Early Identification and Improvement of Variables Related to Course Success

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Carolyn Anne Blondin

August 2013

Copyright © 2012 by Carolyn Anne Blondin
All rights reserved.

To my nephew, Zachary Leo

Acknowledgements

First and foremost, I would like to thank my advisor and committee chair, Dr. Robert Williams. His work ethic, ongoing willingness to help, and value of professionalism are traits I hope to have throughout my career. I am truly grateful for his guidance and encouragement during the last four years, and I attribute my developed skills in effective teaching, applied research, and professional collaboration to him. Secondly, I would like to thank my committee members, Drs. Sherry Bain, Sherry Bell, and Chris Skinner, for their professional support, insight, and helpful suggestions regarding my methodological design and research implications. I am so grateful for my work under each member of the School Psychology faculty. My strong graduate training is attributed to their dedication to the field, sense of humor, and high expectations of me.

I owe great thanks to my friend and cohort colleague, Charles Galyon. His past research on predictors of exam success set a foundation for my dissertation. Charles is always willing to hear my ideas and give feedback, which helped me considerably from dissertation conceptualization to completion. I also thank Kyle Voils, who put in countless hours listening to audio recordings of my intervention sessions and grading homework questions. I would like to extend my appreciation to the members of the Educational Psychology 210 team, Tiffany Best, Brittany Carstens, Jeremy Coles, Eleanore Trant, and Jennifer Wright, whose assistance in contacting participants and sending me grade data was pivotal to my project.

I would like to give a special thanks to Kristin Maurer, Lisa McCleary, and Cora Taylor for their continued emotional and professional support. I am so lucky to have their friendship. Lastly, I would like to thank my family from the bottom of my heart for their love, encouragement, and pride in me.

Abstract

The process of identifying and improving factors related to early exam success or failure in an undergraduate setting (Ed Psych 210) was divided into 2 separate studies. The first study was a retrospective analysis of 2 years' of data that compared high and low performers on the first course exam with respect to their subsequent success in the course. Mean comparison between initially high (N = 158) and low (N = 163) performers revealed significantly higher means for those in the former group across several academic variables (i.e., critical thinking, grade point average, subsequent exams, practice exams, quiz scores). Analyses of results support identifying individuals who initially struggle in a course in order to provide them early intervention services.

The focus of the second study was to evaluate the effectiveness of an out-of-class, tier-level intervention using a multiple-baseline design across 3 course sections. Within the framework of the Response to Intervention model, I used universal screeners in selecting participants and identified progress-monitoring measures (i.e., homework accuracy) to assess response to intervention. Ten out of 42 eligible students elected to participate in the study, and 7 of the 10 students completed the intervention program. Phase means of homework accuracy increased between baseline and treatment across all sections. Visual analysis of individual response to intervention showed some variability in daily performance across phases. At least 70% of treatment data points were higher than the baseline median level for 6 of the 7 participants. Exam performance increased for those who responded to Tier 2 interventions but stayed consistently low for individuals requiring the most intensive tier. Exam gaps, defined as the differences of individual exam scores from class averages**,** decreased following implementation across participants.

Future research suggestions for implementing tier-based interventions at the college level include replicating current procedures to assess effects of intervention on homework accuracy across different types of students; developing interventions, universal screeners, and progress monitoring tools to fit a variety of course contexts; and creating and evaluating universal screeners and progress monitoring tools that can effectively and efficiently assess target skills.

Keywords: Undergraduate Success, Exam Predictors, Response to Intervention

Table of Contents

Chapter I…….. 1
Introduction and Literature Review……...……...…... 1
Predictors of Exam Performance and Improvement……...……..…… 1
Response to Intervention…….…… 3
General Characteristics of Tiers……..... 3
Model of Intervention……...……..... 4
Universal Screening and Progress Monitoring……...……..…. 5
Success Rates……...……..…… 6
RTI in the Secondary School Setting……..……..…… 6
RTI: Application in the Post-Secondary School Setting……..……. 7
Framework of the Current Exam Improvement Project……...……..……... 8
Chapter II……...……..…… 10
General Method……...……..…….. 10
Participants and Course Structure……...…….…….... 10
Measures of Predictor Variables……...……...…….. 11
Chapter III…….…….. 13
Study 1……...……..……... 13
Method……...……...…… 14
Participants……...……...….. 14
Procedures and Data Analysis……..……...……. 15
Results……...…… 15
Pre-Course Variables……...…….. 16
Course Exams……..……. 17
Other Within-Course Variables…...……...…….. 17
Dropout Rates……...……..…….. 19
Discussion……...……..……... 19
Pre-Course Variables……...……...…….. 20
Within-Course Performance……...……...……. 22
Future Directions……...……...……..…… 23
Major Conclusions and Implications...……...…….. 25
Summary……...……..…….. 26
Chapter IV…….…….. 28
Study 2……...……. 28
Method……...……. 29
Participants……...…… 29
Procedures……..……. 30
Research Design……...…… 36
Inter-Observer Agreement……...…… 38
Treatment Integrity……...……..…… 39
Participant Acceptability……...……..……... 40
Results……...……... 41
Across- and Between-Group Data Analysis of Homework Accuracy…...…..……... 41
Within-Group Data Analysis of Homework Accuracy……...……. 43

Changes in Exam Scores……. 49
Discussion…….. 56
Homework Accuracy…….. 56
Exam Improvement……. 57
Progress Monitoring and Intervention……..……. 58
Future Research Directions……..……. 59
Implications…….…….... 65
Chapter V……..……. 67
General Discussion..... 67
Overall Summary and Direction for Future Research…..……..... 67
Implications and Future Directions……... 68
Conclusions……..……..... 71
List of References..... 74

Appendices…….……. 82
Appendix A……..……..... 83
Appendix B……..…… 103
Appendix C……..…… 115
Appendix D……..…… 117
Appendix E……..……. 118
Appendix F……...…… 119
Appendix G……..…… 123
Appendix H……...…… 126
Appendix I……..……...…….. 127
Vita……...……... 128

List of Tables

Table 1 Study 1 Multivariate Analysis of Differences in Means across Pre- and Within-Course Variables of High- and Low-Exam Groups from Fall '09-Spring '11…….…….. 84

Table 2 Study 1 Multivariate Analysis of Differences in Means across Pre- and Within-Course Variables of High- and Low-Exam Groups from Spring '10-Spring '11……..….. 85

Table 3 Study 1 Multivariate Analysis of Differences in Means across Pre- and Within-Course Variables of High- and Low-Exam Groups for Fall '09-Spring '10 and Fall '10-Spring '11……...……. 86

Table 4 Study 1 Multivariate Analysis of Differences in Means across Pre- and Within-Course Variables of High- and Low-Exam Groups for Fall '09 (1), Spring '10 (2), Fall '10 (3), and Spring '11 (4)…… 87

Table 5 Study 2 Participants, Year in School, Critical Thinking Percentile, and Attendance Rate…….……...……..……...…….. 88

Table 6 Tier 2 Intervention Acceptability Survey Response Averages.…..…….... 89

Table 7 Tier 3 Intervention Acceptability Survey Responses……..……...….... 90

Table 8 Homework Accuracy Mean, Range, and Percent of Data Points Exceeding the Median (PEM) Scores for Individuals and Groups across Baseline and Treatment Phases…….….. 91

Table 9 Mean and Range of Percent of Homework Accuracy across Individuals and Groups for Baseline, Tier 1, Tier 2, and Maintenance Phases……...….. 92

Table 10 Mean Percent of Homework Accuracy and Range (in Parenthesis) across Individuals and Groups for Units A-E……...……..…….. 93

Table 11 Mean Percent and Range (in Parenthesis) of Homework Accuracy across Days in Unit A…..……. 94

Table 12 Mean Percent and Range of Homework Accuracy (in Parenthesis) across Days in Unit B..……. 95

Table 13 Mean Percent and Range of Homework Accuracy (in Parenthesis) across Days in Unit C…… 96

Table 14 Mean Percent and Range of Homework Accuracy across (in Parenthesis) Days in Unit D….…….……. 97

Table 15 Mean Percent and Range of Homework Accuracy (in Parenthesis) across Days in Unit E…… 98

Table 16 Average Exam Percentages (out of 100) and Deviations from Class Mean (Dev.) across Baseline and Treatment/Post-Treatment Phases for Individuals and Groups…..…. 99

Table 17 Units A-E Exam Percentages (out of 100) and Deviations from Class Mean (Dev.) across Individuals and Groups……..…….……...…..... 100

Table 18 Average Exam Score (out of 50), Standard Deviation (in parenthesis), and Difference between Study 2 Participants (Part.) and Eligible Students who did participate in Study 2 (Non-Part.), and Study 2 Participants (Part.) and Low-Exam Group of Study 1 (L.E.)..……... 101

Table 19 Units A-E Exam Percentages (out of 100) across Study 2 Participants and Exam Score Differences (Diff.) between Participants and Non-Participants (N = 32)…….. 102

List of Figures

Figure 1. Flow of intervention procedures for Study 2……..…….. 104

Figure 2. Flow of research design procedures for Study 2……..…… 105

Figure 3. Daily Homework-Accuracy Scores in Baseline and Treatment Phases between Section 1, Section 2, and Section 3……..……. 106

Figure 4. Edith's (Section 1) Daily Homework-Accuracy Scores across Baseline, Tier 2, Tier 3, and Maintenance Phases……...……...……. 107

Figure 5. Kristin's (Section 1) Daily Homework-Accuracy Scores across Baseline, Tier 2, and Maintenance Phases……..……...…….. 108

Figure 6. Sally's (Section 1) Daily Homework-Accuracy Scores across Baseline, Tier 2, and Tier 3 Phases……..…… 109

Figure 7. Emma's (Section 2) Daily Homework-Accuracy Scores across Baseline, Tier 2, and Maintenance Phases…..…….……... 110

Figure 8. Elizabeth's (Section 2) Daily Homework-Accuracy Scores across Baseline, Tier 2, and Maintenance Phases……..…...…...……...…… 111

Figure 9. Becky's (Section 3) Daily Homework-Accuracy Scores across Baseline, Tier 2, and Tier 3 Phases…….……...…….. 112

Figure 10. Stephanie's (Section 3) Daily Homework-Accuracy Scores across Baseline, Tier 2, and Tier 3 Phases……..…… 113

Figure 11. Daily Homework-Accuracy Scores across Baseline and Tier 2 Phases for Kathryn, Kylie, and Allison, participants who discontinued intervention services…….. 114

Chapter I

Introduction and Literature Review

Although the number of individuals entering college continues to increase, problems related to academic difficulty and institution dropout remain high (Zajacova, Lynch, & Espenshade, 2005). Between 1997 and 2007, degree-granting institution enrollment rate rose 34% for full-time students, with the current enrollment nationwide at nearly 30 million (U.S. Department of Education, National Center for Education Statistics, 2009). However, approximately 43% of students enrolled nationwide in a 4-year institution failed to graduate within 6 years (U.S. Department of Education, National Center for Educational Statistics, 2011). Factors contributing to college success include student background characteristics, structural characteristics of the classroom and institution, student interactions with faculty and peers, perceptions of the learning environment, and the quality of effort students devote to academic endeavors (Kuh, Cruce, Shoup, & Kinzie, 2008). Thus, early identification and treatment of problems related to students' prematurely leaving school may reduce the dropout rate in American higher education.

Predictors of Exam Performance and Improvement

A significant portion of overall academic performance and progress, particularly in large entry undergraduate settings, is evaluated through multiple-choice examinations that require students to display an understanding of material and critically apply reviewed concepts (Stepp, Schrock, & Coscarelli, 1996). Success on multiple-choice examinations has been frequently investigated, with identified past predictors including critical thinking, homework preparation, rate of reading comprehension, participation in class discussion, and pre-course academic vocabulary (Galyon, Blondin, Forbes, & Williams, in press; Turner & Williams, 2007; Wallace

& Williams, 2003; Williams, Oliver, Allin, Winn, & Booher, 2003; Williams, Skinner, & Jaspers, 2007; Williams & Worth, 2002). Classroom activities linked to improvement on multiple-choice exams at the collegiate level include taking practice exams similar to the actual exams, completing brief daily writing activities associated with issues addressed on exams, taking accurate discussion notes, and participating in classroom discussion (Foster et al., 2009; Hautau et al., 2006; Holtzman, 2008; Krohn et al., 2008; Oliver & Williams, 2005; Turner et al., 2006; Williams & Eggert, 2002).

Struggle on exams may be attributed to several factors. For example, Wallace and Williams (2003) found that incorrect answers on multiple-choice exams were due to flaws in the students' understanding of information related to tested concepts, confused analysis of answer options on exam items, and weaknesses in students' higher reasoning skills. Studies such as Wallace and Williams suggest that academic struggle is not related to one general problem (e.g., a weakness in completing multiple-choice exams) but rather to a combination of factors. In two separate studies, Galyon and colleagues (in press) explored a combination of exam predictors to determine what patterns of academic variables maximize prediction of exam success. Findings of Galyon and colleagues' study suggest that variance in performance on multiple-choice exams is best explained by a combination of several factors within students (e.g., critical thinking, GPA), as well as variables measured as an outcome of the course (e.g., homework accuracy, participation). In a follow-up study, variables found to be most predictive of exam performance included critical thinking and homework accuracy.

Empirical studies such as those cited above have the potential to guide current intervention development at the college level. For example, inasmuch as critical thinking has consistently been found to be strongly related to exam performance, researchers may find merit

in tailoring interventions to include a critical thinking component. Given that homework accuracy has also been found to be a primary predictor for exam success (Galyon et al., in press), those seeking improvement in exam performance might do well to focus on their daily homework preparation.

Response to Intervention

Responsiveness-to-Intervention (RTI) has been defined as a framework that provides high-quality education to all students and tailors interventions to meet the needs of individual students (Reutebuch, 2008). RTI aims to prevent repeated failure by measuring each student's skill attainment and providing supplemental instruction in the least restrictive environment to those identified as struggling. Core concepts in all RTI procedures mandated by federal education regulations include the following guidelines: a) all students receive high-quality, research-based instruction, b) universal screenings of all students are conducted to assess the attainment and maintenance of an academic skill, c) those students struggling on an academic skill receive early intervention that targets their academic weaknesses, d) progress is monitored frequently for those receiving extra instruction to determine responsiveness to intervention, and e) interventions are tier-based and intensified in terms of time or resources for non-responding students (Jimerson, Burns, & VanDerHeyden, 2007; Mellard, McKnight, & Woods, 2009; Reutebuch, 2008). As a diagnostic model, RTI provides the hope that implementing prevention procedures prior to academic failure will decrease the number of individuals being formally evaluated and placed into special education classrooms.

