality in the tails. The standardized skewness ratio indicated left skewed data (z_{skew} = -8.53) and a large standardized kurtosis ratio (z_{kurt} = 4.20). The box plot indicated multiple outliers and Levene's test was statistically significant ($p < .01$). Because of the violations of normality and homogeneity of variance, the 5% trimmed data was examined. The Q-Q Plot for the trimmed data demonstrated deviations from normality, particularly in the upper tail. Although the standardized skewness ratio (z_{skew} = -5.31) indicated the data was left skewed, the standardized kurtosis ratio (z_{kurt} = -0.02) was below the normality threshold. Levene's test was statistically significant ($p < .01$). The smallest group contained 45 elements, which was above the suggestions made by Schmider et al. (2010).

Welch's test for the 5% trimmed data was statistically significant, $F(3, 130.38) = 14.62$, $p < .01$, $\eta^2 = .097$, $\omega^2 = .088$. The results of Welch's test for the complete data set were also statistically significant, $F(3, 140.16) = 12.93$, $p < .01$, $\eta^2 = .077$, $\omega^2 = .069$. Despite the agreement between tests with the trimmed data and the complete data, the desire for consistency of reporting for all tests led to the reporting of confidence intervals for the trimmed data. Because of the violation of the homogeneity of variance assumption, the Games-Howell post hoc test was evaluated. Students in Class 2 (M = 84.28, SD = 12.43) scored statistically significantly higher than students in all other Classes on Test 4. Students in Class 2 scored between 6.72 and 17.10 points higher than students in Class 1(M = 72.37, SD = 17.37, $p < .01$), between 1.18 and 17.51 points higher than students in Class 3 (M = 74.93, SD = 18.89, $p = .02$), and between 4.89 and

17.06 points higher than students in Class 4 (M = 73.30, SD = 14.55, $p < .01$). None of the other Class category differences were statistically significant.

Test 5 Results

A total of 341 students were enrolled in the class at the time of the fifth test and no students skipped the exam. Examination of the Q-Q Plot for the residuals for the complete data set demonstrated some deviations from normality, especially in the tails, and the box plot showed several outliers. Although the standardized skewness ratio (z_{skew} = -6.78) exceeded the normality threshold, the standardized kurtosis ratio (z_{kurt} = 1.25) was below the normality threshold. Levene's test was not statically significant ($p < .01$). Because of the skew, presence of multiple outliers, and variance heterogeneity the 5% trimmed data set was evaluated. The Q-Q Plot for the trimmed data demonstrated better normality and there were no outliers displayed in the box plot. Although still above the normality threshold, the standardized skewness ratio (z_{skew} = -4.37) demonstrated less skew in the data. The standardized kurtosis ratio (z_{kurt} = -1.40) for the trimmed data had changed but was still below the normality threshold. Levene's test was not statistically significant (p = .14), indicating homogeneity of variance.

Welch's test for the trimmed data set was statistically significant, $F(2, 119.01)$ = 3.96, p = .02, η^2 = .028, ω^2 = .022. Welch's test was conducted for the complete data set and returned a statistically significant result, $F(2, 129.53)$ = 9.71, $p < .01$, η^2 = .059, ω^2 = .054. Because Welch's test for the trimmed data set confirmed the result for the complete data set, and because the assumption of homogeneity of variance was violated for the complete data, the Games-Howell post hoc test for the complete data set was investigated. Students in Journal Category 3 (M = 78.73, SD = 17.94) scored between

2.32 and 14.20 points higher than those in Journal Category 2 (*M* = 70.47, *SD* = 21.35), which was statistically significant ($p < .01$). Students in Journal Category 3 scored between 4.03 and 20.62 points higher than those in Journal Category 1 (*M* = 66.40, *SD* = 24.23), which was also statistically significant ($p < .01$).

Evaluation of the residuals for the complete data set for Class category indicated a lack of normality. The Q-Q Plot deviated from normality in the tails and the box plot displayed multiple outliers. The standardized skewness ratio ($z_{skew} = -7.26$) was well above the normality threshold, although the standardized kurtosis ratio ($z_{kurt} = 1.83$), was below the threshold. Levene's test was not statistically significant ($p = .07$), indicating homogeneity of variance was not violated. Because of the skew of the data and the presence of multiple outliers, the assumptions were checked for the trimmed data. The Q-Q Plot was more aligned with normality and the boxplot showed no outliers. The standardized skewness ratio ($z_{skew} = -4.59$) remained above the normality threshold, but was improved, as was the standardized kurtosis ratio ($z_{kurt} = -1.51$). Levene's test returned a non-statistically significant result ($p = .26$).

Welch's test for the trimmed data was not statistically significant, $F(3, 125.57) = 1.72$, $p = .17$, $\eta^2 = .017$, $\omega^2 = .008$. This differed from Welch's test for the complete data, which was statistically significant, $F(3, 144.06) = 2.98$, $p = .03$, $\eta^2 = .023$, $\omega^2 = .014$. Because the Welch test for the trimmed data failed to return a statistically significant result, and in order to report results consistently across all tests, the lack of a statistically significant result was reported.

Test 6 Results

Students in Class 3 and Class 4 were only given five tests each semester. As a result, there were 220 students enrolled in Class 1 and Class 2 at the time of the sixth test. No students missed the test, which meant residuals were evaluated for all 220 students. Investigation of the Q-Q Plot showed the data was well-aligned with normality. The box plot showed four outliers. Although the full data set did not meet the threshold for normality using the standardized skewness ratio (z_{skew} = -4.19), the standardized kurtosis ratio was below the threshold (z_{kurt} = 1.34). Levene's test returned a statistically significant result (p < .01), which indicated a violation of homogeneity of variance. Because of the left skew of the data, the presence of more than two outliers, and the violation of homogeneity of variance, the 5% trimmed data was evaluated. The trimmed data did not violate the normality assumption, as the Q-Q Plot passed a visual inspection, there were no outliers, and both the standardized skewness (z_{skew} = -2.38) and kurtosis ratios (z_{kurt} = 0.71) were below the threshold. The trimmed data did violate the homogeneity of variance assumption (p = .01).

Welch's test with the trimmed data set was statistically significant, F(2, 38.22) = 4.78, p = .01, η^2 = .055, ω^2 = .045. The trimmed data confirmed the statistically significant result for Welch's test with the complete data set, F(2, 51.86) = 10.62, p = .03, η^2 = .126, ω^2 = .120. Because the trimmed data confirmed the results of the complete data, Games-Howell post hoc testing was completed with the complete data set. Students in Journal Category 3 (M = 80.24, SD = 16.24) scored between 1.74 and 14.13 points higher than students in Journal Category 2 (M = 72.31, SD = 18.35), which was statistically significant (p = .01). Students in Journal Category 3 scored between 7.91 and

36.13 points higher than students in Journal Category 1 (M = 58.23, SD = 25.62), which was also statistically significant (p < .01).

Evaluation of the complete data residuals for the Class category demonstrated some deviations from normality. The Q-Q Plot deviated from normality in the tails and the box plot showed four outliers. The standardized skewness ratio (z_{skew} = -5.21) was above the normality threshold but the kurtosis ratio (z_{kurt} = 0.83) fell below the threshold. Levene's test was statistically significant (p < .01). Because of the violations of assumptions and the presence of multiple outliers, the residuals for the trimmed data were evaluated. The Q-Q Plot for the trimmed data was aligned well with normality and the box plot showed only one outlier. Although the standardized skewness ratio (z_{skew} = -3.72) was above the threshold, the kurtosis ratios (z_{kurt} = -0.68) remained below. Levene's test was statistically significant (p = .04).

Welch's test for Class category for the 5% trimmed data, F(1, 154.88) = 3.16, p = .08, η^2 = .017, ω^2 = .010, was not statistically significant. This differed from Welch's test for the complete data set, which was statistically significant F(1, 169.65) = 7.67, p = .01, η^2 = .036, ω^2 = .032. In order to maintain consistency of reporting across all six tests, the trimmed data was used for Class category. Thus, the lack of any statistically significant result was reported.

Research Question 2

Journal scores and corresponding exam scores were gathered for the mixed method ANOVA. First, LPA was conducted to determine how many underlying groups or profiles were appropriate to consider for journal submissions. Students were placed in one of four profiles based on the pattern of their journal submissions, and the four profiles became the between-subjects factors for the mixed model ANOVA. Because not every student took Test 6, only the first five exams were used in the investigation of Research Question 2. Additionally, only students who were enrolled in the class for all of the first five exams were considered because of the repeated nature of the mixed method ANOVA. Eliminating students who dropped the course prior to the fifth exam resulted in a sample of 341 students.

Students were placed into one of four profiles based on patterns of journal submissions found from the LPA (see Figure 2 for graphical representation of journal mean scores for each profile). A total of 241 students were placed in the first profile, labeled Consistently High, which, as a group, had the highest mean score on each of the five journals. Students in Consistently High had mean scores of 14.0 on the first journal, 14.0 on the second journal, 13.1 on the third journal, 13.9 on the fourth journal, and 14.0 on the fifth journal. There were 43 students placed into the second profile, labeled Strong Starters. The journal scores for the Strong Starters began high, with mean scores of 11.4 on Journal 1 and 11.9 on Journal 2. Following the second submission, the scores for Strong Starters began to fall, with mean scores of 8.4 on Journal 3, 8.1 on Journal 4, and 2.5 on Journal 5. Thirty-two students were placed into the third profile, labeled Slow Starters, whose submissions began with low scores and increased on each submission. Slow Starters had mean scores of 5.5 on Journal 1, 5.7 on Journal 2, 7.8 on Journal 3, 10.3 on Journal 4, and 12.0 on Journal 5. Students in the final profile, labeled Consistently Low, and containing 25 students, were characterized by having the lowest mean score for each journal submission. Students in Consistently Low had mean scores of 4.9 on the first journal, 2.6 on the second journal, 3.5 on the third journal, 1.5 on the fourth journal, and 0.5 on the fifth journal.