General characteristics of tiers. At the primary school level, RTI models involve multiple tiers of educational support prior to commencement of a formal special education evaluation or placement. Each model includes a general first tier that provides universal

instruction and services available to all the students. Fuchs, Fuchs, and Compton (2010) recommended that the core instructional program includes research-based teaching methods, opportunities for instructional differentiation, accommodations in the classroom, and problem-solving strategies to address challenges related to academic motivation. The second tier is for those identified as struggling by a universal screener and includes short-term, small-group interventions that provide supplemental help to the identified students. Selection decisions are made by comparing a student's performance on a curriculum measure to that of her or his peers.

Progress is monitored for students at the second tier of RTI, and if they fail to make sufficient progress after a predetermined amount of time, increased targeted instruction is provided. Common practices in Tier 3 include increased time for instruction provided at a one-on-one level with the general teacher or specialist several times weekly (Brozo, 2010). At this level, as teachers are providing the highest quality instruction available for general education students, clearer determinations can be made if the student requires special education services.

Model of intervention. School systems employing RTI typically choose between two models of preventative intervention when developing Tier 2 and 3 procedures: a problem-solving model of intervention or a standard treatment protocol model (Duffy, 2007; Fuchs & Fuchs, 2006; 2007). A problem-solving model intervention is individually-tailored instruction geared toward enhancing already acquired behavioral or academic skills. A standard treatment protocol model is a structured set of evidenced-based interventions with fixed procedures (e.g., duration and frequency of sessions) used across all individuals identified as at-risk (Fuchs & Fuchs 2006). Unlike interventions following the problem-solving model, standard treatment protocol models are designed to promote the acquisition of new skills (Fuchs & Fuchs, 2007). Due to its adherence to rigid procedures, standard treatment protocol models are often preferred for

determining internal validity of an intervention, measuring fidelity of treatment implementation, and assessing the intervention's generalizability across students (Fuchs & Fuchs, 2006).

Universal screening and progress monitoring. RTI programs rely on use of assessment tools to identify individuals needing intervention help and measuring their response to intervention. A universal screener is provided to the entire student body, which increases accountability that each person is receiving appropriate education based on her or his skill level. It also creates a framework of where one student is skill-wise compared to peers in the district. Catts, Petscher, Schatschneider, Bridges, and Mendoza (2009) discussed the importance of administering universal screeners that will accurately measure skills within an appropriate schedule. They described the current challenge of identifying children truly at-risk of possible skill deficits, while limiting the number of individuals who are falsely identified for services. When designing universal screeners, researchers and administrators should remain mindful of the sensitivity of the measure, its reliability, and the optimal time to administer the measure (Catts et al., 2009).

Mellard and colleagues (2009), who recently evaluated the universal screeners and progress monitoring tools used at 41 different school districts, provide general guidelines for identifying struggling students and measuring response to intervention. The study emphasized the inclusion of reliable and valid measures for screeners and progress monitoring tools, as these assessment tools serve as the basis for making placement decisions. Furthermore, Mellard and colleagues encouraged use of parsimonious screeners and progress monitoring tools sensitive to effects of intervention. This arrangement may include different assessments for progress monitoring than were given for screening. Results of a questionnaire provided to school officials

who successfully implemented an RTI program emphasized the importance of progress monitoring and assessment data (Mellard et al., 2009).

Success rates. RTI has received much support from scholars and educational professionals (Burns, Appleton, & Stehouwer, 2005). Numerous system-wide and individual outcome studies have found success using RTI to increase math and reading skills. A meta-analytic review by Burns and colleagues (2005) demonstrated that most research studies measuring success of RTI models showed overall improvements in systemic and student outcomes (mean effect size ranged from .96 to 1.53; standard deviations were .77 for systemic outcomes and 1.02 for student outcomes) and decreases in learning disability diagnoses.

RTI in the secondary school setting. Although a majority of studies on RTI take place at the elementary-school level (Burns et al., 2005), several researchers and educators have recommended applying RTI at the secondary school level (Duffy, 2007). This area of research, currently limited and largely theoretically-based, continues to grow as more researchers conduct empirical studies with this student population. Consensus among researchers is that, similar to practices at the primary school level, RTI decisions at the middle and high school level should be based on multiple sources of empirical data, tiered instruction should include supplemental help in both skill strategies and content, and decisions made on progress and interventions should be based on collaborative problem analysis (Burns, 2008; Shinn, 2008).

While there is a push for RTI implementation in secondary education, modifications of the intervention may be warranted, including changes in the targeted measures used to screen and monitor progress (Duffy, 2007; Ehren, 2008). Intended educational goals will likely be different across academic grades. If the same measures that are used in elementary schools (e.g., words read per minute) are used in high schools, struggling high school students are likely to be

undetected by universal screeners (Shinn, 2008). Students in the secondary schools often have more complex academic concerns that transcend troubles in basic reading or math skills, such as the ability to comprehend domain-specific reading material provided in a variety of courses (e.g., history, social studies, biology, and English).

For the high school population, researchers have recommended that content learning be improved by focusing on both basic and high-level literacy skills, such as proficiency in listening, speaking, reading, and writing (Ehren, 2008; Shinn, 2008). For example, researchers have employed maze-selection measures (which involves reading passages that omit certain words and choosing which word best fits the sentence) to screen and monitor comprehension in the RTI model at the secondary school level (Burns, 2008; Ticha, Espin, & Wayman, 2009).

RTI: Application in the post-secondary school setting. Compared to research at the primary and secondary school levels, no empirical evidence was located indicating that tier-based procedures have been applied at post-secondary levels. Reasons that RTI is not typically considered applicable at the college levels may include the large diversity of topics and course structure across college classes, as well as the amount of face time students have with professors compared to students at the primary and secondary levels.

Unlike RTI interventions provided in primary and secondary schools, participation in a tier-based intervention at the college level will likely be voluntary and therefore student motivation will be a moderating factor. As described by Skinner, Pappas, and Davis (2005), certain teaching strategies will benefit some students more than others depending on the root of their problem. Obstacles to academic success can be split into two broad categories: limited motivation (*won't do*) and limited ability (*can't do*) (Skinner et al., 2005). *Can't do* problems are related to limited resources, understanding, time, or prerequisite skills. Teacher strategies to

help *can't do* problems may include providing extra resources as needed, reviewing or re-teaching course material to enhance understanding, and making modifications or accommodations in the classroom. As opposed to *can't do* issues, when the student is unable to complete a task, *won't do* issues are related to student choice. There are several reasons students might not choose to engage in a task, such as an assigned task requiring too much effort or the credit for doing well on a certain task is devalued by the student (Skinner et al., 2005). RTI interventions target academic skills and, although increased motivation may be an additional outcome, are likely to be more beneficial to college students who are motivated to put forth effort to work through a *can't do* problem. Despite these differences in the academic problems of college students compared to students at other education levels, I expected that RTI, which has been found to be flexible in terms of procedures, measures used, and interventions developed, could be a beneficial framework to identify struggling college students and provide early and appropriate interventions at the post-secondary level.

Framework of the Current Exam Improvement Project

Researchers have examined methods for improving academic success for those struggling in the classroom, and have emphasized the merits of providing prevention and early intervention programs to reduce rates of dropout, skill deficits, and academic failure (Slavin, Karweit, & Wasik, 1992). For example, studies on systematic early reading programs have demonstrated that reading outcome is significantly improved with early intervention delivery (Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998). Nationwide, primary and secondary schools have implemented RTI models in order to detect early difficulties in academic areas such as math and reading, and numerous studies have found this model to increase academic outcome (e.g., Burns et al., 2005). As described in RTI literature, interventions are evidence-based and

routinely checked for appropriateness for each participant. It is likely that college professors who employ early evidenced-based assistance for struggling students will also have more success in improving individual academic outcome than those who provide remedial assistance after repeated failure in a course.

At the post-secondary level, developing interventions that are effective and appropriate first require an overall understanding of what academic factors relate to low versus high performance in a course. Past empirical findings can then inform what academic variables most directly reflect academic progress. Data collected within the course setting that explore what factors appear to account for the relationships between early performance and subsequent academic performance can guide current interventions. This project aimed to enhance current education literature by first exploring retrospective data related to student success and failure and secondly assessing the effectiveness of a tier-based intervention that targets academic struggle at the college level. Specifically, Study 1 focused on identifying what differences across pre-course and within-course variables (e.g., critical thinking, exam performance, quiz performance) existed between college students who did well on the first exam and students who did poorly on the first exam. The aim of Study 2 was to determine if using a tier-based intervention structured after RTI principles would enhance homework accuracy and subsequent exam performance at the college level.

Chapter II

General Method

Participants and Course Structure

Both Studies 1 and 2 were conducted within several sections of a large undergraduate educational psychology course, with each section having approximately 50-55 students enrolled. The course is a requirement for those entering the university's teacher-education program and is recommended for those in their second or higher year in college. Each section of the course was taught by a lead graduate teacher associate. A first-year graduate teacher assistant was also assigned to each section and responsible for record keeping, monitoring attendance, and grading essay quizzes.

In both studies, the general course structure remained largely the same. The course is divided into 5 units (Units A-E) that cover different dimensions of human development, including physical, cognitive, social, psychological, and values development. Each unit is approximately equal in time devoted and work load assigned. Unit formats also remain generally stable during the course: one day is scheduled for video viewing and discussion related to the unit topic, four days are devoted to structured discussion of course issues, one day is allocated to review of an out-of-class practice exam, one day is devoted to an article quiz and discussion of articles, and a final day is reserved for a 50-item multiple-choice exam that covers information related to that unit. The four discussion days include class-wide discussion of homework questions that cover instructor notes. Daily attendance, homework completion, and participation are consistently encouraged across semesters and each variable is enforced by credit contingencies. Inasmuch as the two studies spanned over 5 semesters, some changes to this structure did occur and are discussed within each specific study methodology.

Measures of Predictor Variables

All demographic and academic data were gathered and entered into a database created each semester for student record-keeping purposes. Common variables investigated in both studies included student demographic variables, assessment of quizzes and practice exams, assessment of critical thinking, and evaluation of unit exam scores. Other data collection procedures, such as participation and homework completion, varied across semesters. At the onset of the course, students were asked to fill out a demographic survey that included grade point average, hours of employment per week, and expected grade in the course.

At the beginning of each semester in both studies, students completed the *Watson-Glaser Critical Thinking Test-From-S* (WGCTA, Watson & Glaser, 1994), which assesses several aspects of critical thinking such as inferences, deduction, assumption recognition, and evaluation of arguments. The 40-item test takes 30 min to complete. Each item consists of 2-5 options, with all information needed to answer the item included in the test booklet. Both percentile ranks of raw scores on the WGCTA were established through comparison to a normative sample of college graduates.

Each unit also included a 50-item multiple-choice exam covering information and materials presented during the unit. Based on an in-depth analysis of unit exam items, Wallace and Williams (2003) determined that 58% of the exam items required comprehension to answer the questions, 26% required only direct recall, and 16% were considered a mixture of both direct recall and comprehension. During each testing session, students turned in their exam to their teacher, who immediately scored the test by a test-scanning machine and then provided immediate feedback of the student's performance. The test scanner generated an exam score and

correct options for missed items. Each exam was worth up to 50 points and the combination of the five unit exams approximated 40% of total points in the course.

Short-essay quizzes were given at the end of each unit covering information presented in assigned articles. Total point range was 0-5 and grades were based on accuracy of answer. Practice exams were completed on an out-of-class basis and included 25 multiple-choice items similar in format and content to the unit exams. Students could earn 0-5 points on the practice exam depending on the percentage of correct answers out of 25. Frequency data for participation (e.g., number of comments) and homework questions completed were also recorded each semester.

Chapter III

Study 1

Researchers have found outcome differences across several academic variables between low performers and high performers at the post-secondary level including, but not limited to, critical thinking, homework completion, and participation (Galyon et al., in press; McCleary et al., 2011; Williams et al., 2003; Williams et al., 2004). However, few have examined outcome differences between those identified as high and low performers early in a course. In this current study, I identified differences in performance between these two groups by linking their initial exam performance to pre-course variables and academic outcome within the course.

In the first study, I examined previously collected data of over 600 students recently enrolled in a large undergraduate educational psychology course. I compared differences across several academic variables for those who struggled on the first exam in the course versus those who did well on the first exam. Specific patterns explored included what academic variables differentiated high- and low-exam performance at the outset of the target course (e.g., critical thinking, GPA), what factors were related to preparation for the first exam, how performance on the first exam related to performance on subsequent exams, and how low performance on the first exam linked to preparation strategies for the following exams. Lastly, I compared rates of course dropout for those who struggled early in the course to the total dropout rate for the course across semesters. Based on past research findings on predictors of exam performance in the target course (e.g., Galyon et al., in press), I hypothesized that the means of several variables would be significantly different between low- and high-exam performers.

Method

Participants. Consent to participate in research was obtained for all participants prior to data collection. Two years of data, spanning from Fall 2009 through Spring 2011, were used in analyzing overall performance differences between students who earned a grade of D or F on the first exam and those who earned a B+ or A. The participant pool included 641 students across 12 sections of the large entry-level educational psychology class. The mean class size across all 12 sections was 53 students (range 49-55). Across sections, approximately 78% of the participants were female and 22% were male (503 females and 138 males). With respect to academic classification, the distribution was the following across sections: 42 participants were freshmen (7%), 306 participants were sophomores (48%), 180 of the participants were juniors (28%), 71 of the students were seniors (11%), and 21 of the participants were graduate students (3%). An additional 18 participants (2%) did not indicate academic classification.

Of the 641 participants, 587 students reported their grade point average (GPA). The mean GPA across sections was 3.21, with a range of 1.14 to 4.0. Identification of comparison groups was based on performance on the first course exam. On this exam, 41 of the individuals earned 59% or less of the possible credit (an F), 122 students earned between 60% and 69% of possible credit (a D), 197 students earned between 70% and 79% of possible credit (a C), 121 scored between 80% and 84% of possible credit (a B), 77 students earned between 85% and 89% of possible credit (a B+), and 81 students earned between 90% and 100% of possible credit (an A).

Of the total number of participants ($N = 641$), 319 comprised the High- and Low-Exam Groups. The Low-Exam Group, determined by earning a 34 or lower out of 50 on the first course exam (a D or lower), included 165 individuals (42 males, 123 females). The High-Exam

Group, determined by those who earned a 43 or higher out of 50 (a B+ and higher) on the first exam, had 158 students (24 males, 134 females). Academic classification splits between groups were similar, with the majority of students being sophomores. Within the Low-Exam Group, 6% were freshmen, 46% were sophomores, 35% were juniors, 9% were seniors, and 4% did not report academic year. The High-Exam Group consisted of 6% freshmen, 44% sophomores, 27% juniors, 13% seniors, and 8% who did not report academic year.