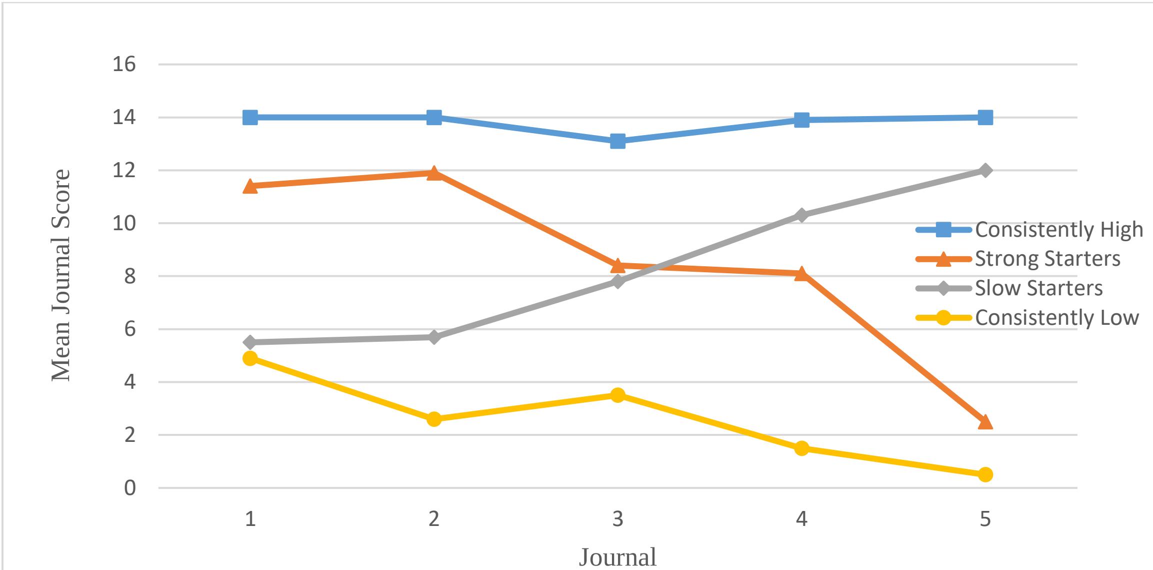

Figure 2. Mean Journal Scores Based on Profile.

Note. Mean journal scores for each submission based on the profile students were placed in.

Assumptions

Mixed ANOVAs incorporate the assumptions of both the within-subject and between-subject ANOVA designs. As with other forms of ANOVA tests, there is an assumption of normality of the residuals (Strunk & Mwavita, 2020). Violations of the normality assumption can result in greater Type I error, particularly when outliers are present or when high degrees of skew are apparent in the data (Keselman et al., 2008). Comparisons of means generated from data containing outliers and large skew may not be accurate because the means are influenced by both heavy skew and the presence of outliers (Keselman et al., 2008). The trimmed mean better represents the center of the data and has been shown to improve the power to identify treatment effects in nonnormal data (Keselman et al., 2008). In repeated ANOVA designs, estimations using the trimmed mean controlled for Type I error better than other outlier mitigation techniques when the normality assumption was violated (Keselman et al., 2008).

The residuals for each test were calculated and normality was assessed through visual inspection of the Q-Q Plots and evaluation of the standardized skewness and kurtosis ratios. If the absolute value of the standardized ratios was less than three then the distribution was determined to be approximately normal. Inspection of the box plot was used to determine whether outliers were present for each test. Examination of the standardized skewness ratio showed a high degree of skew in all tests (see Table 3 for standardized skewness ratios). Of additional concern was the high degree of kurtosis in the majority of tests (see Table 3 for standardized kurtosis ratios). Examination of the box plot showed multiple outliers in the residuals for each test.

Because of the degree of skewness, the high kurtosis in some tests, and the presence of outliers, the data were trimmed in an attempt to correct for the nonnormality of the data. Data were trimmed initially through identification of extreme outliers and the symmetrical trimming of values from the upper and lower ends of the data equal to the number of extreme outliers present in the residuals of each test. When data is deleted from one test of repeated measures ANOVA, SPSS performs a casewise deletion of the remaining data for that participant. Trimming resulted in the removal of data for 32 students, which represented an approximately 10% trimming of data. Of the students removed from the data, the number of students in Consistently High dropped from 241 to 235, the number of students in Strong Starters dropped from 43 to 34, the number of students in Slow Starters dropped from 32 to 25, and the number of students in Consistently Low dropped from 25 to 15.

Table 3

Standardized Skewness and Kurtosis of Residuals

Test	Complete Data		5% Trimmed Data	
	$\frac{\text{Skew}}{SE}$	$\frac{\text{Kurt}}{SE}$	$\frac{\text{Skew}}{SE}$	$\frac{\text{Kurt}}{SE}$
1	-14.88	23.73	-6.70	2.00
2	-10.02	7.85	-7.24	2.32
3	-8.49	5.87	-6.31	1.68
4	-7.49	3.89	-6.60	3.22
5	-7.20	2.06	-6.28	0.20

After trimming the data, the residuals were evaluated for normality through a visual inspection of the Q-Q Plot as well as the calculation of the standardized skewness and kurtosis ratios. The Q-Q Plots for the trimmed residuals demonstrated better alignment with normality, although there was some deviation in the tails for some tests. The trimmed data remained skewed, although this was not unexpected due to the limit placed on the maximum value for each test (Delacre et al., 2019). Although still high and demonstrating left skew in the data, the standardized skewness ratios were greatly improved in the trimmed data (see Table 3 for standardized skewness ratios). The standardized kurtosis ratios were below the normality threshold of 3 for all but one test (see Table 3 for the standardized kurtosis ratio), which was markedly improved when compared to the standardized kurtosis ratios for the complete data. Examination of the box plots showed the existence of a few outliers in the residuals for two of the tests, but no extreme outliers. Because ANOVAs are robust to violations of the normality assumption, particularly in cases in which data is skewed but kurtosis is more normal in shape (Judd et al., 2017) and because the Q-Q Plots and standardized kurtosis ratios fell within acceptable limits, it was determined that the results of the mixed ANOVA could be reliably considered.

The mixed ANOVA has an assumption of homogeneity of variance, which was checked using Levene's test ($\alpha = .05$). Because of the combination of between-subject and within-subject design in the mixed ANOVA, the homogeneity of variance must be checked for each dependent variable, which for this study was the tests (Strunk & Mwavita, 2020). Levene's test refers to a family of tests used to assess homogeneity of variance between two or more populations (Nordstokke & Zumbo, 2010), but can be very sensitive to differences in variance when samples are large and when groups are unequal in size (Field, 2018). The conventional Levene test considers the absolute value of the distance of each score from the sample mean (Nordstokke & Zumbo, 2010). When data is skewed and groups are unequal, the conventional test has inflated Type I error (Nordstokke & Zumbo, 2010). The median-based version, which is a parametric test considering the absolute deviations of scores from the sample median, and the nonparametric version, which is rank-based, both correct for Type I error when data is asymmetric (Nordstokke & Zumbo, 2010). Although both tests have been shown to correct for Type I error well, the nonparametric test has demonstrated greater power than the median-based test (Nordstokke & Zumbo, 2010). Both alternative versions were used to determine the whether the assumption of homogeneity of variance was violated. The results of the median-based Levene test showed homogeneity of variance for Test 2 (p = .81), Test 3 (p = .90), and Test 5 (p = .14). The median-based Levene test for Test 1 (p =

.01) and Test 4 ($p < .01$) were statistically significant, which indicated heterogeneity of variance for those tests. Homogeneity of variance for Test 2 ($p > .99$), Test 3 ($p = .82$), and Test 5 ($p = .24$) was confirmed through the nonparametric Levene test and the results indicated variance homogeneity for Test 1 ($p = .06$) as well. Test 4 ($p < .01$) continued to demonstrate heterogeneity of variance. Examination of the box plot for Test 1 and Test 4, disaggregated by Profile, appeared to confirm the results of the nonparametric Levene test. Because four of the five tests demonstrated homogeneity of variance, the Tukey post hoc test was used to make pairwise comparisons.

The third assumption of the mixed ANOVA is that of sphericity, which was checked with Mauchly's test ($\alpha = .05$). Sphericity is a determination of whether the error variances are approximately equal for all pairs of the within-subject variable (Strunk & Mwavita, 2020). If Mauchly's test returned a statistically significant result, indicating that the sphericity assumption had been violated, then the Greenhouse-Geisser or Huynh-Feldt correction was applied (Strunk & Mwavita, 2020). Both corrections represent adjustments to the degrees of freedom, which controls Type I error (Field, 2018). The choice of the correction to be used is based upon which estimate for sphericity (ε) provided by SPSS is larger. Mauchly's test returned a statistically significant result ($p < .01$), which led to the need to examine the Greenhouse-Geisser ($\varepsilon = .94$) and Huynh-Feldt ($\varepsilon = .96$) corrections. Because the Huynh-Feldt correction was larger, the within-subject ANOVA results were based on this correction.

Mixed ANOVA

The mixed ANOVA returned statistically significant results for the within-subject and between-subject tests. The between-subjects effect to determine whether there was a difference in the mean score based on the profile a student was placed into was statistically significant, $F(3, 305) = 13.85$, $p < .01$, $\eta^2 = .136$, $\omega^2 = .126$. Additionally, the within-subjects effect of Test was evaluated with the Huynh-Feldt correction applied and was statistically significant, $F(3.84, 1169.91) = 19.53$, $p < .01$, $\eta^2 = .064$, $\omega^2 = .061$.

The mixed ANOVA returned a statistically significant interaction effect of Test * Profile. The interaction, with the Huynh-Feldt correction applied, was statistically significant, $F(11.51, 3544.16) = 4.90$, $p < .01$, $\eta^2 = .048$, $\omega^2 = .032$. The interaction indicated that students' performance on the exams varied based on their profile. Further examination of each test, disaggregated by the profile a student was placed in, provided more insight into differences in student scores based on profile. See Figure 3 for a graphical representation of the mean scores on each test disaggregated by profile.

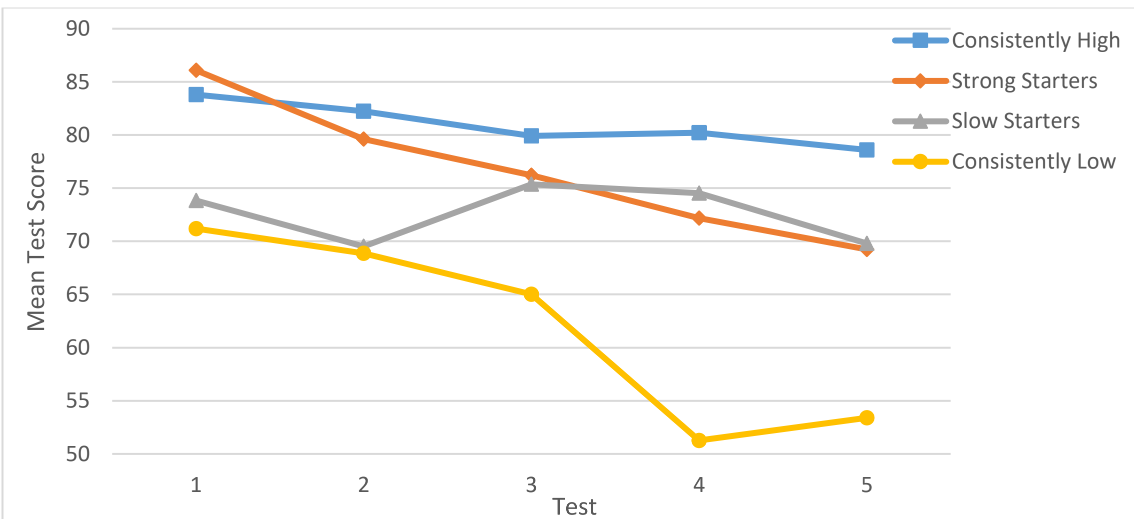

Figure 3. Mean Test Scores by Profile**.**

Note. Mean scores for each test disaggregated by the profile students were placed in.

Student performance on Test 1 showed several statistically significant differences between mean scores for students in different profiles. The mean score for students placed into Consistently High (*M* = 83.80, *SD* = 11.23) was statistically significantly higher than that for students placed into Slow Starters (*M* = 73.84, *SD* = 13.26, $p < .01$) and Consistently Low (*M* = 71.20, *SD* = 18.52, $p < .01$). Students in Consistently High scored between 3.48 and 16.43 points higher than those in Slow Starters and between 4.39 and 20.80 points higher than those students in Consistently Low. Similarly, students in Strong Starters (*M* = 86.09, *SD* = 8.66) scored between 4.14 and 20.36 points higher than students in Slow Starters, which was statistically significant ($p < .01$). Students in Strong Starters scored between 5.34 and 24.43 points higher than those students in Consistently Low, which was statistically significant ($p < .01$). There was no statistically significant difference in Test 1 score between students who were placed in Consistently High and those placed into Strong Starters. There was no statistically significant difference in score for students in Slow Starters and those in Consistently Low.

The results for Test 2 were similar to those for Test 1 when considering the mean scores for students in different profiles. Students placed into Consistently High (*M* = 82.24, *SD* = 13.56) scored between 4.85 and 20.59 points higher than those students in Slow Starters (*M* = 69.52, *SD* = 15.37, $p < .01$) and between 3.42 and 23.34 points higher than those in Consistently Low (*M* = 68.87, *SD* = 15.85, $p < .01$). Students placed in Strong Starters (*M* = 79.62, *SD* = 15.85) scored statistically significantly higher ($p = .04$) than those students in Slow Starters. There was no statistically significant difference in mean test score for students placed in Consistently High and those in Strong Starters. There was no statistically significant difference between scores for students placed in Strong Starters and those in Consistently Low, and there was no statistically significant difference in performance for students in Slow Starters and those in Consistently Low.

On Test 3, the mean score for students placed into Consistently High (M = 79.91, SD = 16.16) was statistically significantly higher (p = .01) than for those students in Consistently Low (M = 65.00, SD = 20.54). Students in Consistently High scored between 3.26 and 26.56 points higher. There were no statistically significant differences in mean scores for students in Strong Starters (M = 76.21, SD = 17.06) or Slow Starters (M = 75.36, SD = 15.96) and any other profiles.

On the fourth test the trimmed mean score for students in each profile was statistically significantly higher than scores for students in Consistently Low. Students in Consistently High (M = 80.21, SD = 15.44) scored between 16.93 and 40.96 points higher than students in Consistently Low (M = 51.27, SD = 29.49, $p < .01$). Students in Strong Starters (M = 72.18, SD = 20.37, $p < .01$) scored between 6.93 and 34.89 points higher and students in Slow Starters (M = 74.52, SD = 16.26, $p < .01$) scored between 8.52 and 37.99 points higher. There were no other statistically significant differences in profile scores on Test 4.

On Test 5, the trimmed mean score for students placed into Consistently High (M = 78.60, SD = 17.48) was statistically significantly higher than mean scores for students placed into Strong Starters (M = 69.24, SD = 22.27, p = .04) and Consistently Low (M = 53.40, SD = 22.77, $p < .01$). Those students in Consistently High scored between 0.40 and 18.32 points higher than students in Strong Starters and between 12.19 and 38.20 points higher than students in Consistently Low. The mean score for students in Strong Starters was between 0.70 and 30.97 points higher than that for students in Consistently Low, which was a statistically significant difference (p = .04). Additionally, the mean score for students in Slow Starters (M = 78.60, SD = 18.31) was between 0.45 and 32.35

points higher than for students in Consistently Low, which was a statistically significant difference (p = .04). There were no other statistically significant differences.

CHAPTER V:

Discussion

The majority of college students recognize that note taking is an essential skill and strive to take good notes during class (Bonner & Holiday, 2006; Williams & Eggert, 2002). Researchers have shown college students struggle to capture more than one-third of key components in a lecture (Bonner & Holliday, 2006; Chen, 2013; Luo et al., 2016; Williams & Eggert, 2002). For developmental students, the amount of key information recorded in class notes can be far lower (Eades & Moore, 2007; King, 1992). Because lecture remains the most prevalent instructional method used in the college classroom (Badger et al., 2001; Bonner & Holiday, 2006), the inability to correctly and completely capture content negatively impacts students' ability to prepare for exams (Williams & Eggert, 2002). A student who captures one-third of the key concepts from class only possesses one-third of the key concepts for studying.

Many college students choose not to review their notes immediately after class, waiting until prior to an exam to interact with their notes, by which time the engagement will be less effective (Cafarella, 2014). Additionally, the passive nature of the interaction, which usually encompasses rereading (King, 1992) or recopying (Bjork et al., 2013) their incomplete notes, is less than optimal (Cafarella, 2014; Williams & Eggert, 2002). Review is an essential component of the note taking process (Al-Musalli, 2015; Boch & Piolat, 2005; Williams & Eggert, 2002). In order to be effective, review needs to occur shortly after a lesson, and notes need to be revised to improve the completeness of the record and to ensure concepts were recorded correctly (Al-Musalli, 2015; Kobayashi, 2006). Students' failure to review their notes often results in their overestimating the level of understanding they have with a concept, which also negatively impacts their test results (Nordell, 2009; Zimmerman et al., 2011). Teaching students how to better gauge their level of understanding, and providing them opportunities for self-reflection, can lead to improved academic performance (Zimmerman et al., 2011).

There are teaching techniques instructors can use to improve the quality of students' class notes, such as providing students with complete or partial notes (Boch & Piolat, 2005; Cardetti et al., 2010; Kiewra et al., 2018; Williams & Eggert, 2002). Additionally, instructors can incorporate instructional practices that help students improve note taking by giving students verbal (Titsworth & Kiewra, 2004) or visual cues (Austin et al., 2004) when key concepts are addressed, or by providing breaks to allow students to collaborate on their notes (Luo et al., 2016). Although all have been shown to improve student performance, all require instructor intervention. These instructor-driven interventions fail to provide students with a framework they can build upon and use in future courses to improve the quality of their class notes. Additionally, none encourage active note review after class or help students to recognize which concepts they understand when preparing for a test.