Procedures and data analysis. Between-group differences among high and low performers were examined per semester for all of the academic variables collected (i.e., critical thinking, reported GPA, hours of weekly employment, hours of course enrollment, practice-exam scores, homework completion, participation, quiz scores, and exam performance). Between-group analysis between high- and low-performing groups across all semesters was conducted for any variable that was consistently assessed across semesters (i.e., critical thinking, GPA, quiz scores, practice-exam scores, and subsequent exam scores). Rates of dropout from the course prior to the end of the semester were compared between low- and high-performing groups, as well as between the low-performing group and others taking the course.

Results

Several multivariate analysis of variance (MANOVA) tests were computed to identify the differences between High- and Low-Exam Groups across pre-course, exam, and within-course variables, with a significance level determined at $p < .05$. Pre-course variables included critical thinking, GPA, enrolled course hours for the current semester, and reported work hours. Exam variables included unit-exam scores, average exam scores across all units, and average exam scores following the first exam. Within-course variables compared were average quiz scores, average practice-exam scores, attendance rates, total homework completion, and total frequency

of comments made during class. Because data of several of the dependent variables analyzed were inconsistently measured across the four semesters included in Study 1, I separated MANOVA tests for each set of common dependent variables. Differences between High- and Low-Exam Groups on selected variables were first compared across the four semesters combined (Table 1), across three of the four semesters (Table 2), by academic year (2 semesters per year; Table 3), and by each semester separately (Table 4). The analysis produced a statistically significant difference between exam groups on the dependent variables measured ($p < .001$) across each time frame assessed.

Pre-course variables. Table 1 provides information on the differences between the Low-and High-Exam Groups for pre-course variables, including critical thinking and self-reported GPA, across the four semesters combined. Across 2 years' of data ($N = 280$), the High-Exam Group scored significantly higher on the critical thinking test than the Low-Exam Group [F (1,278) = 67, $p < .001$]. The Low-Exam Group's Critical Thinking Score mean was M = 24.11 (5th percentile), while the High-Exam Group's Critical Thinking Score mean was M = 29.35 (45th percentile). The High-Exam Group's mean GPA ($M = 3.46$) was significantly higher than the average Low-Exam Group's reported GPA, $M = 2.99$ [F (1,278) = 79, $p < .001$].

Information on reported course hours and hours of weekly employment were available for three of the four semesters analyzed (Spring 2010-Spring 2011), and differences of means between exam groups for these variables are presented in Table 2. While the means of critical thinking scores and reported GPA were significantly different by group, course hours and weekly employment hours were not. Reported enrolled hours were similar for Low- and High-Exam Groups (M = 15.24 and 15.37, respectively). Reported work hours per week was the only variable that yielded higher means for the Low-Exam Group ($M = 9.58$ hours) than the High-

Exam Group (M = 7.71). Although this is a notable difference, it was not significant at the $p <$ .05 level.

Course exams. Table 1 compares the mean scores across all unit exams, average scores for combined exams, and average exam scores following Unit A across the four semesters. Analysis of Table 1 suggests that, on average, the two groups respectively performed about the same on subsequent unit exams as they did on the first exam across units. The High-Exam Group performed significantly higher ($p < .001$) than the Low-Exam Group across each unit's exam, with F score ranges of F (1, 278) = 118 (Exam C) to F(1, 278) = 1478 (Exam A). The largest difference in means existed between the scores on Exam A (Low M = 30.94, equivalent to 61%; High M = 44.94, equivalent to 90%), while the smallest difference in means was between Exam C scores (Low M = 38.02, equivalent to 76%; High M = 43.86, equivalent to 88%). Unit C, with its emphasis on cooperative learning, was the only unit in which students were grouped heterogeneously to include low to high performers on the prior Exam B. The groups were instructed to work together and help one another master difficult concepts. Historically, low-performing students on previous exams generally improve their scores significantly on the Unit C exam (Carroll, Williams, & Hautau, 2006). Total average exam score for the Low-Exam Group was 34.85 out of 50 (equivalent to 70%), while average exam score was 43.68 (88%) for the High-Exam Group. Differences in average exam scores were significantly different [F (1,278) = 537, $p < .001$]. Average exam scores of Units B-E were also significantly higher [F (1,278) = 282, $p < .001$] for the High-Exam Group (M = 43.37, approximately 87%) than the Low-Exam Group (M = 35.83, equivalent to 70%).

Other within-course variables. Tables 1-4 also include comparisons of means between Low- and High-Exam Groups across other academic variables measured in the course. Some

variables (i.e., quiz average, practice-exam average) were consistently measured across semesters and are reported in Table 1. Other academic factors were assessed and recorded differently depending on the year and/or semester. These variables were compared either by academic year (i.e., attendance, homework completion) as displayed in Table 3, or by semester (i.e., participation frequency) as presented in Table 4. Although variables such as homework completion, participation, and attendance were reported slightly differently across semesters, each semester's variables represented quantity (e.g., number of items completed, frequency of attendance) as opposed to quality (e.g., accuracy of comments or homework answers).

Quiz means were significantly different between groups [F (1,278) = 68, $p < .001$], with an average of 3.70 out of 5 for the Low-Exam Group and 4.52 out of 5 for the High-Exam Group. During the semester, a practice exam was completed prior to each unit exam, and graded on a 5-point rubric. Average practice-exam scores for the High- and Low-Exam Groups were significantly different [F (1, 278) = 62, $p < .001$], with the Low-Exam Group's mean at 4.05, and the High-Exam Group's mean at 4.67.

Attendance rates and homework completion scores were graded differently during the 2009 and 2010 academic years and mean differences are presented in Table 3. While analysis of the Fall 2009-Spring 2010 academic year yielded significantly higher attendance rates [F (1, 135) = 20, $p < .001$] for the High-Exam Group, groups were not significantly different on this measure during the Fall 2010-Spring 2011 academic year. Similarly, the High-Exam Group had significantly higher homework-completion scores [F (1,135) = 22, $p < .001$] than the Low-Exam Group during Fall 2009-Spring 2010, but this significant difference between groups was not sustained during the following year.

Measurements used to assess participation varied by semester and differences between groups per semester are presented in Table 4. The number of comments made was significantly lower for the Low-Exam Group than the High-Exam Group ($p < .05$) for three of the four semesters analyzed. Across the four semesters, F scores ranged from F (1, 69) = 2 ($p = .210$) in Fall 2010 to F (1, 76) = 18 ($p < .001$) in Fall 2009.

Dropout rates. The overall course dropout rate for all students enrolled in the course in the last 2 years was 2.0% (15 individuals). Of the 15 individuals who dropped out between Spring 2009 to Spring 2011, 67% of the students earned a 34 or lower on the first exam. Rates of course dropout were compared across the identified High- and Low-Exam Groups. Approximately 6% of the Low-Exam Group (10 individuals) dropped out prior to the end of the course: 4 students dropped out prior to Unit D, 5 students dropped out prior to Unit E, and 1 student dropped out prior to the final exam. None of the individuals within the High-Exam Group dropped out during the semester that data were collected.

Discussion

Study 1 explored potential factors that lead to academic difficulty in a large-entry college course. Specifically, personal and academic patterns were compared between individuals who did well early in a large-entry college course ($n = 158$) and individuals who struggled early in this course ($n = 165$). While differences between low and high performers at the college level have been identified by previous researchers (see Galyon et al., in press; McCleary et al., 2011; Williams et al., 2003; Williams et al., 2004), this study focused on outcome differences across only those groups that did poorly or well on the first exam. By doing so, the current study aimed to provide evidence that students who initially struggle in a course are more likely to continue to have difficulty in the course compared to those who initially do well. By comparing groups,

Study 1 also aimed to identify performance variables that differed significantly between High- and Low-Exam Groups.

Previous studies conducted in the targeted course have identified several predictors of academic success in the course, including critical thinking, homework preparation, reading comprehension, vocabulary, and participation (Galyon et al., in press; Turner & Williams, 2007; Wallace & Williams, 2003; Williams, Oliver et al., 2003; Williams, Skinner, & Jaspers, 2007; Williams & Worth, 2002). This study adds to the current literature on predictors of course performance by connecting academic difficulty in a college course to initial exam performance. Findings of this study revealed marked differences between groups across numerous pre-course and within-course academic variables, including exam performance subsequent to Unit A, critical thinking, reported GPA, quiz performance, and practice-exam performance during the last 2 years.

Pre-course variables. Findings support the hypothesis that the means of several pre-course variables, such as critical thinking and GPA, would be significantly different between initially low- and high-exam performers. At the onset of the semester, students who did poorly on the first exam had, on average, lower critical thinking skills than those who did well on the first exam. This finding was to be expected, given that the multiple-choice exams in this course require students to critically apply unit concepts to answer items correctly. High critical thinking skills have previously been associated with the ability to analyze and organize information, problem-solve, effectively communicate information, and work well with others (Pithers & Sodden, 2000). In the college setting, critical thinking has been found to be a predictor of course success in a variety of courses (Williams et al., 2004). Although low critical thinking skills reduce the possibility of doing well in a course that emphasizes critical application of material,

some low critical thinkers are still able to do well. In Williams and Stockdale's (2003) study comparing high-performing, low critical thinkers and high critical thinkers, they found that individuals in the former group put forth better work habits (e.g., note-taking skills). Conclusions of their study were that low critical thinkers may need to both study harder and more effectively to be successful in the targeted course.

Students who did poorly also, on average, reported lower grades in their past courses than individuals who initially did well. Predictors of college GPA include willingness to study, amount and quality of study time, organization, persistence, time management, encouragement from parents, perceived competence of academic skills, and reading skills (Ransdell, 2001). A lower GPA may be related to poor study habits, low academic motivation, or disorganization. In combination, low-performing, low critical thinkers not only need to work harder than high critical thinkers to succeed in a course, but perhaps also need to be explicitly taught study habits and preparation strategies (Williams & Stockdale, 2003). Early identification and intervention services to these individuals allow opportunities for students to receive the help recommended by Williams and Stockdale.

Two pre-course variables that yielded little difference in means between High- and Low-Exam Groups were course and employment load. Both groups of students took about 15 hours of course work in the semester of data collection, suggesting that most students enrolled in this course are full-time students. Individuals in the Low-Exam Group reported higher employment hours per week than the High-Exam Group, but the difference was not statistically significant. It does not appear, based on these findings, that one group had significantly higher employment or course demands than the other.

Within-course performance. Within-course performance across several variables was also significantly lower for the Low-Exam Group. Differences suggest that, on average, those who did poorly on the exam at the beginning of the course continued to do poorly in exams and made lower scores on other academic variables. Compared to the other variables measured, subsequent exam performance yielded the largest differences in group averages. Given its weight in the course grade-wise (e.g., cumulative exam scores comprise 40% of the course grade), exam improvement would seem a likely top priority for students. However, students in the Low-Exam Group remained, on average, low performers across subsequent exams. Consistently low-exam performance might have accounted for the 10 individuals from the Low-Exam Group to withdraw from the course (67% of the total dropout rate). Reasons for low-exams grades are likely related to characteristics related to pre-course variables (e.g., low critical thinking scores, poor study habits), as well as poor performance across other within-course variables related to exam preparation (e.g., quiz, practice exam).

The Low-Exam Group also had lower practice-exam and quiz averages during the course. Similar to exams, scores on these variables evaluated the students' knowledge and ability to critically think about course topics. The quizzes required that students answer essay-type questions about course readings. Questions were provided prior to each quiz, so the students had notice as to what questions to be prepared to answer in class. Therefore, quiz scores relied on preparing information prior to the quiz and writing an organized, short essay that would completely and accurately answer the each of the possible questions. Analysis of results suggest that, similar to subsequent exam success, students in the Low-Exam Group struggle on preparing, memorizing, and organizing information to complete the quiz compared to the High-Exam Group. The practice exam, an open book, out-of-class assignment, requires students to

critically apply course knowledge. Low practice-exam scores suggest that students may have difficulty thinking critically about concepts, differentiating incorrect from correct answers, and accurately understanding course material when it is presented to them in class discussions and print material. Lower practice-exam scores could suggest that students in the Low-Exam Group did not allocate the same amount of time or effort to completing assignments as the High-Exam Group.

Several academic variables assessed in this study, including participation, homework completion, and attendance rates, were not consistently measured across both years studied, and therefore comparisons on these measures were categorized by year (attendance rates, homework completion) or by semester (participation). Multivariate analyses demonstrated variable differences between High- and Low-Exam Group means depending on the semester. The smaller and more unbalanced Low- and High-Exam Groups per semester might account for some of the inconsistent differences on some variables across semesters (e.g., inconsistent differences for attendance). For example, while the Low- and High-Exam groups were similar in number when semesters were combined, many more students who comprised the Low-Exam Group were from the Fall 2010-Spring 2011 academic year (N = 100) than the Fall 2009-Spring 2010 year (N = 52). Furthermore, variation in scores across these variables during the previous 2 years might be due to the different types of point contingencies used to encourage participation, homework completion, and attendance from semester to semester.

Future directions. Though the current study presents patterns of differences between High- and Low-Exam Groups, researchers should continue to study the relationship and impact of poor performance on the first exam on subsequent exam preparation, total performance in the course, and success in other courses. The current study aimed to identify patterns of

performance between two groups rather than infer causality for patterns. Two pre-course variables that did not yield difference between groups were course and work load. How the two groups balance work and course load might have affected exam performance. Although the number of hours enrolled in courses or working per week did not differ statistically, a comparison of such measures as time management, study habits, or locus of control might suggest differences between the two groups that would affect academic performance. To infer directionality of relationship, researchers could study the impact of targeting certain academic variables to see if an intervention produces eventual change on exam or course success.

Study 2 is a single-subject experimental study that aimed to enhance homework accuracy skills in individuals who have been identified by poor performance on the first exam. Although Study 1 did not investigate Study 2's target skill, homework accuracy, variables related to homework accuracy, including critical thinking, and other preparation variables (e.g., homework completion, quiz performance, and practice-exam scores) were assessed. To best determine if homework accuracy differs between early high- and low- exam performers, researchers should assess the accuracy of answers to homework questions.

While the Low-Exam Group had significantly lower scores across a number of variables compared to the High-Exam Group, no comparisons between the Low-Exam Group and those who earned average grades on the exam were made. Future researchers can compare group performance across a number of variables based on exam letter grades earned (e.g., compare those who failed on the first exam with those who earned a D, C, B, or A). Furthermore, researchers could identify patterns of differences between those who initially did poorly in the exam and those who did well on the first exam but had poor final marks in the course. It is

predicted that these two groups would have similarly low patterns of pre-course and within-course scores.