Researchers have demonstrated that the use of active learning techniques improves performance when compared to passive engagement (Chi & Wylie, 2014; Freeman et al., 2014; Riley & Ward, 2017). When incorporated in STEM classes, active learning has been shown to lead to increased student retention, increased academic performance, and decreased failure rates (Freeman et al., 2014). These results are disproportionately higher for students from disadvantaged backgrounds and female students (Freeman et al., 2014). In their ICAP framework, Chi and Wylie (2014) showed that as learning activities progressed from passive, to active, to constructive, to interactive, learning and performance improved. Students need to be taught strategies for incorporating active learning into their review and preparation for tests (Eades & Moore, 2007). Teaching students a methodology they can use to improve the quality of their note taking provides them with a constructive level of review and helps them recognize which concepts they are struggling with, both of which should lead to improved performance.

This research study was conducted in an attempt to gain greater insight into the degree to which the completion of a constructive note review and note revision intervention leads to improved performance for community college students in developmental mathematics courses. Students who were enrolled in one of four developmental mathematics courses, ranging from one to four levels below transfer, were assigned to complete a note revision assignment prior to each course exam. The journal intervention required them to define key terminology, write out step by step methodologies, and complete sample problems illustrating their methodologies. Student scores on the journal, as well as scores on the corresponding course exams, were recorded and used to consider the research questions.

Research Question 1

Research Question 1 considered to what extent students' performance on each journal intervention impacted their performance on the corresponding exam. In order to consider the question, journal scores (See Appendix A for a sample rubric used to score the journals) were placed into one of three categories. Journal Category 1 was used for students who scored between zero and five points, which typically indicated a student had not turned in the journal or had completed the glossary and less than half of the problems.

Students in this category rarely completed any part of the methods section. Journal scores between six and 13 points were placed into Journal Category 2. Students in this category usually completed the glossary and the assigned problem set, but either failed to start the methods section or, more frequently, would only write methods for some of the outlined procedures. Students who scored between 14 and 15 points on the journal were placed into Journal Category 3. The journals scored in this category were correct and were either complete or were missing the methods for only one procedure.

Conclusions

Importance of Methods Section

On the first day of the semester, students were given a list of the terms to define, methods to explain, and problems to complete for each test. The methods section was the most difficult for students to complete. The glossary terms were available for students to find in the textbook, in their notes, or online, and as a result, was the section most often completed. Although the problem sets required work for students, the completion of assorted problems was more in line with what students had been asked to do in previous math classes. The methods section required students to construct a general strategy for completing a problem, which required students to (a) consider what they understood about the family of problems, (b) assimilate what they understood about all problems of that type, and (c) use their understanding to create a holistic strategy capable of guiding them through completion of problems. Thus, completion of the methods section required more time, thought, and effort than the other sections.

The results indicated the methods section of the journal had the greatest impact on test scores. Based on the scoring rubric, only students in Journal Category 3 could be determined to have mostly, or entirely, completed all elements of the journal. Because of the complexity of the methods section, it was the section least likely to be completed by students who did not complete the journal. It was not unusual for students to turn in complete glossaries and problem sets, and either turn in no methods or incomplete methods. It would have been highly irregular for a student to be in Journal Category 2 with their scoring coming from completing the glossary and methods, and not completing the practice problems. Students in Journal Category 1 typically had not completed the problem sets and had not begun the methods section. The percentage of the problems completed would have been the factor that moved student scores out of Category 1 and into Category 2. Students in Journal Category 2 would have completed the glossary and the majority of the problems. The aspect of the journal that differentiated students in Journal Category 2 and Journal Category 3 was the percentage of the methods section they completed.

Students in Journal Category 3 outperformed the other groups on every test. Because the most notable difference in journal performance between those students and students in the other two categories was the percentage of the methods section completed, the results indicated the methods section as the most valuable for impacting test scores. Developmental students need to be provided with a way to more honestly diagnose their understanding of ideas (Nordell, 2009). The complex nature of writing the methods required students to recognize which concepts they understood and could thoroughly explain and which they needed to work on. Metacognition, which refers to “people’s knowledge of their own information-processing skills, as well as knowledge about the nature of cognitive tasks and of strategies for coping with such tasks” (Schneider &

Artelt, 2010, p. 149), is an essential component of this process. Throughout the semester, time was spent helping students learn how to better construct their own methodologies to improve students' metacognition. Interventions, like the journal, that are designed to improve metacognition in students have been linked to improved performance (Langdon et al., 2019; Nordell, 2009). Constructing a methodology required students to reflect upon how well they truly understood a concept and provided a mechanism for improving their understanding. This is an essential and often overlooked aspect of improving metacognition.

The impact of the methods section on academic performance is a result that makes sense, as the methods section is the most constructive portion of the journal, requiring students to develop written explanations of the content. Based on the ICAP theoretical framework, the methods section should promote the greatest learning (Chi & Wylie, 2014). Completing the glossary and problem sections are, at best, active engagements, but the amount of self-explaining necessary for completing the methods is a powerful constructive engagement (Chi, 2009; Chi & Wylie, 2014). Chi and Wylie (2014) acknowledged that activities may fall on the boundaries between two categories, with classification depending on observation of the overt activities displayed by students. A key differentiation between active and constructive engagement is the ability for students to create new learning from the activity (Chi, 2009). In the context of the journal, the glossary and problems were review of concepts and each would have helped in the review process, but the level of engagement needed to write out the methods was greater and incorporated a higher level of metacognition.

Benefits of Partial Completion

Interestingly, there were no statistically significant difference in performance for students in Journal Category 1 and Journal Category 2 on any of the six tests. Although there were no differences in mean exam scores, the results indicated there were benefits associated with completion of a greater proportion of the journal. Students in Journal Category 1 either failed to turn in a journal or completed the glossary and a small portion of the problems. Students in Journal Category 2 had completed the glossary, at least half of the problems, and might have begun the methods. The ICAP theoretical framework could be effectively used to differentiate active learning activities and to predict which activity would lead to greater learning (Chi & Wylie, 2014). Chi (2009) was also able to use the framework to explain why different learning activities that were classified at the same level showed no appreciable difference in learning. Defining terms and solving the problems, both of which involved review, would both be considered active learning, particularly because there was limited new learning occurring. This would explain the lack of a statistically significant difference in scoring for students in the two groups.

Although there was no statistically significant difference, the results indicated a practically important difference in scoring for students in Journal Category 2. Mean test scores on every test were highest for students in Journal Category 3 and lowest for students in Journal Category 1. Students in Category 2 scored almost a full letter-grade higher than those in Category 1 for the six tests. Although not statistically significant, this is a notable difference for students in developmental classes. Average exam scores for students in Journal Category 2 were above 70% on all but one test, with the Test 4 mean grade being 69.85%. Average scores for students in Journal Category 1 were below 70%

on every test except Test 1, for which the average score was 72.00%. The 70% threshold is an important one for students as it represents the cutoff between a C and a D. Because many developmental courses are required prerequisites for taking future classes, this is an essential threshold for students to meet. The question as to why there was an important difference in scoring is worth considering. Students in Journal Category 2 completed a greater proportion of the journal than those in Journal Category 1 and there was value in the proportion they completed. Students in both groups defined the key terminology but students in Journal Category 2 completed at least half of the assigned problems and might have begun the methods section. Chi and Wylie (2014) recognized that there was a continuum within each classification, with some active learning being closer to passive and some being closer to constructive. Based on the ICAP framework, defining the glossary terms may have been more passive if students were copying the definitions verbatim from the textbook or their notes (Chi & Wylie, 2014). Solving the assigned problems would have been classified as being more on the constructive end of the continuum. Students might have used the problems to build connections between different concepts. Had they created new learning, the task would be more constructive in nature. Additionally, the completion of any part of the methods section would be constructive. It is also possible there was not sufficient power in the ANOVA to identify a difference in scoring because of the number of students in each group (i.e., unbalanced design).

Consistent Benefits Across Courses

For the majority of tests there was no difference in performance for students based on the course they were enrolled in. This was important because the courses for the study ranged from one to four levels below transfer and represented both traditional and accelerated courses. The variation in course level and delivery method indicated the value of the journal intervention for the majority of students, regardless of the level of the course. Providing developmental students with a study technique they can use to effectively prepare for exams for courses at any level is a valuable outcome of teaching students this intervention.

It is important to note the two tests in which there was a noted difference in exam performance. Students in Algebra 1 and Algebra 2 both outperformed students in Arithmetic on Test 1. This result can be explained by the nature of the first test in each course. Students in Arithmetic tended to underprepare for the first exam because the content is operations with whole numbers, a familiar concept they have spent considerable time working on in their academic career. This tended to make students complacent in their preparation, which the findings indicated did not occur on future tests. On Test 4, students in Algebra 1 outperformed students in all other courses. The fourth test in Algebra 1 tests students' understanding of the rules of exponents and operations with polynomials. These topics can be improved with practice, and substantial time was dedicated to practicing these concepts during class. The test concepts in the other classes required more application of concepts, which is often more difficult for students.

Research Question 2

Research Question 2 addressed to what extent the students' journal profile impacted their performance in the class. In order to consider this question, students were placed into one of four profiles based on the pattern or their score on each of the journals.

Because not every class was given six exams, only the first five exams were considered for this question. Students in Consistently High had the highest average journal score in each test period. The Strong Starters began with high journal scores, but scores fell as the semester progressed. The Slow Starters began with low journal scores but increased their scores throughout the semester. The average journal scores for students in the Slow Starter group surpassed those in the Strong Starter group after the third test. Students in Consistently Low had the lowest journal scores throughout the semester.

Conclusions

Improved Journals Led to Improved Tests

Students in the Consistently Low profile averaged less than five points on the journal on each of the five exams. Slow Starters had a slightly higher average on the first two journals, averaging less than six points on each journal. On the first three tests there was no difference in performance for students in these two groups. On the fourth and fifth tests, students in Slow Starters improved their average journal score to above 10 points and subsequently outperformed the Consistently Low students by a statistically significant amount.