Although Study 1's participant pool included over 600 students, the majority of participants were Caucasian, female, and sophomores or juniors, with the intention of entering the teacher-education Program. To determine generalizability of results, researchers should look at the mean differences of academic variables across different student groups and courses. A specific group to study in the future is college freshmen. Implications from Study 1 include conducting further studies that incorporate the use of prevention and early intervention programs in a college setting. Therefore, a comparison of performance between first-year students who do well versus first-year students who do poorly initially might best inform intervention development for core-requirement classes.

Results came from data collected across 12 course sections of the same large-entry course. Analyses of results of this study, as well as results of previous studies conducted within the target course (e.g., Galyon et al., in press; Wallace & Williams, 2003), provide directions for the creation of an intervention program that fits the current course context. Although critical thinking, test-taking, and homework-accuracy skills predict exam success in this course, different variables might be higher predictors of performance in other courses. Before selecting and applying universal screeners, progress monitoring tools, and intervention procedures in a college course, past performance data of the specific course should be analyzed in a similar fashion as performance data in Study 1.

Major conclusions and implications. Findings of this study provide the framework for Study 2 by highlighting differences between initially High- and Low-Exam Groups and presenting patterns of academic variables of low-performing students. In similar courses, these

findings can assist teachers in working with students by alerting them to past predictors of exam performance. Instructors might also emphasize to the student that, in the past, initial exam performance has been related to subsequent and total exam performance. This information may help motivate students to put forth more effort in preparation for subsequent exams, as opposed to perceiving poor performance on the first exam as an isolated event.

Describing differences in performance between groups can help researchers designing interventions to choose which skills are most related to early exam performance. Study 1's analyses suggest that doing poorly on the first exam is not an isolated event but often is a precursor to lower scores on other academic measures. This pattern is consistent with past researchers' claims that test-taking success or failure is related to a combination of factors (e.g., Galyon et al., in press; Wallace & Williams, 2003).

Moreover, the findings that those who do poorly on the first exam continue to do poorly in the future exams, have lower scores on exam preparation grades (e.g., quiz, practice exam), and come to the course with lower critical thinking and grade records support the use of early intervention. Researchers should continue to identify the best universal screeners to pinpoint possible early academic problems in courses. Although early exam performance was a predictor of later exam performance in this study, it might not adequately predict subsequent course performance across courses. Inasmuch as researchers have reported the benefits of providing prevention and early intervention programs to reduce rates of academic failure (e.g., Slavin et al., 1992), college instructors should focus on finding measures that accurately identify who might struggle in a course within a timeframe that allows early intervention.

Summary. Study 1 provides information on the specific gaps in academic dimensions between initially low and high performers. Some differences were pre-existing (e.g., critical

thinking levels), while others pertained to outcomes within the course (e.g., exam and quiz performance). Analyses of results of Study 1 can help identify universal screeners to determine who needs intervention. Furthermore, understanding what factors differ between those who do well versus those who do poorly on the first exam can inform what variables should be targeted in an intervention program. For example, findings of Study 1 suggest that interventions could focus on test-taking strategies, ways to enhance critical thinking skills, and strategies to complete quizzes and practice exams. Researchers should continue to explore patterns of academic performance amongst those who struggle initially on exams in different courses to determine if generalizable trends apply across college settings.

Chapter IV

Study 2

With Study 2, I focused on improving overall exam and homework performance using a tier-based intervention program that included early identification of low performers and use of progress monitoring tools to measure response to intervention (RTI). A multiple-baseline design across participants was used to evaluate the intervention's effect on homework accuracy. I designed the selection and implementation of interventions based on several philosophies underlying the RTI model. Most consistent with the current study were RTI strategies that included providing research-based services to the entire classroom, assessing students' academic deficiencies by using universal-screening measures to select those in need of early intervention, providing a standard tier-based intervention, using progress monitoring measures to assess response to intervention, and making data-based intervention decisions when students do not respond adequately to the standard intervention. One to two interventions, selected on the basis of need and responsiveness, were implemented on an out-of-class basis. The first level of intervention focused on increasing test-taking and homework strategies by reviewing past exam performance and pinpointing areas to strengthen. If little to no response occurred to the first intervention, as determined through a combination of subsequent homework accuracy ratings and unit exam improvement, the intervention was intensified to include several tutoring sessions targeting overall understanding of key concepts outlined in the unit and methods to improve thinking critically about each concept.

With Study 2, I hoped to expand upon current RTI literature by evaluating the practicality of applying a standard treatment protocol that included RTI core concepts and recommendations in a large undergraduate teacher-education course. By doing so, I sought to find what RTI

components are feasible in the college setting and what areas need adjustment. Inasmuch as Galyon and colleagues (in press) have established that homework accuracy is a strong predictor of exam success, I hypothesized that homework accuracy would be an appropriate academic skill to monitor academic progress. Furthermore, I hypothesized that applying the core concepts and several recommendations of RTI at the college level will not only be feasible, but will also be an effective guide to improve overall understanding and critical application of course material.

Method

Participants. Study 2 was conducted in three course sections (Sections 1, 2, and 3) of the same undergraduate human development course as used in Study 1. Consent to participate in research was obtained from each section at the beginning of the semester. As was the case in Study 1, most of the students intended to enroll in the university's teacher-education program and class size averaged 54 students per section (Section 1- 56 students; Section 2- 54 students; Section 3- 52 students). Across sections, approximately 72% of the students were female and 28% were male (168 and 66, respectively). With respect to academic classification across sections, 5 students were freshmen (2%), 112 students were sophomores (47%), 70 of the students were juniors (29%), 33 of the students were seniors (14%), 8 of the students were graduate students (3%), and 2 students were non-degree seeking (< .1%). An additional 18 students (2%) did not indicate academic classification. The average reported GPA across sections was 3.22.

The number of intervention participants was expected to be low (between 10 and 20 students across all sections), as only those exhibiting significant problems on the first exam were asked to participate and participation was voluntary. A score of D or F on the first exam was Study 2's identified curriculum-based universal screener. Consistent with Study 1's

identification of low performers, students who scored a 34 or lower on the first exam were invited to participate in an exam improvement option. Although administrators who employ RTI in the schools typically make selections based on normative comparisons, I selected a criterion measure because students earning a C or higher on the first exam may not view their performance as low enough to pursue the exam-improvement option.

Across the three sections, 42 students were eligible and invited to participate in the study; 13 students from Section 1, 22 students from Section 2, and 7 students from Section 3. Approximately 24% (10 students) of those contacted participated in the study; 26% (11 students) expressed interest but were not able to attend the intervention sessions or did not follow up with scheduling; and 50% (21 students) declined to participate in the study. The ratio of the study's participants and students eligible varied across sections: Of the 13 students invited to participate from Section 1, 5 students elected to participate in the study (38% participation rate); of the 23 students eligible to participate from Section 2, 3 became participants (13% participation rate); and of the 7 students contacted from Section 3, 2 elected to participate (29% participation rate). Table 5 contains demographic and descriptive information of the 10 participants. For confidentiality purposes, the names of intervention participants in Study 2 have been changed.

Of the 10 participants, 3 participants (Kylie and Kathryn in Section 1 and Allison in Section 2) dropped out of the study prior to meeting the criteria to enter the maintenance phase. Each participant dropped out prior to beginning a more intensive tier of intervention. Two of the students, Kathryn and Allison, indicated interest in continuing with the study but were unable to schedule times to meet during the allocated time period.

Procedures. The intervention program implemented was modeled after RTI standard protocols. Figure 1 displays the flow of intervention in terms of participant selection, tiered

intervention, and progress monitoring. A multiple-baseline design across participants was implemented to determine effects of interventions and rule out intervening variables of setting (e.g., instructor, time in semester of initial implementation). Described below are the procedures implemented for Tier 1 (the general classroom), Tier 2, and Tier 3 phases. Included is information on the tools used to identify target participants and assess response to intervention, the implementation steps for Tier 2 and 3 interventions, and the performance criteria set for data-based decision making.

Tier 1. A tenet of RTI is that instruction strategies in the general setting are data-based and that some services are provided to the whole class. Tier 1, which consists of strategies provided within the classroom to all students, was conducted by three lead graduate teaching associates in the targeted courses. The overall structure of the current educational psychology course has been developed to include several evidence-supported practices, some of which are research-evaluated each year: stable and randomized credit contingencies to encourage attendance, homework completion, and participation in class discussion; immediate feedback on examinations and participation; the assignment of homework and inclusion of class discussion that targets thinking critically about core concepts; practice exams to help students prepare for unit examinations; the use of cooperative learning groups to provide assistance to peers, especially those performing poorly on the previous exam; short written assignments covering reading material; the expectation of accurate homework and participation; and organized, outlined Instructor Notes that included information synthesized from peer-reviewed journals.

There were several sets of material provided to each student related to exam preparation. Students received tips on how to do well in the course, outlined in their syllabus (e.g., review answers to discussion questions at the end of each unit and compare to instructor answers).

Students were also encouraged to arrange for help sessions with the instructor if they experienced any course-related difficulties. Each discussion day was audio-recorded and available for review at the course website. At the end of each unit, but prior to the exam, two documents related to the unit were emailed to students and posted on the course website. The first document was a set of complete instructor answers to all the homework questions. The second document was an Exam-Sources Document, which identified specific sources within course information for each exam item. Instructors encouraged all students to reference both documents in preparations for the exam.

Tier 2. Eligible students received notification from their lead teacher regarding the exam- improvement option, which included information on the potential benefits of participating (see Appendix A for the scripts of participant invitation to Tiers 2 and 3). Adherence to intervention guidelines was contingent on the student's willingness to participate in all sessions. Upon identification of participants for Tier 2, the primary researcher evaluated each participant's homework accuracy across select questions in the first unit and continued to assess homework accuracy in successive units.

Tier 2, provided to those individuals selected for the exam improvement option, was conducted by the primary researcher and commenced at differing times depending on the section in which students were enrolled. In order to provide direct instruction of test-taking skills, teach tactics for reading exam questions, and review how the homework material was connected to exam items, the primary researcher asked students to meet with her in small groups of approximately 2-4 students. Students were grouped based on their availability to meet and were asked to bring course materials including their Instructor-Notes handbook, the Exam-Sources document, and instructor answers to the discussion notes. The primary researcher brought the

tests and each student's scan forms to all sessions. Each meeting took place in a private room and lasted approximately 30 min. Only one session was scheduled per unit per individual.

To adhere to the standard treatment protocol, the primary researcher followed a procedural checklist when implementing Tiers 2 and 3 (see Appendices B and C). For Tier 2, the primary researcher asked the students to discuss what difficulties they experienced while taking the first test (e.g., problems with comprehension of questions, low-content mastery), as well as discuss how the students prepared for the exam. The primary researcher and the students then reviewed frequently missed exam items, identifying any gaps between knowledge of a certain concept and an exam item targeting that concept. Using the Exam-Sources Document and instructor answers to the discussion questions, the primary researcher demonstrated how the missed exam items corresponded to homework questions and instructor notes. Strategies for answering exam questions accurately were reviewed (e.g., understand what the question is asking, eliminate unsupportable options, and minimize choice of options with extreme terminology such as "always" or "never"). With remaining time, the researcher provided general tips for completing homework (e.g., answer each question completely and include all relevant concepts) and referred to references available (e.g., Exam-Sources document).

Progress monitoring. Improvement in content comprehension was determined by accuracy of daily homework across discussion-question items that were most representative of items on the upcoming exam. Inasmuch as Galyon and colleagues (in press) concluded that homework accuracy was a significant and, in certain situations, primary predictor of exam performance, I expected Daily Homework-Accuracy Score, to be the best representation of content mastery. Daily homework was assigned to all students on the four discussion days of each unit and graded by the primary researcher on percent of concepts accurately addressed in

student answers to each targeted homework question. Accuracy percentage was computed by dividing the number of concepts accurately represented in a student's response to a homework question by the number of concepts reflected in the official answer to that question.

Assessment of homework accuracy followed procedures similar to those conducted by Galyon and colleagues (in press). Accuracy scores were computed for 9 to 12 selected homework questions per unit (see Appendix D for list of questions selected by day and unit). Selection of the targeted homework questions was based on the two criteria outlined by Galyon and colleagues: the number of exam questions corresponding to a particular homework question and total number of discrete concepts covered within each homework-question answer. Homework questions relating to the most exam items and covering the most discrete unit concepts were selected for accuracy scoring. To provide highly accurate answers to discussion questions, students had to process information at the highest levels of Bloom's Taxonomic dimensions (1956), which included applying, synthesizing, and evaluating course material.

Evaluation of homework questions consisted of comparing the number of concepts accurately explained in a student's answer to the total number of concepts in the official answer to each selected homework question (Homework-Accuracy Score) (see Appendix E for a sample of the grading rubric adapted from Galyon et al., in press). A high Homework-Accuracy Score indicated that the student provided an answer that included a high percent of accurate core concepts represented in the official answer, while a low Homework-Accuracy Score indicated that very few concepts were accurately reflected in the student's answer. Homework-accuracy data were collected each discussion day and consisted of the average percent of accuracy and thoroughness across evaluated homework items completed by the students. In past research, Galyon et al. found that students seldom included all the concepts for each homework question;

consequently, a higher score simply indicated a more complete and accurate answer than a lower score on the same question. An omitted homework question was not recorded in the database, whereas an answer that included none of the rubric concepts was recorded as a zero for that item. Data were not collected on days that homework was not submitted.

Data-based decision making. Homework-Accuracy Scores were used to evaluate the effectiveness of the intervention implemented and provide a framework for decisions on student placement in the tiers. Following typical RTI procedures, decisions to keep the student in her or his current level of intervention, cease providing intervention aside from what is universally available, or to move the student to a more intensive intervention were contingent on the student's making adequate progress on benchmark assessments. The following criteria were used in making intervention decisions in this setting: (a) if there was no adequate improvement on overall accuracy of homework (e.g., the average increase is lower than 10% on the Homework-Accuracy Score compared to the student's homework average of the previous unit), then the participant was placed in a more intensive tier; (b) if an individual's Homework-Accuracy Score average was equal to or higher than 10% of the previous accuracy average but the student still earned a D or F on the present unit's exam, he or she remained in the current tier; (c) if the student achieved at least a 10% increase on homework accuracy from the previous unit and earned a C or higher on the next unit exam, then that student met the criteria to return to the general tier. Maintenance data on homework accuracy and exam performance were collected for any participant who successfully graduated back to the general tier.