Similar results can be seen when investigating student performance in the Consistently High and Strong Starters profiles. Students in both groups averaged over 12 points on the first two journals and there was no statistically significant difference in their scores on Test 1 or Test 2. The gap between the average journal scores for students on the third and fourth tests increased, with Consistently High scores averaging between 13 and 14 points, and the journal scores for Strong Starters remaining above eight points. On Test 5 students in Consistently High maintained an average journal score near 14 points, while the average journal score for students in Strong Starters fell to below four points. Unlike on the first four tests, there was a statistically significant difference in test scores between the two groups on Test 5.

The question remains as to why the decrease in journal scores on Test 3 and Test 4 for Strong Starters did not lead to a difference in test scores. Completion of the glossary and problems sections would have earned students nine points. An inspection of the journal scores for students in Strong Starters on Test 3 and Test 4 showed two-thirds of the students in this profile earned nine points or more, indicating they had completed the glossary and problem sections. Scores for the journals further indicated that more than 40% had begun the methods section. The difference in scoring that caused average scores for Strong Starters to decrease came from a marked increase in the number of students who failed to turn in the journal for Test 3 and Test 4. On the first and second tests combined, only one journal was not turned in by Strong Starters. On Test 3 and Test 4, though, Strong Starters failed to turn in six and eight journals, respectively. The prevalence of students who had begun the methods section may explain the lack of a statistically significant difference in exam scores despite the drop in average journal score created by the increase in students who did not turn in the journal.

The findings indicated an impact on exam score associated with journal performance. Although it is possible there were confounding factors, the difference that arose between students in Consistently High and Strong Starters after Test 4, and the difference that arose between students in Slow Starters and Consistently Low after Test 3, was the change in journal performance. This is important because the two groups of students began with similar journal and test scores at the start of the semester. When the gap between students' journal performance increased, the difference in test scores became significant. This was a key finding as it demonstrated the benefit associated with the intervention for students who early in the semester scored similarly on both the journals and the tests.

There is a lack of longitudinal studies investigating the impact of note taking or note review interventions that are implemented and removed for the same students (Williams & Eggert, 2002). Cohen et al. (2013) investigated the impact of a note revision intervention students completed during specific assigned weeks and did not complete in other weeks. Their research indicated test scores rose in the weeks corresponding to the application of a note review intervention and fell in weeks in which the intervention was removed. Similar results were seen in this study, in which students' grades fell when they chose not to complete the intervention and rose when they completed it. Further, comparisons of the performance gains for students indicated that benefits were roughly equivalent for students at the top, middle, and bottom of course grades (Cohen et al., 2013).

Importance of a Strong Start

The findings indicated the difficulty in making up for a slow start in a developmental mathematics course. As seen in Figure 2 and Figure 3, the Slow Starter group began the semester, when the foundational concepts were introduced, not performing well on the journal or the tests. As the semester progressed, their performance on the journal and the tests improved noticeably. By the end of the semester, students in the Slow Starter profile were outperforming students in the Consistently Low profile, who did not improve their journal performance during the semester. While students in the

Slow Starter profile improved their journal performance, students in the Strong Starter profile saw their journal performance slide. Despite the almost inverse curves of the Slow Starter and Strong Starter journal performance, the Slow Starters were never able to significantly outperform the Strong Starters on the exams.

Developmental mathematics courses are typically sequential in nature, and concepts introduced in the beginning of the semester become the foundation for learning later in the course. The lack of an effective journal from the beginning of the semester negatively impacted performance later in the semester. Students who had created an effective journal early in the semester had a useful tool for review when they forgot something. Cornelius and Owen-Deshcryver (2008) found an encoding benefit for students who actively engaged with guided notes from the start of the semester, and found those students saw improved results during testing late in the semester and in a cumulative final exam. The ability for students to access methodologies written in their own words may have provided an important resource when questions arose later in the semester. Menekse et al. (2013) concluded that the consideration of deep concepts will usually promote greater learning for students because they require students to build connections between new concepts and their old understandings. The researchers found that there was a cumulative effect to learning, through which students who engaged with content at a constructive level possessed a greater base upon which to build connections with future ideas. Not having access to those procedures may have meant students in the Slow Starter profile could begin many problems, but then could not successfully complete the problems because they lacked the foundational concept. This may be further indicated by the decrease in mean exam grade on Test 5 for students in the Slow Starter profile despite the increase in journal grade. As concepts became more difficult, being able to make up for missed concepts early in the semester becomes more unlikely.

Implications for Practice

ICAP Theoretical Framework

This research confirmed the ICAP theoretical framework that as learning activities develop from passive, to active, to constructive, to interactive, learning increases (Chi, 2009; Chi & Wylie, 2014). Of the journal components, only the creation of the methodologies would rise to the level of constructive learning. The defining of glossary terms and the completion of the practice problems are active engagements, particularly in light of the requirement that constructive activities involve the development of new knowledge. The increase in mean grades on the exams for each successive group indicated the importance of constructive engagement for increased learning. Because time was spent during instruction working on the development of the methodologies, and that time was spent with students working together to refine their methods, this engagement might have reached the level of being classified as interactive. Whether students maintained an interactive level of engagement outside of class cannot be known. Importantly, only students in Journal Category 3 completed the majority of the methods, which means they were the students who spent the most time interacting at a constructive level with the course content.

Chi and Wylie (2014) stressed the importance of classifying activities based on the visible actions of students. The requirement that classifications be determined from what can be seen is important because it is not possible to understand how students engage with content they internalize. The journal provides an opportunity to assess a written output and the action of writing it down serves an important purpose. Creating a document allows students to reduce the memory requirement of learning because they are creating a written record (Chi & Wylie, 2014). The written journal “provides students the opportunity to monitor whether in fact they know the material” (Chi & Wylie, 2014, p. 224). Creating an external document provides a chance for students to assess and improve their metacognition.

Journal as Guided Notes

The conclusions reached in response to both research questions indicated a positive impact for completing the methods section of the journal on student performance, which is a relationship worth considering. Each student was provided a summary of the chapter, which included the terms to be defined, the methods to be expressed, and a list of the practice problems to be completed (See Appendix C for an example). Because students were provided the summary of all chapters at the start of the semester and were encouraged to use the summary to guide their note taking, it is possible that students used the chapter summary as a form of guided notes. Thus, students were able to pay closer attention to specific concepts during class because they understood in advance which concepts were most important. Although it is not possible to know whether students chose to use the journal summary to guide their note taking during class, to inform their note editing after class, or only considered the summary in preparing the journal, the summary of important methodologies led to stronger preparation for students who chose to consider them.

Guided notes improve the quality of note taking beyond that of student generated notes, which improves the quality of the study resource (Kiewra et al., 2018). Students have noted they follow lessons more attentively when provided guided notes and are able to use the guided notes to better organize themselves when reviewing and studying (Cardetti et al., 2010). Cornelius and Owen-DeSchryver (2008) noted that students who were provided guided notes were able to more easily learn because the partial notes encouraged them to actively engage with the course content, which led to improved performance. Summarizing content by explaining it, similar to the process students go through in completing the methods section, is characterized as constructive engagement based on the ICAP framework (Chi & Wylie, 2014). This engagement creates strong learning that, in turn, builds a solid basis for building future learning (Menekse et al., 2013) and better prepares students to see connections between concepts learned throughout the semester (Cornelius & Owen-DeSchryver, 2008).

Limitations

There were several limitations in this study. One threat to internal validity was the mortality threat. The number of enrolled students fell for each of the five exams taken by all classes. Although the sample remained large for each of the Welch ANOVA, the loss of students may have resulted in the loss of valuable data. The loss of students due to withdrawal led to a reduction of 50 participants for the mixed ANOVA. Although it is not possible to know why students chose to withdraw from a course, it is likely that many of these students were underperforming in the course and chose to drop because of poor grades. The loss of struggling students posed a threat to internal validity as it may have impacted the differences between groups. If students who withdrew were the lowest performing students and were disproportionately represented in the smallest samples, then their loss may have decreased the power of the test and muted the ability to identify statistically significant differences.

A second threat to internal validity is selection bias (Fraenkel et al., 2015). It is known that at the time of this study it was not uncommon for students to be placed into classes below the level at which they could have been successful (Rutschow & Mayer, 2018). If students were under-placed, it is possible they could have done well on the exams despite not completing the journal, which may have made differences in groups more difficult to identify for students who were placed at the correct level. It is also possible that students who were under-placed completed the journal and performed well on the exams, not because of the benefits of journal, but because of their comfort with the material. This may have increased the average exam grades for the high performing students and could have made differences in group performances more extreme.

In addition to the internal validity threats, there were limitations in the study that impacted external validity. The use of a convenience sample may impact the generalizability of the results, but this external validity threat can be protected against if the results are demonstrated over a wide variety of conditions (Shadish et al., 2002) and are repeated multiple times with similar samples in order to demonstrate consistent results (Fraenkel et al., 2015). In this study, students were enrolled in several levels of developmental mathematics classes and the sample represented students over a three-year time period. A more balanced design would have helped with the control of Type I error and may have provided more power to each test, which may have allowed for the determination of additional significant results. In particular, the size of the Consistently Low group, which after trimming contained 15 students, was small.