Tier 3. If participants continued to show little to no improvement in exam and homework accuracy, they received more intensive instruction that targeted accuracy of homework. Students were asked to meet with the primary researcher twice during the current unit for a 45-min

session, once early in the unit (Day 1) and once prior to the last discussion day (Day 3). During these sessions, student answers to select homework questions were discussed and the instructor identified areas where the student's answers were incomplete or inaccurate. At this time, students were able to solicit instructor clarification about any question and/or concept within the homework. Goals of this intervention were to help the student be more thorough and accurate in answering questions and, as a result, gain an accurate and solid understanding of course material.

Given that homework questions required students to consider concepts critically (e.g., compare and contrast concepts, develop examples of concepts), students should be able to better understand exam items that also require higher-order thinking. Each intervention session adhered to a procedural checklist, and any additional questions covered at the request of the student were recorded (see Appendix C). Criteria for tier placement following the next exam were similar to the decision criteria previously used: the student remained in Tier 3 if homework accuracy did not improve, returned to Tier 2 if homework accuracy improved but exam performance remained low, or graduated back to Tier 1 if both homework accuracy and exam performance improved.

Research design. A multiple-baseline across subjects design was used to evaluate the effectiveness of the interventions, as measured by homework accuracy (See Figure 2 for flowchart of research procedures). Given that carry-over effects were expected once the intervention was withdrawn, a multiple-baseline design was considered a more appropriate design compared to other single-subject designs (e.g., withdrawal, reversal) for showing a cause-and-effect relationship between the independent and dependent variables (Barger-Anderson, Domaracki, Kearney-Vakulick, & Kubina, 2004). Although the criterion for selection was low performance on the first exam, the point at which individuals were asked to participate differed

depending on the section in which they were enrolled: students in Section 1 were invited to participate immediately following the first exam; those in Section 2 were invited to participate immediately following the second exam; and individuals from Section 3 were invited to participate immediately following the third exam. Performance on the Unit A exam was the only screener for participant eligibility. Course sections were identified at random prior to the Unit A exam.

Consistent with the multiple-baseline design, this staggered application of RTI determines if improvement on homework accuracy was likely due to the onset of the intervention. Baseline for participants in Section 1 was the average of each Homework-Accuracy Score per day on homework items selected in the first unit (4 data points per unit). Baseline for participants in Section 2 was the Homework-Accuracy Scores for items across the first two units (8 data points pending completion of each homework item assessed). Lastly, baseline for participants in Section 3 was all scores from the first three units (up to 12 data points). Although length of time in each intervention varied per individual responsiveness, each participant had at least 4 data points represented during the Tier-2 intervention. Furthermore, each participant had opportunity to receive Tier-3 instruction if they did not respond to Tier-2 intervention. Baseline data following Unit A were assessed for final selection of research participants. If the baseline was declining or stable in trend, participant data were included. If the baseline data were increasing in trend, treatment was still offered but the student was not included in the study.

A time-series graph was constructed to display data of Homework-Accuracy Scores across baseline, intervention, and maintenance (when available) phases for each individual (see Figure 3). Visual and trend analyses were used to interpret these graphed data by examining

changes (e.g., immediacy, trend, level, non-overlapping data points) across individual phases. Visual inspection was used to interpret data across multiple baselines.

Inter-observer agreement. In the current study, a total of 52 questions were evaluated for homework accuracy, with a range of 9 to 12 questions chosen per unit (9 questions for Unit A, 9 questions for Unit B, 11 questions for Unit C, 12 questions for Unit D, and 11 questions for Unit E) (see Appendix D). The mean number of questions rated per day was 2.6 (mode of 2 questions daily), with a range of 1 to 4 questions per day. In a previous study, Galyon and colleagues (in press) reported 84% inter-rater agreement for overall tabulation of exam-homework correspondence and 78% agreement for concepts covered in the homework answer. Many targeted homework questions were previously selected by Galyon and colleagues, with modifications made to accommodate new course material. New questions were selected by both the developer of the original Homework-Accuracy rubric and the primary researcher of the current study. Discussion questions selected corresponded highly with exam items and contained many unit concepts (see Appendix E for sample grading rubric).

Inter-rater agreement data pertaining to evaluation of the Homework-Accuracy Scores were collected following previously delineated procedures (see Galyon et al., in press). Prior to inter-rater data collection, the developer of the homework accuracy rubric taught grading methods to the primary researcher and the comparison rater, who was an undergraduate research assistant familiar with the course material. Consistency of rater scores was checked for a non-participating student's homework in one unit. In this current study, participant Homework-Accuracy Scores on individual items were evaluated for inter-rater agreement for approximately 27% of the questions selected per unit (14 out of the total 52 questions per participant evaluated). The number of questions rated by the comparison rater ranged from 2-4 questions per unit.

Percent inter-observer agreement was calculated by dividing the number of evaluated questions in agreement by the total number of questions rated and multiplying by 100. Inter-observer agreement averaged 85% across the 5 units, with agreement ranges between 70% in Unit B and 93% in Unit E.

Treatment integrity. Prior to intervention implementation, the primary researcher rehearsed both Tier 2 and Tier 3 intervention procedures with two graduate teaching assistants. The primary researcher conducted a total of 18 intervention sessions; 12 were the 30-min Tier 2 sessions and 6 were the more intensive, 45-min Tier 3 sessions. Across units, the primary researcher conducted four Tier 2 sessions in Unit B, five in Unit C, two in Unit D, and one Tier 2 session in Unit E. These sessions, designed for small groups, generally had 1-2 participants in attendance. Three participants met the homework criterion for the Tier 3 intervention. Each participant met for two sessions per unit: two participants met individually with the primary researcher twice in Unit D and one participant attended two sessions in Unit E.

Treatment integrity was monitored by the primary researcher's completing one of two self-designed checklists immediately following each session (see Appendices B, C). The Tier 2 checklist included 7 steps and the Tier 3 checklist included 5 steps. Checklists completed by the primary researcher revealed that she implemented 100% of the steps correctly on all 18 sessions.

To further assess treatment integrity, each intervention session was audio-taped. An independent observer randomly picked and listened to approximately 33% of the total intervention sessions (three Tier 2 sessions and three Tier 3 sessions) and used the same checklists to rate the researcher's adherence to the standard procedure. Checklists completed by the independent observer, an undergraduate research assistant familiar with the course material,

confirmed that the primary researcher implemented 100% of the steps correctly on the six recorded sessions.

Participant acceptability. Student acceptability data for Tier 2 and Tier 3 were collected in a 6-item Likert survey designed by the primary researcher and given to participants at the end of each unit prior to the unit exam (see Tables 6 and 7). Possible responses ranged from 1 (*Strongly Disagree)* to 6 (*Strongly Agree*). Surveys were intended to be anonymous and could be turned into the student's teacher on the day of the exam. Seven surveys were turned in at the end of the semester: six rated the Tier 2 intervention and one rated the Tier 3 intervention. Overall, participants' answers were favorable towards both interventions and perceived them as helpful. The numbers of Tier 2 intervention surveys turned in across units were two prior to Unit B's exam from Section 1, two prior to Unit C's exam from Section 2, and two prior to Unit D's exam from Section 3. Averages and ranges across the six surveys completed for Tier 2 were computed for each question and presented in Table 6. Answers across questions averaged 5.1 (*Agree*) and ranged from 4.5 to 5.83 out of 6. The highest endorsed item referred to using exam-taking strategies discussed during the session prior to the next exam (M = 5.83). Participants also rated the item, "I will use these study and test-taking tips in my other classes" highly (M = 5.3), suggesting that they perceived the intervention's test-taking tips learned as generalizable. Following the intervention, one individual from Section 3, Stephanie, contacted the primary researcher to say that the Tier 2 intervention helped her greatly during the exam.

Answers to the Tier 3 survey completed by one participant from Section 1 prior to Unit D's exam are presented in Table 7. Answers ranged from 4 (*Slightly Agree)* to 6 (*Strongly Agree*), with a mean of 5.3 out of 6. The participant reported that the intervention changed how she completed her homework and helped her prepare for the exam. She indicated that she would

use these preparation tips in her other classes and would recommend this intervention to others struggling on exams.

Results

Changes in Daily Homework-Accuracy Scores, the target dependent variable for Study 2, were interpreted at both a group (Sections 1, 2, 3) and individual level. Daily Homework-Accuracy Scores were analyzed via visual analysis across baseline and treatment phases, comparison of phase mean levels, and percentage of data points exceeding the median of the baseline (PEM; Ma, 2006). Non-overlap indices such as PEM have been considered more robust measures of shift between phases than mean or median level analysis in single-subject designs alone (Parker, Vannest, & Davis, 2011). This measurement index was selected due to the high sensitivity to potential outliers within the progress monitoring measure, Daily Homework Accuracy. To determine PEM, the percentage of treatment phase data above the baseline median level is computed (Ma, 2006). Ma (2006) recommends interpreting PEM scores according to the interpretation criteria for effect sizes suggested by Scruggs, Mastropieri, Cook, & Escobar (1986): a PEM score over 90% is considered highly effective, a PEM score between 70% and 90% suggests that an intervention is moderately effective, and a PEM score less than 70% is considered questionable or not effective. A PEM score was calculated to determine effect of the intervention across groups and individually.

Across- and between-group data analysis of homework accuracy. Table 8 provides average homework accuracy between baseline and treatment phases. Baseline Homework-Accuracy data were variable across Sections 1, 2, and 3, with average baseline scores ranging from 12% to 27% (Section 1- 12%; Section 2- 27%; Section 3-15%). Combined mean across sections was an average Homework-Accuracy Score of 18% in baseline phases, with the average

accuracy increasing to 34% in treatment/post-treatment phases. Data in the treatment phase also showed variability in homework scores, ranging from 23% to 44% across sections. Table 9 provides information on range of Homework-Accuracy Scores within phases across individuals and sections. Across all three sections, Daily Homework-Accuracy Scores had some variability during both phases, with daily percentage ranges between 30 (Section 1) and 59 (Section 2) during baseline phases and percentage ranges between 35 (Section 2) and 75 (Section 1) in treatment/post-treatment phases.

Figure 3 displays the Homework-Accuracy Scores between baseline and treatment phases across sections of target students. Visual analysis of figures revealed some variability within baseline and treatment phases, particularly for Section 3. Due to this variability across groups and individuals, an analysis of mean levels was conducted. Comparison of mean levels between phases showed higher mean levels in treatment phases than baseline phases for each section (see Figure 3). While Sections 1 and 2 both had substantial increases in treatment/post-treatment phases (an increase of 22 and 17 percent points, respectively), individuals in Section 3 had a smaller change between phases (an increase of 8 percent points). PEM scores across group sections ranged from 62.5% to 100% (Section 1- 93%, Section 2- 100%, Section 3- 62.5%) (see Table 8). This pattern suggests that, as a group, the intervention was highly effective for Sections 1 and 2, and slightly effective for individuals in Section 3, who received the intervention only in Units D and E of the course.

Table 10 presents information on the mean and range of Homework-Accuracy Scores across course units for individuals in Sections 1, 2, and 3. The multiple-baseline design staggered time of treatment implementation based on course unit; individuals in Section 1 started the tier treatments in Unit B, Section 2 participants began intervention in Unit C, and Section 3

participants started the intervention in Unit D. All three sections had higher average Homework-Accuracy Scores in each of the units during and following treatment than in baseline phases. Gains between baseline and treatment unit ranged from a 4% gain (Section 3) to a 37% gain (Section 1)

Within-group data analysis of homework accuracy. As expected in an RTI study, students responded to the interventions differently. Individual data were collected and analyzed for each participant across Sections 1, 2, and 3. Figures 4 through 9 present results of each individual's Daily Homework-Accuracy Average across baseline, treatment (i.e., Tier 2 and Tier 3 phases), and, when applicable, maintenance phases. Four of the 7 participants (Kristin in Section 1; Emma and Elizabeth in Section 2; Stephanie in Section 3) made substantial gains in homework accuracy following the Tier 2 intervention, and met the exam criterion to go to the maintenance phase after one course unit of treatment. The remaining 3 participants (Edith and Sally in Section 1; Becky in Section 3) required a more intensive tier. Two of the participants, Edith and Sally, initially made substantial improvements under Tier 2 but did not sustain this improvement in the consecutive unit. Becky failed to respond to the Tier 2 intervention based on the primary researcher's criteria to determine progress, but made gains in homework accuracy during Tier 3.

As indicated in Figures 4 through 9, 6 of the 7 participants made an immediate increase in homework accuracy following treatment implementation. Tables 11 through 15 provide additional information on individual and group Homework-Accuracy percent and range for each of the 20 homework days within the study. Visual analysis of mean levels in Homework-Accuracy Scores shows that mean levels were higher following implementation of the Tier 2 intervention for 6 of the 7 participants. The 3 participants who received the Tier 3 intervention

made immediate increases in homework accuracy. Interpretation of PEM scores suggests that the tiered intervention was highly effective for 4 of the individuals, moderately effective for 3 of the individuals, and not effective for one individual.

Table 9 presents mean and range data across baseline, Tier 2, Tier 3, and maintenance phases for each student and section. Average baseline Homework-Accuracy Scores ranged from 7% (attained by Sally of Section 1) to 27% (attained by Emma of Section 2). Average Homework-Accuracy Scores ranged from 7% (Becky of Section 3) to 39% (Stephanie from Section 3) during Tier 2 and ranged from 21% (Becky) to 51% (Edith) during Tier 3. Maintenance data, which were collected for 71% of the participants, ranged from 26% (Stephanie, Section 3) to 50% (Elizabeth, Section 2).

Section 1. Section 1 participants were the first to receive the tiered intervention. Three of the 5 original participants completed the tier program during the semester. Substantial gains were made by this group between baseline and treatment phases, with only two overlapping data points between phases (See Figure 3). The PEM score for Section 1 was 93%, considered highly effective. The section below includes results for each participant in Section 1.

Edith. Figure 4 displays Edith's Daily Homework-Accuracy Scores across baseline, Tier 2, Tier 3, and maintenance phases. Visual analysis of Figure 4 found that Edith made the most improvement in her homework accuracy in Tier 3, and was able to maintain a higher Homework Accuracy Mean Score during her maintenance phase than baseline and Tier 2 phases. Following the Tier 3 intervention, Edith only had two overlapping data points, with her highest point in baseline. Edith's overall PEM was 73% between baseline and treatment phases, which is considered moderately effective.