The use of Welch ANOVA in considering Research Question 1 required the division of students into journal categories. The selection of the scoring ranges for the journal categories was based on the historical practice used by the researcher when explaining the journal to students. This was not an arbitrary choice, and reflected the scoring tendencies based on the rubric, but was not the only choice. Shadish et al. (2002) warn against the interaction of the causal relationship with units, which may have meant the results would be different had alternate ranges of scores been chosen.

Another concern is that all students came from the same college and had the same instructor. This provided consistency in course content, testing style, and teaching style, which protected against concerns of the external validity threat of treatment variations, but there may be concerns as to whether the results were due more to pedagogy than the intervention. The number of students who were in Journal Category 3, as well as the number of students in the Consistently High group may speak more to the emphasis placed on the journal by the instructor than to buy-in by the students. In order for any intervention to have value, students must embrace it and recognize its benefit. Gathering information about student perceptions in this study may have shed some light on the student perspective.

Recommendations for Future Research

The results in this study indicated there was more impact on student performance from the completion of the methods section than either the glossary or the associated problem set. Future research should be conducted to investigate this relationship more fully. There would be value in conducting an experiment in which students were randomly assigned to one of four groups to determine the extent to which the benefit of the journal is associated with one section or the completion of the full journal as assigned in this study. This would provide insight into whether there was greater impact from the combination of improved familiarity with the language, metacognitive development from the methods section, and practice from completing the problem set, if the completion of any one section led to similar positive impact, or if the completion of any one section provided minimal impact to academic performance.

The benefit to students of any intervention is to provide them with a framework they can use immediately, and in future courses, to improve their learning and likelihood for success. It is important students recognize the value in an intervention so they will continue to use the new skill when it is no longer graded or being actively taught as a part of the course. A qualitative study to consider student perceptions of the benefit of completing the journal intervention would be valuable and may provide insight on which aspects students find the most helpful, as well as which aspects they believe they will continue to use in future courses.

The improved performance seen from students who completed the journal in this study indicated the value in the journal for developmental students. Because the result was consistent across various levels of developmental mathematics courses there is evidence that the skills learned by students could be applied in courses beyond what they were currently enrolled in. Further examination of the value of the journal intervention for students in transfer level mathematics courses would be positive. One goal of developmental instructors should be to provide students with the skills needed to be successful in transfer level courses. If the results of this study can be repeated for college level mathematics courses, then the introduction of the journal to students has value far beyond the immediate impact in developmental coursework.

Conclusion

The primary focus of this study was the degree to which an active note review and revision intervention would impact student performance in developmental mathematics classes. The findings indicated value in the completion of the journal for courses that ranged from one to four levels below transfer. That students who successfully completed the journal consistently outperformed students who failed to turn in completed journals speaks to the potential value of the intervention in helping students learn mathematical concepts and succeed in their developmental coursework. The introduction of the journal as a valuable study tool must be done early in the semester, as students who started strong with their journal performance continued to see a latent impact of their early work, even when their journal results fell off. Students who started slowly were unable to surpass students who started the semester strong, despite their journal scores being far superior at the end of the semester.

REFERENCES

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods, 16*(2), 270-301. https://doi.org/10.1177/1094428112470848

Al-Musalli, A. M. (2015). Taxonomy of lecture note-taking skills and subskills. *International Journal of Listening, 29*, 134-147. https://doi.org/10.1080/10904018.2015.1011643

Austin, J. L., Lee, M., & Carr, J. P. (2004). The effects of guided notes on undergraduate students' recording of lecture content. *Journal of Instructional Psychology, 31*(4), 314-320. https://doi.org/10.1177/1077695816653857

Badger, R., White, G., Sutherland, P., & Haggis, T. (2001). Note perfect: An investigation of how students view taking notes in lecture. *System, 29*(3), 405-417. http://dx.doi.org/10.1016/S0346-251X(01)00028-8

Barnett, J. E. (2003). Do instructor-provided online notes facilitate student learning? *The Journal of Interactive Online Learning, 2*(2), 1-7.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64*(1), 417-444. https://doi.org/10.1146/annurev-psych-113011-143823

Boch, F., & Piolat, A. (2005). Note taking and learning: A summary of research. *The WAC Journal, 16*, 101-113.

Bohay, M., Blakely, D. P., Tamplin, A. K., & Radvansky, G. A. (2011). Note taking, review, memory, and comprehension. *American Journal of Psychology, 124*(1), 63-73. https://doi.org/10.5406/amerjpsyc.124.1.0063

Bol, L., Campbell, K. D. Y., Perez, T., & Yen, C.-J. (2015). The effects of self-regulated learning training on community college students' metacognition and achievement in developmental math courses. *Community College Journal of Research and Practice, 40*(6), 480-495. https://doi.org/10.1080/10668926.2015.1068718

Bonner, J. M., & Holliday, W. G. (2006). How college science students engage in note-taking strategies. *Journal of Research in Science Teaching, 43*(8), 786-818. https://doi.org/10.1002/tea.20115

Bui, D. C., Myerson, J., & Hale, S. (2012). Note-taking with computers: Exploring alternative strategies for improved recall. *Journal of Educational Psychology, 105*(2), 299-309. https://doi.org/10.1037/a0030367

Cafarella, B. V. (2014). Exploring best practices in developmental math. *Research & Teaching in Developmental Education, 30*(2), 35-64.

Cardetti, F., Khamsemanan, N., & Orgnero, M. C. (2010). Insights regarding the usefulness of partial notes in mathematics courses. *Journal of the Scholarship of Teaching and Learning, 10*(1), 80-92.

Castelló, M., & Monereo, C. (2005). Students' note-taking as a knowledge construction tool. *L1-Educational Studies in Language and Literature, 5*(3), 265-285. https://doi.org/10.1007/s10674-005-8557-4

Chen, P.-H. (2013). The effects of college students' in-class and after-class lecture note taking on academic performance. *Asia-Pacific Education Researcher, 22*(2), 173-180. https://doi.org/10.1007/s40299-012-0010-8

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*(1), 73-105. https://doi.org/10.1111/j.1756-8765.2008.01005.x

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning techniques. *Educational Psychologist*, *49*(4), 219-243. https://doi.org/10.1080/00461520.2014.965823

Cohen, D., Kim, E., Tan, J., & Winkelmes, M. (2013). A note-restructuring intervention increases students' exam scores. *College Teaching*, *61*(3), 95-99. https://doi.org/10.1080/87567555.2013.793168

Conference Board of the Mathematical Sciences. (2016). *Active learning in post-secondary mathematics education*. http://www.cbmsweb.org/2016/07/active-learning-in-post-secondary-mathematics-education/

Cornelius, T. L., & Owen-DeSchryver, J. (2008). Differential effects of full and partial notes on learning outcomes and attendance. *Teaching of Psychology*, *35*(1), 6-12. https://doi.org/10.1080/00986280701818466

Cribbie, R. A., Fiksenbaum, L., Keselman, H. J., & Wilcox, R. R. (2012). Effect of non-normality on test statistics for one-way independent group designs. *British Journal of Mathematical and Statistical Psychology*, *65*(1), 56-73. https://doi.org/10.1111/j.2044-8317.2011.2014.x

Delacre, M., Leys, C., Mora, Y. L., & Lakens, D. (2019). Taking parametric assumptions seriously: Arguments for the use of Welch's *F*-test instead of the classical *F*-test in one-way ANOVA. *International Review of Social Psychology*, *32*(1), 1-12. https://doi.org/10.5334/irsp.198

Eades, C., & Moore, W. M. (2007). Ideas in practice: Strategic note taking in developmental mathematics. *Journal of Developmental Education*, *31*(2), 18-26.

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th Ed.). SAGE Publications.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2015). *How to design and evaluate research in education* (9th ed.). McGraw-Hill Education.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 1-6. https://doi.org/10.1073/pnas.1319030111

Grabe, M. (2004). Voluntary use of online lecture notes: Correlates of note use and note use as an alternative to class attendance. *Computers and Education*, *44*(4), 409-421. https://doi.org/10.1016/j.compedu.2004.04.005

Grissom, R. J., & Kim, J. J. (2012). *Effect sizes for research: Univariate and multivariate applications* (2nd Ed.). Routledge.

Howard, L., & Whitaker, M. (2011). Unsuccessful and successful mathematics learning: Developmental students' perceptions. *Journal of Developmental Education, 35*(2), 2-16.

Hudesman, J., Crosby, S., Ziehmke, N., Everson, H., Isaac, S., Flugman, B., Zimmerman, B., & Moylan, A. (2014). Using formative assessment and self-regulated learning to help developmental math students achieve: A multi-campus program. *Journal on Excellence in College Teaching*, *25*(2), 107-130.

Jager, J., Putnick, D. L., & Bornstein, M. H. (2017). More than just convenient: The scientific merits of homogenous convenience samples [Monograph]. *Monographs of the Society for Research in Child Development, 82*(2), 13-30. https://doi.org/10.1111/mono.12296

Judd, C. M., McClelland, G. H., & Ryan, C. S. (2017). *Data analysis: A model comparison approach to regression, ANOVA, and beyond* (3rd Ed.). Routledge.