During the semester, Edith received two Tier 2 sessions (across Units B and C) and two Tier 3 sessions (during Unit D). Following the first Tier 2 session, Edith made immediate and substantial gains in her homework accuracy on the first two treatment days. For example, she had an increase of 40 percentage points on Day 7 compared to her baseline mean. Due to Edith's Unit B exam score, she continued receiving the Tier 2 intervention. During this unit, Edith's homework scores fell below baseline levels. Due to a decrease in Homework-Accuracy Scores during this unit, Edith was asked to participate in Tier 3 interventions. Edith came well prepared to discuss selected homework items for both Tier 3 sessions, having all her questions completed prior to meeting. Edith's Homework-Accuracy Scores rose immediately and substantially in this unit (see Figure 4). Her homework and exam performance met criteria for her to continue to maintenance phase, where her mean Homework-Accuracy Score level decreased slightly but remained higher than baseline and Tier 2 levels.

Kristin. Kristin only met with the primary researcher once for a Tier 2 session before she met the homework and exam criteria to continue in maintenance. Figure 5 presents Kristin's performance across baseline, Tier 2, and maintenance phases. Kristin's homework accuracy rose immediately and substantially following the first day of treatment. Kristin scored highly on the first homework day during baseline (44%), and she had several data points overlapping with this highest baseline point during treatment and post-treatment phases. However, Kristin's PEM was 100%, considered highly effective. Kristin continued to make gains in homework accuracy in her maintenance phase, although her performance was slightly more variable than in Tier 2.

Sally. Figure 6 displays Sally's Daily Homework-Accuracy Scores across baseline, Tier 2, Tier 3, and Tier 2 phases. Visual analysis of Figure 6 suggests that, like Edith, Sally made the most improvement in her homework accuracy in Tier 3. Across all treatment phases, Sally's

PEM was 100%, considered highly effective, and she only had one overlapping data point with her highest baseline score. Visual analysis and the high PEM percentage suggest that the tier interventions improved her homework accuracy from baseline levels. During the semester, Sally attended three Tier 2 sessions (across Units B, C, and E) and two Tier 3 sessions (during Unit D). Figure 6 shows that, following the first Tier 2 session, Sally made immediate and substantial gains in her homework accuracy across the first four homework days in the phase. Due to Sally's Unit B exam score, she continued receiving Tier 2 intervention. Across Tier 2, Sally had a large range of Homework-Accuracy Scores (0 to 74) with a mean of 32% (See Table 9). Due to a decrease in Homework-Accuracy level during the last three days of Unit C (Days 10-12), Sally began Tier 3 interventions. Sally attended two Tier 3 sessions in Unit D, and came prepared to discuss most homework questions. Compared to Edith, Sally came in with a lower understanding of several of the unit's concepts. Sally's Homework-Accuracy Scores rose immediately and substantially during Tier 3. Due to her low Unit D exam score, Sally returned to Tier 2 interventions for the final four days of the study. Her homework performance during this phase returned to approximate equal mean levels of the first Tier 2 treatment phase. Compared to Sally's first Tier 2 phase, her homework performance in the Tier 3 and final Tier 2 phases was much less variable.

Section 2. Three students originally comprised Section 2, with one participant withdrawing from the study after one unit. Similar to Section 1, the participants in Section 2 made gains in the treatment phase on average, but had more variability across both baseline and treatment phases. Below is the individual homework performance of individuals in Section 2. The PEM score for Section 2 was 100%. Both Section 2 participants, Emma and Elizabeth, presented similar trends across phases and therefore are described together.

Emma and Elizabeth. Both Emma and Elizabeth met with the primary researcher as a pair once for a Tier 2 intervention in Unit C before meeting exam and homework criteria to begin maintenance phase (See Figures 7 and 8). Emma and Elizabeth's baseline performances were largely inconsistent, with a baseline range of 58 and 60 percentage points (See Table 9). Figures 7 and 8 show that no data points in Tier 2 and maintenance phases overlapped with the baseline mean level line of 27% for either participant. Both individuals increased their Homework-Accuracy mean levels following treatment phases from an approximate average mean level of 26.5% in baseline to 37% accurate following the Tier 2 intervention session. Moreover, both participants continued to make improvements in Homework-Accuracy Scores in maintenance; Emma had a mean level of 44% accurate and Elizabeth had a mean level of 50% accurate across questions during the maintenance phase. Compared to the high variability in baseline, scores within Tier 2 and maintenance phases for both participants were more consistent (see Figures 7 and 8).

Section 3. Section 3 participants, invited at the beginning of Unit D, were the last group to receive intervention services. Compared to Section 1 and 2, Section 3 made much smaller gains in treatment phases. Section 3's PEM was 62.5%, which is considered questionable; however, individual performance showed different trends in homework accuracy between Becky and Stephanie.

Becky. Figure 9 displays Becky's Daily Homework-Accuracy Scores across baseline, Tier 2, and Tier 3phases. Visual analysis of Figure 9 shows that Becky made the most improvement in her homework accuracy in Tier 3 and had the lowest performance in Tier 2. Baseline data show much inconsistency in homework accuracy across days. For example, she received a score of 0% accurate on Days 1, 5, 7, 9, but scored higher than 20% on Days 4, 8, and

10 within baseline. Following implementation of the Tier 2 small-group intervention in Unit D, Becky's scores across 3 of the 4 days in the phase remained low (a 5% or lower accuracy). Due to her decrease in homework performance, Becky was asked to participate in the more intensive intervention. During these Tier 3 sessions, Becky reported that she was experiencing much stress due to many obligations with her other classes and extracurricular activities. She appeared rushed and unprepared at times while meeting with the primary researcher. For instance, she would come to the sessions without completing many of the questions that would be discussed in the session. The primary researcher noted that several of Becky's homework questions would be omitted when turned in, even after discussing homework concepts during the Tier 3 intervention sessions. Becky appeared to respond to the Tier 3 intervention, which had a mean increase level of 14% points from the Tier 2 phase and 10% points from baseline. She made immediate gains following the implementation of Tier 3 and responded more consistently to homework answers during this phase of treatment. The PEM score for Becky was 37.5%, which classifies the intervention as questionably effective. Conversely, when analyzing Tier 3 in isolation from Tier 2 in Figure 9, Becky had no data points overlapping with her mean and median baseline levels (11% and 7%, respectively).

Stephanie. Figure 10 presents Stephanie's Homework-Accuracy performance across baseline, Tier 2, and maintenance phases. While Becky made little improvement in homework accuracy following Tier 2 compared to baseline, Stephanie made substantial gains in homework accuracy after the Tier 2 intervention was implemented. Compared to Stephanie's mean level during baseline, her average Homework-Accuracy Score rose by 20%. Stephanie's homework and exam performance following Unit D met criteria to stop intervention and move to maintenance. During maintenance phase, Stephanie's homework performance decreased, but

remained higher than her baseline mean level. Stephanie's performance across phases remained variable. During treatment/post-treatment phases, Homework-Accuracy Scores on Day 16 (Tier 2) and Day 19 (maintenance) were notably lower than the six other days within these phases. Stephanie's overall PEM was 75%, which Ma (2006) would classify as moderately effective.

Dropout participants. Figure 11 provides the Homework-Accuracy Scores of the 3 participants, Kathryn, Kylie, and Allison, who withdrew from the study. Upon departure from the study, their data were removed from group Homework-Accuracy totals. Kylie (Section 1) and Allison (Section 2) both stopped meeting with the primary researcher following one Tier 2 session. Although both individuals had higher average gains in homework accuracy during treatment than baseline, the lower level of improvement (i.e., less than 10%) met criteria for Tier 3. While Allison expressed interest in continuing to attend sessions, she was unable to schedule a time to meet. Kathryn attended two Tier 2 intervention sessions across Units B and C. During Unit B, Kathryn made immediate and substantial gains in homework accuracy, with an initial increase of approximately 75% between Days 4 and 5. Because of low-exam performance following Unit B, Kathryn continued receiving the Tier 2 intervention. Her Homework-Accuracy mean across days in Unit C did not improve compared to the preceding unit but remained higher than baseline mean levels. She was invited to participate in Tier 3, but was unable to continue with the sessions due to scheduling conflicts. PEM scores were 33%, 75%, and 86% for Kylie, Allison, and Kathryn, respectively.

Changes in exam scores. Insomuch as Galyon et al. (in press) reported homework accuracy to be a strong predictor of exam performance, I expected that by targeting homework accuracy as a dependent variable, exam performance would also improve. Table 16 displays group and participant data on exam grade averages and deviations from course section mean

exam grades between baseline and treatment phases. Combined, exam points across all 7 participants averaged 32.7 points out of 50 (a grade of 65.4%) in units during baseline phase and exam averages increased to 37.5 points (75%) in the units occurring during treatment phases. Percentage increases of exam scores between baseline and treatment phases approximated +8% in Section 1, +11% in Section 2, and +10% in Section 3. Table 17 presents participant exam grades for Units A-E course exams, as well deviations from class means. Across units, participants generally performed highest on the Unit C exam, which is consistent with class trends. Compared to other units, cooperative learning was emphasized during this unit, and bonus points were offered as incentives for individual and group improvement on the Unit C exam.

Patterns of changes in exams scores were largely heterogeneous across participants by unit: while some participants were able to make noticeable gains following Tier 2 interventions (Kristin, Stephanie, Emma, Elizabeth), others made questionable gains or losses in exam scores (Edith, Sally, Becky). Of the 7 participants, 5 individuals made improvement in exams following the treatment phase: Emma and Elisabeth made an overall improvement of +4.1 points (an 8.2% grade increase) following baseline, while Kristin and Stephanie made gains averaging +9.4 points (an 18.8% grade increase). Edith increased her average exam grade minimally (+0.5 points; 1% grade change). Conversely, Sally and Becky both showed slight drops in exam performance, averaging -1.25 points (-2.5% grade change) following baseline. A comparison of changes in homework accuracy performance to changes in exam grades revealed that those requiring the more intensive, Tier 3 intervention (Edith, Sally, and Becky) made little improvement on exams following baseline; whereas individuals who showed immediate

improvement and maintained responses in homework following the initial Tier 2 intervention also made notably higher increases on exams scores during and following treatment phases.

Section 1. As a group, the participants' average baseline exam score (Unit A) was 32 out of 50 (64%), which deviated from the class average by -7 points, or 14% lower than average. Following treatment, exam scores increased to a range of 35-38 points out of 50 across unit exams, with an average score of 36 points (72%). The gap of participant exam scores with class means decreased from -14% in Unit A to -4% in Units B and D and -6% in Units C and E.

Across participants in Section 1, Kristin made the largest gains on unit exams following treatment. Kristin's exam scores between Unit A, occurring in baseline phase, and Unit B, her first unit following treatment, rose 22% (66% in Unit A, 88% in Unit B). Her exam scores deviated from class mean by -12% in Unit A to +14% in Unit B. Following the Unit B exam, Kristin entered her maintenance phase. During this phase, Kristin's scores ranged from 39 to 44, with an average of 41.3 out of 50, or an 82.7% grade across exams. Deviation from class average scores in maintenance ranged from -4% to +12%.

Conversely, Edith and Sally's test scores did not show much improvement between baseline, Tier 2, Tier 3, and maintenance phases. Edith's Unit A exam score was a 33 out of 50 (66%, deviation of -12% from class means), while her average exam scores following the treatment phase was a 33.5 out of 50 (a grade of 67%). Edith's exam scores while receiving the Tier 2 intervention were 32 in Unit B (64%, deviation of -10% points from class means) and 35 in Unit C (70%, deviation of -12%). Following the Tier 3 intervention, her exam score remained at 35, although the gap between her score and the class average decreased to -6%. During maintenance phase, her exam score was 32 out of 50 (64%). Sally's exam scores within the treatment phase ranged from a 26 to a 38 out of 50, with an average exam score of 30.3 (a grade

of 60.5%). Unpredictably, the exam following the most intensive tier, Unit D, was Sally's lowest exam grade (26 out of 50, or 54%).

Section 2. As a group, participants in Section 2 increased their average exam score from a mean score of 31.5 (63%) during baseline (Unit A and B) to an average score of 37 out of 50 (74%) across treatment phases (Units C, D, and E). The gap between Section 2 exam performance and class averages decreased from a mean of -7% in Units A and B to a 0% difference across Units C, D, and E.

Between the two participants in Section 2, Emma showed the most consistent gains. Emma's exam scores during baseline averaged 33.5 out of 50 (67%) and increased to an average of 37 points (74%) across treatment and maintenance phases. Compared to class averages, Emma's deviations in scores from class means changed from -3% in baseline to +1% in treatment phases. She made her highest exam score (39 out of 50; a grade of 78%) in the unit that Tier 2 was implemented. Her scores on exams during the maintenance phase (a 36 in Unit D and a 38 in Unit E) were slightly lower, but remained higher than baseline. On average, Elizabeth's exam scores were a mean of 31 out of 50 (62%) in baseline phases and 35 out of 50 (70%) following treatment. Elizabeth also made her best score in the unit that Tier 2 was implemented (40 out of 50, or 80%) and made lower grades during maintenance phase (Unit D-68% and Unit E-62%). Deviations from class means decreased from -8% on exams in baseline to -3% in treatment phases.

Section 3. Section 3 had the highest average baseline exam score of the three sections, with an average of 34.7 out of 50 (69.3%). During baseline units, Section 3's deviation from class means ranged from -16% to -8% (M = -11%). Section 3 participants also made the largest gains in exam scores during treatment compared to the other sections, with an average of 39.5

out of 50 (79%). Compared to their class average, deviations of exams occurring in baseline suggest no exam score gap.

Becky and Stephanie performed differently across baseline and treatment exams. Becky's average exam score during baseline was 36 out of 50 (72%), with a range of 64% in Unit B to 88% in Unit C. Her average exam score during treatment was 34.5 out of 50 (69%). Deviations of exam scores from class means averaged -8% in units during baseline and -10% in units following treatment. Becky's score during Unit C is noticeably higher than her other exam scores. This is likely explained partly by the cooperative learning incentive. Becky's exam scores during treatment phase were higher than the first two units in baseline. Comparatively, Stephanie made substantial gains between baseline and treatment phases. During the three exams occurring in baseline, she attained an average exam score of 33.7 out of 50 (67.3%), with an average deviation from class percent means of -13%. Following treatment condition, Stephanie's exam scores averaged 43.5 out of 50 (87%). Compared to the class average, these scores were about 8% higher. Stephanie's scores ranged from 43 out of 50 (86%) during maintenance phase to 44 out of 50 (88%) following the Tier 2 intervention.

Dropout participants. Table 16 also describes the outcome of the 3 participants who received some help during the semester but elected not to continue with the sessions throughout the semester. Although Kylie's scores do not show sustained improvement in homework accuracy in the unit that she attended a Tier 2 intervention session, she increased her overall exam score from a 62% in baseline to an average of 82% in units following treatment. Allison also made substantial gains between baseline and treatment phases, with an average of 59% between her first two tests and an average 74% during treatment phases. Kathryn made much smaller gains in exams between baseline (64%) and treatment phases (67%).