Keselman, H. J., Algina, J., Lix, L. M., Wilcox, R. R., & Deering, K. N. (2008). A generally robust approach for testing hypotheses and setting confidence intervals for effect size. *Psychological Methods*, *13*(2), 110-129. https://doi.org/10.1037/1082-989X.13.2.110

Keselman, H. J., Kowalchuk, R. K., Algina, J., Lix, L. M., & Wilcox, R. R. (2000). Testing treatment effects in repeated measures designs: Trimmed means and bootstrapping. *The British Journal of Mathematical and Statistical Psychology, 53*(2), 175–191. https://doi.org/10.1348/000711000159286

Kiewra, K. A. (2016). Notetaking on trial: A legal application of note-taking research. *Educational Psychology Review*, *28*(2), 377-384. https://doi.org/10.1007/s10648-015-9353-z

Kiewra, K. A., Colliot, T., & Lu, J. (2018). Note this: How to improve student note taking. *The IDEA Center, IDEA Paper #73*. https://www.ideaedu.org

King, A. (1992). Comparison of self-questioning, summarizing, and note-taking review as strategies for learning from lectures. *American Educational Research Journal, 29*(2), 303-323. http://dx.doi.org/10.3102/00028312029002303

Kobayashi, K. (2006). Conditional effects of interventions in note-taking procedures on learning: A meta-analysis. *Japanese Psychological Research, 48*(2), 109-114. https://doi.org/10.1111/j.1468-5884.2006.00311.x

Landrum, R. E. (2010). Faculty and student perceptions of providing instructor lecture notes to students: Mix or match? *Journal of Instructional Psychology, 37*(3), 216-221.

Langdon, J., Botnaru, D. T., Wittenberg, M., Riggs, A. J., Mutchler, J., Syno, M., & Caciula, M. C. (2019). Examining the effects of different teaching strategies on metacognition and academic performance. *Advances in Physiology Education, 43*(3), 414-422. https://doi.org/10.1152/advan.00013.2018

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Houghton Mifflin.

Liao, H., Li, Y., & Brooks, G. (2016). Outlier impact and accommodation methods: Multiple comparisons of Type I error rates. *Journal of Modern Applied Statistical Methods, 15*(1), 452-471. https://doi.org/10.22237/jmasm/1462076520

Lix, L. M., Keselman, J. C., & Keselman. H. J. (1996). Consequences of assumptions violations revisited: A quantitative review of alternatives to the one-way analysis of variance *F* test. *Review of Educational Research, 66*(4), 579-619. https://doi.org/10.3102/00346543066004579

Luo, L., Kiewra, K. A., & Samuelson, L. (2016). Revising lecture notes: How revision, pauses, and partners affect note taking and achievement. *Instructional Science, 44*(1), 45-67. https://doi.org/10.1007/s11251-016-9370-4

Masyn, K. E. (2013). Latent class analysis and finite mixture modeling. In T. D. Little (Ed.) *The Oxford handbook of quantitative methods: Statistical analysis* (Vol. 2, pp. 551-611). Oxford University Press.

Menekse, M., Stump, G. S., Krause, S., & Chi, M. T. H. (2013). Differentiated overt learning activities for effective instruction in engineering classrooms. *Journal of Engineering Education*, *102*(3), 346-374. https://doi.org/10.1002/jee.20021

Mesa, V., Celis, S., & Lande, E. (2014). Teaching approaches of community college faculty: Do they relate to classroom practices? *American Educational Research*, *51*(1), 117-151. https://doi.org/10.3102/0002831213505759

Murrar, S., & Brauer, M. (2018). Mixed model analysis of variance. In B. B. Frey (Ed.) *The SAGE encyclopedia of educational research, measurement, and evaluation* (pp. 1075-1078). SAGE Publications. https://doi.org/10.4135/9781506326139.n436

Nordell, S. E. (2009). Learning how to learn: A model for teaching students learning strategies. *Bioscene: Journal of College Biology Teaching*, *35*(1), 35-42.

Nordstokke, D. W., & Zumbo, B. D. (2010). A new nonparametric Levene test for equal variance. *Psicológica, 31*(2), 401-430.

Raver, S. A., & Maydosz, A. S. (2010). Impact of the provision and timing of instructor-provided notes on university students' learning. *Active Learning in Higher Education, 11*(3), 189-200. https://doi.org/10.1177/1469787410379682

Riley, J., & Ward, K. (2017). Active learning, cooperative active learning, and passive learning methods in an accounting information systems course. *Issues in Accounting Education*, *32*(2), 1-16. https://doi.org/10.2308/iace-51366

Rutschow, E. Z., & Mayer, A. K. (2018). *Early findings from a national survey of developmental education practices.* Center for the Analysis of Postsecondary Readiness.

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology*, *6*(4), 147-151. https://doi.org/10.1027/1614-2241/a000016

Schneider, W., & Artelt, C. (2010). Metacognition and mathematics education. *ZDM Mathematics Education*, *42*(2), 149-161. https://doi.org/10.1007/s11858-010-0240-2

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Skidmore, S. T., & Thompson, B. (2013). Bias and precision of some classical ANOVA effect sizes when assumptions are violated. *Behavior Research Methods*, *45*(2), 536-546. https://doi.org/10.3758/s13428-012-0257-2

Strunk, K. K., & Mwavita, M. (2020). *Design and analysis in educational research: ANOVA designs in SPSS.* Taylor & Francis.

Stutts, K. J., Beverly, M. M., & Kelley, S. F. (2013). Evaluation of note taking method on academic performance in undergraduate animal science courses. *NACTA Journal*, *57*(3), 38-39.

Tein, J.-Y., Coxe, S., & Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. *Structural Equation Modeling*, *20*(4), 640-567. https://doi.org/10.1080/10705511.2013.824781

Titsworth, B. S., & Kiewra, K. A. (2004). Spoken organizational lecture cues and student notetaking as facilitators of student learning. *Contemporary Educational Psychology*, *29*(4), 447-461. https://doi.org/10.1016/j.cedpsych.2003.12.001

Vandehey, M. A., Marsh, C. M., & Diekhoff, G. M. (2005). Providing students with instructors' notes: Problems with reading, studying, and attendance. *Teaching of Psychology*, *32*(1), 49-52.

Van der Stel, M., Veenman, M. V. J., Deelen, K., & Haenen, J. (2010). The increasing role of metacognitive skills in math: A cross-sectional study from a developmental perspective. *ZDM Mathematics Education*, *42*(2), 219-229. https://doi.org/10.1007/s11858-009-0224-2

Van Meter, P., Yokoi, L, & Pressley, M. (1994). College students' theory of note-taking derived from their perceptions of note-taking. *Journal of Educational Psychology*, *86*(3), 323-338. https://doi.org/10.1037/0022-0663.86.3.323

Wiggins, B. L., Eddy, S. L., Grunspan, D. Z., & Crowe, A. J. (2017). The ICAP active learning framework predicts the learning gains observed in intensely active classroom experiences. *AERA Open*, *3*(2), 1-14. https://doi.org/10.1177/2332858417708567

Wilkinson, L., & Task Force on Statistical Inference, American Psychological Association, Science Directorate. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*(8), 594–604. https://doi.org/10.1037/0003-066X.54.8.594

Williams, R. L., & Eggert, A. C. (2002). Notetaking in college classes: Student patterns and instructional strategies. *The Journal of General Education, 51*(3), 173-199. https://doi.org/10.1353/jge.2003.0006

Williams, G. A., & Kibowski, F. (2016). Latent class analysis and latent profile analysis. In L. A. Jason, & D. S. Glenwick (Eds.), *Handbook of methodological approaches to community-based research: Qualitative, quantitative, and mixed methods* (pp. 143-152). Oxford University Press. https://doi.org/10.1093/med:psych/9780190243654.003.0015

Williams, W. H., Weil, T. M., & Porter, J. C. K. (2012). The relative effects of traditional lectures and guided notes lectures on university student test scores. *The Behavior Analyst Today, 13*(1), 12-16. https://doi.org/10.1037/h0100713

Zientek, L. R., Ozel, Z. E. Y., Fong, C. J., & Griffin, M. (2013). Student success in developmental mathematics courses. *Community College Journal of Research and Practice, 37*(12), 990-1010. https://doi.org/10.1080/10668926.2010.491993

Zimmerman, B. J., Moylan, A., Hudesman, J., White, N., & Flugman, B. (2011). Enhancing self-reflection and mathematics achievement of at-risk urban technical college students. *Psychological Test and Assessment Modeling, 53*(1), 141-160.

APPENDIX A

Name ____________________

Journal Rubric

Glossary Section (3 total points)

Criteria	Journal 1	Journal 2	Journal 3	Journal 4	Journal 5
Terms spelled correctly (+1)	☐	☐	☐	☐	☐
Terms defined correctly (+2)	☐	☐	☐	☐	☐
No glossary (0)	☐	☐	☐	☐	☐
Incomplete glossary (maximum 2 pts)	☐	☐	☐	☐	☐

Methods Section (6 total points)

Criteria	Journal 1	Journal 2	Journal 3	Journal 4	Journal 5
Methods have titles (+1)	☐	☐	☐	☐	☐
Written step-by-step explanation (+1)	☐	☐	☐	☐	☐
Methods are correct (+2)	☐	☐	☐	☐	☐
Methods have correct examples (+2)	☐	☐	☐	☐	☐
No methods (0)	☐	☐	☐	☐	☐
Incomplete methods (max 3 pts)	☐	☐	☐	☐	☐

Chapter Review Section (6 total points)

Criteria	Journal 1	Journal 2	Journal 3	Journal 4	Journal 5
Problems completed (+6)	☐	☐	☐	☐	☐
Most problems completed (+4)	☐	☐	☐	☐	☐
Half of problems completed (+3)	☐	☐	☐	☐	☐
Less than half of problems completed (+1)	☐	☐	☐	☐	☐
No Problems (0)	☐	☐	☐	☐	☐
A few incorrect solutions (−1)	☐	☐	☐	☐	☐
$>$ 10 incorrect solutions (−3)	☐	☐	☐	☐	☐
FINAL SCORE					

Comments

Journal 1	Journal 3

Journal 2	Journal 4

Journal 5

Figure A1. Sample Journal Rubric.