Comparison of homework performance between participants and non-participants.

Table 18 provides the means and standard deviations of Study 2's participants, eligible students who elected not to participate (non-participants), and individuals comprising the Low-Exam Group from Study 1 (e.g., individuals who earned a 34 or less on the Unit A exam during the last 2 years in the course). Inasmuch as the participants who dropped out of the study still attended one or more Tier 2 intervention sessions, they were also included in the Study 2 participant group for exam comparison. The means of the Study 2 participants and these two groups (Study 1's Low-Exam Group and Study 2's non-participants) were computed. Between the participants and the non-participants, performance on Unit A was the closest, with a difference of 0.19. The exam means were higher for participants across Units B-E, with a difference between groups range of .97 (Unit C) to 6.04 (Unit B). The average test score for non-participants was 33.12 out of 50 and the average test score for participants was 35.26. The average total test score was 17.24 points higher for participants (M = 176.30) than non-participants (M = 159.06).

Comparison of means between participants in Study 2 and the Low-Exam Group showed closer exam performance across units (a range difference of -.19 to 2.19), with the participant group having slightly higher means for each unit besides Unit E (-0.19). Participants had an average of 9.33 points higher on their total test score than the Low-Exam Group (M = 166.97).

As far as exam success, the test average for Units B-E for individuals who did poorly on the first exam during the last 2 years was 35 out of 50 (70%). For Study 2, students eligible to participate who elected not to participate had an average test score across Units B-E of 34 out of 50 (68%). Ten out of the 31 students (32% of group) earned an average grade of C or higher on Units B-E (M = 74%), while 21 out of 31 (68%) students earned an average grade of D or lower across Units B-E (M = 64%). Study 2's participants had an average of 36 out of 50 (72%) across

Units B-E. Six out of the 10 individuals earned a C or higher across Units B-E (M = 78%), while 4 out of 10 earned a D or lower (M = 66%). As a group, Study 2's participants earned about 13 more points across units B-E than non-participants (see Table 18).

Table 19 displays the differences in exam grades between individual participants in Study 2 and eligible non-participants from the current semester. Inasmuch as the non-participants had the same teachers, resources, and course material during the semester as the participants, I suspected that this would be a better comparison than the Low-Exam Group. Differences between exam scores varied greatly across participants. Average differences in exam grades per units for individuals who received treatment compared to those who did not were as follows: Unit B--M = 12% higher than non-participators (range = +1% to +30%); Unit C--M = 1% higher than non-participators (range = -7% to + 11%); Unit D--M = 8% higher (range = -15% to +21%); Unit E--M = 4% higher than non-participators (range = -15% to +17%). Kylie, Kristin, and Stephanie had the highest differences between class means following treatment, with a difference as high as 30% (Kristin, Unit B). Emma also performed higher than the comparison group in treatment, but had similar differences from the class means during baseline as well.

Demographics and dropout rate. Overall differences in patterns of outset demographic variables (i.e., GPA, critical thinking, academic classification) and within-course variables (i.e., attendance rates, participation, quiz performance, practice exam performance) were not detected between individuals who responded well to tiers and/or made improvements on exams following treatment (e.g., Kristin, Stephanie, Emma, Elizabeth) and those who showed more variable exam performance or accuracy of homework responses (e.g., Edith, Sally, and Becky). During the semester Study 2 took place, 3 individuals across the morning sections dropped out of the course.

Two of these individuals earned a D or lower on the first exam, but neither students elected to participate in the second study.

Discussion

Previous researchers have provided evidence to support the use of RTI practices at the elementary levels (Burns et al., 2005), and have recommended its use at the secondary level (Burns, 2008; Shinn, 2008). This study expands current literature on RTI practices by describing a methodology and results of a tier-based intervention following RTI core principles catered to individuals at the post-secondary level. A multiple-baseline design was implemented to assess the effects of an out-of-class intervention program on homework accuracy in a college course across seven participants. A multi-level intervention system was developed and implemented based on the RTI standard protocol model (i.e., use of research-based academic services in the general classroom, universal screeners, standard tier-based interventions, and progress monitoring measures to determine response to intervention and make placement decisions). The focus of the two interventions, though both emphasized completing homework accurately and completely, were slightly different. The goal of the first intervention, intended for small groups, was to increase knowledge of exam and test-taking strategies. The goal of the more intensive, one-on-one intervention was to increase understanding of core concepts, which were targeted in homework assignments and evaluated across five course exams.

Homework accuracy. Overall, analyses of the study provide evidence that an intervention program following RTI core concepts is not only feasible at the college level, but can lead to improvements in homework accuracy. Visual analyses of results were variable day to day, but show higher mean increases in Homework-Accuracy Scores for 6 of the 7 participants between baseline and treatment phases. The multiple-baseline design between participants

reveals the intervention's effect on homework accuracy when the intervention's initial implementation is staggered across the semester. Across tiers, there was some variability of homework accuracy, but higher mean levels were met during Tier 2 and Tier 3 than in baseline. Indices of effect, measured by percentage of data points exceeding the median (PEM), suggest that the intervention was moderately to highly effective for 6 out of 7 participants.

Analyses of results demonstrate that the tier program was effective for each participant in terms of changes in homework accuracy, with some participants responding to the less intensive tier, and the rest making improvements in the more intensive tier. Across tiers, 4 out of 7 participants made substantial gains in homework accuracy following one Tier 2 session during treatment phases, and were able to maintain Homework-Accuracy Scores at levels higher than baseline. Three participants met with the researcher one-on-one to receive more intensive tier sessions. These three participants made higher gains in homework accuracy in Tier 3 than in Tier 2 or in baseline. The participant who had questionable gains between baseline and treatment phases displayed much higher and more consistent gains in Tier 3; this pattern indicates that, while Tier 2 was ineffective, Tier 3 was highly effective for her.

Exam improvement. Overall improvements in exams between baseline and treatment phases occurred for the combined participants, as well as for most participants individually. Participants not only made improvements relative to their past performance, but also increased their normative performance. As a group, exam grades improved by an average of 10% between baseline and treatment phases. Furthermore, exam gaps, defined as the differences of individual exam scores from class averages, were smaller during treatment than baseline phases. Score comparisons with those who did not choose to participate show that students who received early intervention performed better on subsequent exams than non-participators.

At the individual level, the intervention had more variable effects on exam performance. While individuals who responded to Tier 2 were able to maintain both homework accuracy and exam improvement trends during the semester, those who required the Tier 3 intervention made little to no improvement on exams following baseline. Even though the Tier 3 intervention helped improve homework accuracy, these participants continued to struggle on the exams. Similar to tier intervention systems at the secondary level, some individuals did not make improvements following intervention, even at the most intensive tier. Because only five exams were given during the semester, changes in exam performance between baseline and intervention phases were not as strong of an indicator of treatment effects as the Daily Homework-Accuracy Scores.

Progress monitoring and intervention. Ehren (2008) and Shinn (2008) describe challenges in selecting progress monitoring tools that are appropriate for higher-level skills and can be quickly and repeatedly measured. This study describes a systematic method to evaluate and target homework accuracy, a skill purported to be related to exam success (Galyon et al., in press). Homework-Accuracy Scores were intended to represent a daily reflection of course understanding and ability to think about items critically. Interventions were designed to enhance homework accuracy as a means to better prepare for exams.

Mellard et al. (2009) recommended that progress monitoring tools be valid and reliable measures of target skills. The current progress monitoring tool was selected based on Galyon et al.'s (in press) findings that the skill of completing homework accurately is a strong predictor of exam success. Similar to findings in Galyon et al.'s study, the grading of homework items in the current study were consistently measured across raters. Furthermore, the homework questions challenged students to display an understanding of material and critically apply course concepts,

a similar task used to complete each multiple-choice exam. For several participants, particularly those who responded to Tier 2, targeting and measuring homework accuracy translated into performance on subsequent exams. These analyses partially support Galyon et al.'s conclusions that homework accuracy is strongly related to exam performance, and that promoting homework-accuracy skills can enhance exam performance. Furthermore, homework accuracy may serve as a reliable and valid progress monitoring tool to determine response to intervention and make data-based decisions on tier placement.

Future research directions. Analyses of the current study suggest many potential benefits of implementing RTI procedures in order to enhance the academic performance of college students who struggle in a course. To my knowledge, this study is the first of its kind in implementing and evaluating the use of an RTI-based model in a college classroom. Insomuch as current research on RTI programs at the college level is limited, there are many avenues for growth in this area. To strengthen the support of using interventions that follow RTI concepts at the college levels, researchers should continue to investigate the effects of RTI interventions with college students. Goals for future research in this area are similar to directions suggested for RTI implementation at the secondary level (see Duffy, 2007): identify appropriate intervention models that can work across subjects, find and assess screening and progress monitoring tools appropriate for different subject areas, consider the unique issues of the settings and participants, and provide structural support for collaboration across college professionals. In order to design RTI programs that are flexible enough to fit multiple college settings, researchers should continue to assess the use of various procedures, universal screeners, and progress monitoring tools in a college course, described in more detail below.

Participant and settings. In a previous study, Abrams and Jernigan (1984) found a strong connection between motivation and college course success among at-risk students. Specifically, they found one of the strongest indicators of first-semester college GPA measured among high-risk students to be their willingness to seek help from their teachers. Similarly, the current study relied on the student's willingness to attend sessions on an out-of-class basis. Conversely, RTI programs at the primary and secondary levels often require participation of the entire student body and are implemented during school hours. Therefore, individuals who need early intervention at this level receive services whether they are motivated to receive them or not. Unlike these RTI interventions, participation in the current study was optional. As expected, only a small percent of students who were asked to participate elected to receive services. Of the 10 who elected to proceed with the intervention, 3 individuals were unable to maintain intervention attendance needed to remain in the study. The low participation rate suggests that motivation was a likely mediating factor driving efforts for improvement in exams. An additional intervening variable between participants and eligible non-participants could be available time for study. Although results were favorable in terms of improvements in homework and exam performance following treatment, it is unclear whether the same effects would occur with eligible individuals who did not participate in the study.

Progress monitoring tools. Research on progress monitoring tools conducted at the secondary level suggests that more complex skills (e.g., comprehension) are harder to evaluate than skills measured at the elementary level (e.g., reading fluency, basic calculation) (Duffy, 2007; Ehren, 2008). In this study, visual analysis did not reveal clear, immediate changes in Homework-Accuracy Scores between phases. The large range of Homework-Accuracy Scores across days in both baseline and treatment phases for the participants allude to the sensitivity of

the current measure to outlier scores. Even though questions were carefully selected and reliably graded, some questions are likely to be more difficult to answer than others. This difference may be due to a number of factors including the particular concept, level of critical thinking required to answer the question, or how the question was worded. In the current study, multiple questions per day were evaluated, which helped limit the power of one particularly challenging or easy question from unduly deflating or inflating the accuracy score. Although the target skill appeared to guide and evaluate the intervention, concerns about its fit in other courses and its efficiency as a tool remain. Researchers should continue to evaluate progress monitoring tools that are shown to reflect target skills in the college classroom and are sensitive to responses to intervention, while avoiding measures that produce inconsistent scores regardless of the student's skill level.

Mellard et al. (2009) suggest that progress monitoring tools should be appropriate indicators of skill level and efficient to remain. In a typical college classroom, finding a tool that can be used repeatedly to measure current skill level and evaluate progress toward educational goals might be difficult. However, many instructors, such as those in the current study, give daily homework assignments. The current study used a grading rubric developed by Galyon et al. (in press) to measure homework accuracy across pertinent questions. Each day, approximately 2-3 questions were evaluated, with a total of nearly 400 questions rated during the study. In large-entry classrooms, using the current grading rubric to evaluate answers to questions may be time and labor intensive. Continued research on progress monitoring tools that are both connected to course exam and efficient should be conducted. For example, future researchers could explore progress monitoring tools that are already evaluated as part of the curriculum or are efficient to obtain (e.g., one question rated per day).

Intervention procedures. In order to increase the likelihood that interventions will be used in practice by instructors, intervention procedures should be perceived as feasible and flexible (Skinner & Skinner, 2008). Another area of this study that should be targeted in future research is the time invested to provide intervention sessions. Although the procedures were designed so that the less-intensive, Tier 2 intervention would be offered in small groups, the majority of Tier 2 sessions had only one or two individuals in attendance. In the later units, only one person required Tier 2. In the earlier units, coordinating times that fit with multiple college students' schedules was a challenge. As a result, I had to schedule more Tier 2 sessions per unit than expected. Although I did not observe performance differences between attendees of group sessions and attendees of one-on-one sessions, it may have affected homework accuracy or exam performance. Future researchers should try to limit the amount of sessions offered by providing a sign-up sheet with instructor availability to help organize scheduling.

Three of the 7 participants, those who required Tier 3 services, made little changes in exams between baseline, Tier 2, and Tier 3 phases. Although the students eventually became more accurate in homework accuracy following the Tier 3 intervention, exam performance did not improve. This pattern is not largely inconsistent with Galyon et al.'s results (in press), who found that while homework accuracy was the largest predictor of exam success, other variables (e.g., critical thinking, participation) also accounted for some of the variability in exam scores. Unclear in this study is whether students did not respond because the overall difficultly level of the course, because some students required a more-intensive intervention than offered, because the intervention's target did not cater to the student's specific needs, or due to a combination of these factors. Compared to those who responded well to Tier 2, these participants may have benefitted more from a tier-based intervention following a problem-solving model rather than the

standard protocol format used in this study. In problem-solving models, interventions are developed based on the specific needs and responses of the individual (Fuchs & Fuchs, 2006; 2007). Unlike the standard protocol model, which uses preselected, evidence-based instructional strategies to develop new skills, the problem-solving model is individually designed to enhance the student's current skills (Fuchs & Fuchs 2007). Fuchs and Fuchs (2007) recommend that RTI programs integrate both models, with the standard treatment model implemented for academic problems and the problem-solving model used for behavioral problems. In the college setting, a problem-solving model may target specific behavior obstacles (e.g., poor time management, poor note-taking skills, and test anxiety) that are inferring with a student's academic performance. Future research directions may include comparing the effects and feasibility of problem-solving models versus standard protocol models at the college level. Researchers should continue to evaluate possible out-of-class interventions that can raise performance at the college level.

The intervention in the current study was designed to be course specific. For example, procedures prepared students for the upcoming unit exam by using course resources. In order to determine if effects of a tier-based intervention can be generalized across classrooms, it might be beneficial for interventions to be designed to target broader test-taking, time management, and preparation skills. For example, Tier 2 services at the college level can be broadened to provide general test-taking strategies, and then the more intensive tiers can focus on specific performance variables in the target course. Similar to targeting reading fluency skills at the primary level, effects of interventions that give broader instruction might generalize better to the participant's future courses.