APPENDIX B

Journal Guidelines

Math 151-Bloom

The journal is an ongoing project that will span the entire semester. It will summarize the course and provide you with a valuable study resource. The journal should become a steady work in progress. It provides an opportunity for you to review each class, pull out the most important concepts, organize your thoughts, and develop questions you can ask in class. Often the chapter review exercises will require you to put the concepts together and will lead to questions that require you to go more in depth investigating a topic.

The journal will be turned in prior to all exams.

- You must turn in your journal in a **separate** folder.
  - The journal is not your class notes and needs to be separate from your notes.
  - The folder can be of any type, but three ring binders work best.
  - You may organize your journal in any way that makes sense to you, but each section must be clearly labled.
- The journal will have three sections:
  1. **Glossary**

     Each chapter contains terms which will be defined. For each term you will need to provide a definition. You can use the textbook definitions, my definitions, or put the definitions in your own words.

     Example: Set- A set is a collection of objects, called elements, enclosed in braces {}
  2. **Methods**

     In this section you will be responsible for detailing IN WORDS the steps involved in solving the specific types of problems. I will provide a list of those methods that must be included. You are free to include more if you would like. For each method you will write out step-by-step instructions and present an example that illustrates that method. The methods will be discussed in class and you are free to use my methods or those from the book. The example, however, must be one of your own creation, not one from me or the text. The methods should not be just an example or the steps necessary to solve your specific example. The methods should tell the general steps for solving ANY problem.

     The method section must be laid out as follows:
     - Title-The method to be outlined.
     - Steps-Numbered instructions for how to solve the type of problem.
     - Example-Laid out beside the corresponding instruction illustrating the method being described.

To Add Fractions Without a Common Denominator

Steps	Example
	$\frac{2}{3}+\frac{3}{4}$
1. Identify the Least Common Denominator	$LCD \Rightarrow 12$
2. Multiply each fraction by "fancy one"	$\frac{2}{3}\cdot\frac{4}{4}+\frac{3}{4}\cdot\frac{3}{3}$
3. Multiply and rewrite fractions over LCD	$\frac{8}{12}+\frac{9}{12}$
4. Add numerators and write sum over common denominator	$\frac{17}{12}$
5. Simplify if possible	

3. **Chapter Test**
   In this section you will be responsible for doing all of the indicated chapter test exercises.
   - To receive credit you must write the original problem and **you must show your work.**

   The purpose of this section is to supplement the daily homework. Math is a skill and, as such, needs to be practiced to be perfected. The best strategy is to work on the review a little at a time as we move through the chapter. You will find that the chapter review is reflected on exams.

- There are several benefits for completing the journal.
  - Class notes can become cluttered with doodles or can become disorganized. The journal will be a collection of the most important information from the class.
  - Writing the methods and defining the terms will help to solidify the concepts in your mind prior to exams.
  - At the end of the semester you will need to review for the final and it will be easier to review 25-30 pages from your journal instead of picking out the important information from the text or your notes.
  - At the end of the course you will move on to the next class and may need to go back and review a concept that you have seen before but don't remember.

This is a valuable tool that should be given time and taken seriously. Every semester I see a direct correlation between the effort put in to the journal and the grades on the test. Taking the time to organize your work, pay attention to details, and complete the journal is an invaluable tool for preparing for the next day's class, the tests and the final. Give the journal the time and effort it requires.

Figure B1. Example of Journal Guidelines.

Note. **S**tudents were given guidelines for completing the journal on the first day of class.

APPENDIX C

Chapter 3 - Introduction to Graphing

Glossary:

- Rectangular Coordinate System
- Origin
- Axis
- Graph
- Coordinate Pairs (Ordered Pairs)
- Quadrants of the Rectangular Coordinate System
- Intercepts
- Rate
- Slope
- Slope-Intercept Form of a Line
- Point-Slope Form of a Line

Methods:

- Plotting a Point in the Rectangular Coordinate System
- Graphing a Linear Equation Using a Table (Pick and Plug)
- Determining the x-intercept and y-intercept
- Graphing a Linear Equation Using Intercepts
- Graphing a Vertical or Horizontal Line
- Calculating the Slope of a Line
- Graphing Using Slope-intercept Form
- Determining the Equation of a Line using Point-Slope

Chapter Test: Page 227: 1-29 all

Figure C1. Example of Chapter Summary.

Note. Students were given summaries of each chapter to guide their journal.

APPENDIX D

Sam Houston
STATE UNIVERSITY
A Member of The Texas State University System

Date: Jun 5, 2020 10:24 AM CDT

TO: Andrew Bloom Susana Skidmore
FROM: SHSU IRB
PROJECT TITLE: Effect of an Active Note Restructuring Intervention on Developmental Mathematics Students' Test Performance
PROTOCOL #: IRB-2020-141
SUBMISSION TYPE: Initial
ACTION: Exempt
DECISION DATE: June 5, 2020
EXEMPT REVIEW CATEGORY: Category 4. Secondary research for which consent is not required: Secondary research uses of identifiable private information or identifiable biospecimens, if at least one of the following criteria is met:

(i) The identifiable private information or identifiable biospecimens are publicly available;

(ii) Information, which may include information about biospecimens, is recorded by the investigator in such a manner that the identity of the human subjects cannot readily be ascertained directly or through identifiers linked to the subjects, the investigator does not contact the subjects, and the investigator will not re-identify subjects;

(iii) The research involves only information collection and analysis involving the investigator's use of identifiable health information when that use is regulated under 45 CFR parts 160 and 164, subparts A and E, for the purposes of "health care operations" or "research" as those terms are defined at 45 CFR 164.501 or for "public health activities and purposes" as described under 45 CFR 164.512(b); or

(iv) The research is conducted by, or on behalf of, a Federal department or agency using government-generated or government-collected information obtained for nonresearch activities, if the research generates identifiable private information that is or will be maintained on information technology that is subject to and in compliance with section 208(b) of the E-Government Act of 2002, 44 U.S.C. 3501 note, if all of the identifiable private information collected, used, or generated as part of the activity will be maintained in systems of records subject to the Privacy Act of 1974, 5 U.S.C. 552a, and, if applicable, the information used in the research was collected subject to the Paperwork Reduction Act of 1995, 44 U.S.C. 3501 et seq.

Greetings,

Thank you for your submission of Initial Review materials for this project. The Sam Houston State University (SHSU) IRB has determined this project is EXEMPT FROM IRB REVIEW according to federal regulations.

We will retain a copy of this correspondence within our records.

*** What should investigators do when considering changes to an exempt study that could make it nonexempt?**

It is the PI's responsibility to consult with the IRB whenever questions arise about whether planned changes to an exempt study might make that study nonexempt human subjects research.

In this case, please make available sufficient information to the IRB so it can make a correct determination.

If you have any questions, please contact the IRB Office at 936-294-4875 or irb@shsu.edu. Please include your project title and protocol number in all correspondence with this committee.

Sincerely,

Chase Young, Ph.D.
Chair, IRB
Hannah R. Gerber, Ph.D.
Co-Chair, IRB

VITA

Andrew Bloom

EDUCATION

EdD in Developmental Education Administration, Sam Houston State University, July 2017 – present. Dissertation Title: "Effect of an active note restructuring intervention on developmental mathematics students' test performance"

Masters of Arts in Secondary Education, Wake Forest University, 1998

Bachelors of Science in Mathematics, Wake Forest University, 1996

ACADEMIC EMPLOYMENT

Associate Professor, Mathematics Department, Ohlone College, August 2012 – present. Responsibilities include: teaching developmental and transfer-level mathematics courses, course coordinator for Trigonometry, Algebra I, Algebra II, Arithmetic & Pre-Algebra, develop new credit and noncredit curriculum, serving on committees, participating in course and program review.

Math Department Coordinator, Ohlone College, August 2016 – present. Responsibilities include: scheduling department courses, coordinate adjunct hiring, facilitate completion of department paperwork.

Assistant Professor, Mathematics Department, Shasta College, August 2009 – May 2012. Responsibilities include: teaching developmental mathematics courses, serving on committees, participating in program review.

COMMITTEE WORK

Ohlone College:

- Chair, Math Full-Time Faculty Hiring Committee, Spring 2017
- SEM Division Dean Hiring Committee, Spring 2017
- Math Full-Time Faculty Hiring Committee, Spring 2015
- Distance Education Committee, 2017-present
- Student Equity and Achievement Committee, 2018-present
- Summer Bridge Development Committee, 2019-present
- College Council, 2020
- Basic Skills Committee, 2017-2018
- Faculty Professional Development Committee, 2012-2018
- Equivalency Committee, 2014-present
- Student Equity Workgroup, 2015-2016

Shasta College:

Academic Senate, 2010-2012
Basic Skills Committee, 2009-2012
Student Success Committee, 2010-2012

GRANTS

Bay Area Math Consortium, Steering Committee, 2019-present

Greater Bay Area P20 Consortium, Steering Committee, 2017-2019

Faculty Contributor, WIB Biotechnology Grant, 2013-2014

Principal Investigator, CA Math and Science Partnership Grant, Tehama County Department of Education, Red Bluff, CA 2010-2014

PRESENTATIONS

Bloom, A. "Supporting Student Study Skills to Improve Learning," Promoting Active Engagement in the Mathematics Classroom, January 2018

PROFESSIONAL MEMBERSHIPS

American Mathematical Association of Two Year Colleges
Mathematical Association of America