Universal screener. The universal screener used in the current study, first exam performance, was supported by Study 1 as an adequate determination of who may need help in

the course. Using the first exam as a universal screener has its pros and cons. The universal screener served two purposes in that it was a tool for identification and assessed understanding of the course's first unit. The screener was completed within a 50-min time limit and identification of eligible students was quickly and easily determined. Lastly, because the screener was part of student's grade, students were likely to perform the best they could.

Catts et al. (2009) discuss the challenge of finding the optimal point to provide initial universal screeners, or the point at which groups being screened have enough experience and ability to produce adequate variance on the screening scale. Screeners that provide a lot of false positives (e.g., students who do not need the intervention), will take up resources and time. Although it appeared evident that the current screener adequately identified who needed help, unknown is the best time to include the universal screener at the college level. Waiting until after the first exam, for example, may negatively impact students struggling in the course by lowering overall course grades and increasing feelings of discouragement. While some expressed motivation to improve subsequent exam scores after doing poorly on the first exam, others might perceive initial failure as a fluke occurrence or, conversely, that improvement is unobtainable due to external controls (e.g., perception of unfair testing or teaching). By identifying whom to target and offering services prior to the first exam, students might be comfortable using effective in-class and out-of-class strategies to prepare for exams. Universal screeners that are unrelated to course grades, are efficiently measured, and assess accurate skill levels as early as possible should continue to be explored.

Data-based decision making. While RTI practices typically recommend measuring rates of learning curve to evaluate progress, the tier placement decisions in this study were based on both exam and homework performance criteria (e.g., an average of 10% increase in homework

accuracy, a C or higher on the exam). The criteria were set prior to the course with the hope that they accurately reflected who no longer needed intervention, who needed to continue similar services, and who may need more intensive intervention. The criteria were aimed to be challenging but attainable goals for participants. However, the criteria set may have not been sensitive enough to determine intervention need with certainty. For example, some individuals were able to make substantial improvement on homework accuracy on the first day of a unit, but their homework performance dropped toward the end of the unit. For individuals who had outlier scores, a median unit score may have been a more appropriate homework criterion measure than a mean. Nevertheless, there was a distinct pattern in exam performance of those who needed Tier 3 interventions versus those who did not. Researchers should continue to assess the best methods to determine progress of intervention and make data-based decisions on treatment placement in colleges.

Implications. Although RTI programs have been developed at the primary and secondary levels, I found no studies tailoring an RTI program to meet the needs of college students. Analyses of results of Study 2 establish that interventions featuring RTI concepts are possible to implement in the college setting if universal screeners are in place, appropriate tools to measure progress are repeatedly used, time and resources are available to help monitor progress and implement interventions, and college students are willing and able to receive out-of-class services. The program was implemented systematically, with treatment integrity and inter-rater checks to support the reliability and validity of Study 2's results. The current program was well-liked by participants, who reported that they used preparation skills taught for the upcoming exam and plan to use these strategies in future classes.

Determining the appropriate level of out-of-class intervention may be a challenge to implement logistically. Both teachers and students may have limited time and resources to provide or receive out-of-class exam help. By adopting RTI strategies at the college level, professors would be able to provide interventions to individuals who need it and at intensity levels specific to the needs of the student. Preventive and early intervention measures can detect and address academic problems without the undue stress and resources associated with remedial efforts.

The current intervention procedures and the level of intensity were appropriate to each participant in terms of improving homework accuracy; participants who did not increase and/or sustain homework accuracy performance following the less intensive tier were able to make these gains by a more content-focused, intensive tier. This pattern of results provides evidence that, similar to primary and secondary levels, a tier-based intervention program that systematically identifies individuals who are struggling, provides progress monitoring tools to assess response, and makes evidence-based decisions on intervention needs can be successfully implemented at the post-secondary level. Future research supporting successful implementation of tier-based intervention programs at the college level have the potential to increase its likelihood to be further accepted and followed university-wide.

Chapter V

General Discussion

Overall Summary of Findings

With more students than ever entering college, it is not surprising that academic failure remains a common concern at the higher-education level. The current studies aimed to expand upon literature by identifying characteristics of those at-risk of course failure and developing an intervention program that is systematic and feasible to implement with college students. Whereas Study 1 was an investigative study comparing exam groups on past performance, Study 2 was a study supporting the efficacy of an intervention program for initially low-exam performers. The combination of studies follows Fuchs and Fuchs's (2006) guidelines that every student receives research-based teaching starting at the general tier. The first study compared the subsequent success in the course of individuals who did well on the first exam and those who did poorly on the first exam. This comparison was conducted to identify patterns of overall academic outcome between groups and to highlight possible contributors to poor exam performance. Conclusions of the first study were that those who do poorly on the first exam in a course continue to do poorly on subsequent exams, practice exams, and quizzes than those who initially do well in the course. Furthermore, individuals who do poorly on the first exam enter the course with lower critical thinking skills and reported GPAs on average.

Analyses of results in Study 1 support using early identification measures to provide interventions to college students in a similar course, which was the objective of Study 2. Patterns of differences between the groups partly set the groundwork for Study 2 in terms of target variables, intervention procedures, and universal screeners. For example, the comparison groups in Study 1 were based on initial exam performance, which served as Study 2's universal

screener. Study 2's interventions focused on exam improvement via targeting preparation strategies, which complements Study 1's findings that subsequent exam success and scores on exam-preparation variables were significantly lower for those in the Low-Exam Group. In combination, the studies describe the execution of a tier-based intervention that focused on improving academic skills with which students in the past have struggled. By analyzing data of past performance in a course, empirically-sound interventions and target variables can be designed to fit the specific needs of the course. In creating an RTI program, Study 1 supports the use of performance on the first exam as a universal screener, given that analyses of results showed significantly lower performance on many academic factors for those who did poorly versus well on this exam. Study 2 supports the use of a tier-based intervention to help those who initially do poorly on the first exam.

Implications and Future Directions

Although the studies spanned 15 course sections, Studies 1 and 2 were conducted in the confines of one educational psychology course. Therefore, though the combined studies were directed toward identifying and improving factors related to exam improvement, results might not generalize to courses organized quite differently from the target course. Not all instructors keep and have access to a database that contains demographic and academic variables spanning the previous 2 years at the detail available with this course. Thus, identification of performance gaps between past High- and Low-Exam Groups might not be as easily accessible. It is likely that most instructors, however, still keep grade books and a listing of final grades, which could help inform academic achievement and progress of those who do poorly on the first exam. Future researchers should continue to compare academic and demographic variables of initially

low performers. This information could help support current findings and guide possible tier intervention programs to address the academic needs of struggling students.

Settings for future studies. With increased research and application of tier-based interventions at the college setting, similar programs as described in Study 2 may eventually be suitable and beneficial across several types of college courses. Inasmuch as the scope of the current research is limited, researchers might aim to identify what types of courses are currently more conducive than others to a tier-based intervention adhering to RTI core concepts. Courses that have access to previous performance data across several academic factors can narrow focus on what skills need to be targeted and how to conduct universal screeners. Furthermore, courses that have resources available, such as graduate teaching assistants, collaborative professional help, recording devices, and room availability, might be good prospects for tier-intervention programs. The current study was labor intensive and could not have been conducted without the assistance of inter-raters, access to room availability, and cooperation of instructors and teaching assistants. Just as RTI programs at the primary and secondary level rely on resources and assistance from an RTI team to collect skill data, make empirically-based decisions, and implement interventions, programs at the college level will likely also run much more smoothly with these aids in place.

Large-entry courses, such as the one described in this study, are recommended settings for future studies aimed to provide early interventions to students struggling from the outset of a course. First of all, struggling students may be more likely to be overlooked by their instructor in large-entry courses than in smaller, more advanced courses. Dillon, Kokkelenberg, and Christy (2002) found an inverse relationship with lower grade point averages and increased class size at the college level. Potential reasons for this inverse relationship include that teachers

simply cannot create a close working relationship with each student in a large course, students are more likely to skip class in large lecture courses, and exams might be quickly assessed and recorded by instructors without much attention to past performance. For example, Chapman and Ludlow (2010) found overall lower student perceptions of the importance of attendance to student learning in larger than in smaller college class rooms.

Secondly, by assessing interventions in largely populated classes, researchers might be able to have a larger pool of participants. Only 6% of the 162 students taking the course were eligible and participated in the current study. Unless participation in the study became a mandatory part of the course, it is likely the number of research participants will stay low, even in larger classes. Thirdly, compared to 300 or 400 level courses, 100 and 200 level courses often offer more grades within the course. More grades in a course allow opportunities for past data to be used to inform patterns of academic problems, students to be identified early enough to improve, and progress monitoring tools to be administered as a part of the course. Lastly, instructors teaching several sections of the same large-entry course might be more likely to follow a similar curriculum across sections and semesters than teachers who occasionally teach a course or have a more advanced, less-populated course. If course sections are similarly taught, then teachers might be better able to predict what skill to target in the current class by looking at past performance.

Classes catering to freshmen and sophomores in college, such as core requirement courses for majors, are also recommended settings for future studies. The current tier intervention program was designed as a measure of early intervention to thwart academic problems in a course in which the majority of students enrolled were sophomores and juniors. By implementing RTI-like programs in courses for freshmen and sophomores, students might be

able to alter and establish exam preparation and work habit skills that can carry over to other courses. Interventions provided to students in their first 2 years may decrease dropout rate by addressing academic vulnerabilities before they become problematic and contribute to failure. Not only can general test-preparation and test-taking skills be taught, the student might also be able to more clearly and completely understand core concepts in the course. By building a stronger academic base in general, entry-course concepts, students are likely to better apply more advanced concepts taught in higher-level courses within their major.

Previously, researchers have found that student choice of college major is largely influenced by time and experience within an introductory course (Mauldin, Crain, & Mounce, 2000; Keilor, Bush, & Bush, 1995), as well as the student's perceived assessment of her or his strengths and abilities within that field (Beggs, Banthan, & Taylor, 2008). To a much smaller extent, students also choose majors based on the ease of the major and ability to maintain a high grade point average (Adams, Pryor, & Adams, 1994). Early intervention aims to provide the best practices needed to meet the student's needs. If interventions are provided in large-entry courses typically recommended for first- or second-year students, the student receiving intervention might be better able to discern if the current major is the best fit for her or him. Furthermore, if the student receiving the most intensive tier of help at the introductory level continues to make little progress in the course, then he or she might decide to switch to a major that better complements her or his pattern of skills.

Conclusions

The current studies are small steps toward preventing college dropout and improving academic outcome of vulnerable students, with the hope that future research will build upon the current findings. To increase graduation rates, researchers should continue to find ways to

identify and help struggling students. Even by providing course-specific intervention help to those who initially struggle, students might be able to acquire study and preparation tips that will transfer to other courses.

The current studies focus on the performance of one identified group of students in one course. Aside from information collected on reported GPA, little information is known of the previous and subsequent academic performance across other courses. Beyond the scope of this project is exploring identified students' performance in other courses and cumulative GPA at graduation. In order to determine collegial outcome for students who struggle initially and receive services in one course, academic records would need to be available and analyzed longitudinally.

One area to further identify factors leading to academic failure is to look at the records of those who drop out of college system-wide. For example, researchers could see if patterns of low grades were consistent across years of attendance for dropout students, if grades deteriorated during college attendance, or if variables unrelated to course grades led to dropout. If grades are consistently low across years of attendance, then administrators might choose to implement intervention programs such as Study 2 for individuals who struggle in entry-level courses (e.g., freshmen and sophomores) as opposed to more advanced courses. If tier-based interventions were established in college classrooms, total outcome in terms of attendance, cumulative GPA, and completion of major requirements would inform the interventions' possible effects.

The future of research in integrating RTI programs in the college setting has many avenues. By identifying possible patterns and producing early interventions at the college level, teachers can provide services to help build skills associated with exam preparation and higher-order thinking for struggling students in their courses. College-wide implementation of a tier-

based intervention program can potentially produce an increase in overall cumulative GPA and graduation rates. One of the more important factors in its future applications is whether instructors are willing and equipped to implement such intervention programs in their classrooms.

List of References

Abrams, H.G., & Jernigan, L.P. (1984). Academic support services and the success of high-risk college students. *American Educational Research Journal, 21,* 261-274.

Adams, S.H., Pryor, L.J., & Adams, S.L. (2004). Attraction and retention of high-aptitude students in accounting: An exploratory longitudinal study. *Issues in Accounting Education, 9*, 45-58.

Barger-Anderson, R., Domaracki, J. W., Kearney-Vakulick, N., & Kubina, R. M. (2004). Multiple baseline designs: The use of a single-case experimental design in literacy research. *Reading Improvement, 41,* 217- 226.

Beggs, J.M., Banthan, J.H., & Taylor, S. (2008). Distinguishing the factors influencing college students' choice of major. *College Student Journal, 42,* 381-394.

Bloom, B. (1956). *Taxonomy of Educational Objectives, Handbook I. The Cognitive Domain.* New York: David McKay.

Brozo, W. G. (2010). Response to intervention of responsive instruction? Challenges and possibilities of response to intervention for adolescent literacy. *Journal of Adolescent & Adult Literacy, 53*, 277-281.

Burns, M.K. (2008). Response to intervention at the secondary level. *Principal Leadership, 8,* 12- 15.

Burns, M.K., Appleton, J.J., & Stehouwer, J.D. (2005). Meta-analytic review of responsiveness to-intervention research: Examining field-based and research-implemented models. *Journal of Psychoeducational Assessment, 23*, 381-394.

Carroll, E., Williams, R. L., & Hautau, B. (2006). Cooperative learning contingencies: Unrelated versus related individual and group contingencies. *Journal of Behavioral Education, 15,* 191-202.

Catts, H. W., Petscher, Y., Schatschneider, C., Bridges, M. S., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of Learning Disabilities, 42,* 163-176.

Chapman, L., & Ludlow, L. (2010). Can downsizing college class sizes augment student outcomes? An investigation of the effects of class size on student learning. *Journal of General Education*, *59*, 105-123.

Dillon, M., Kokkelenberg, E., & Christy, S. (2002). The effects of class size on student achievement in higher education: Applying an earnings function [Report]. New York: Cornell University. (ERIC Document Reproduction Service No. ED482389.)

Duffy, H. (2007). *Meeting the needs of significantly struggling learners in high school: A look at approaches to tiered intervention.* National High School Center, American Institutes of Research.

Ehren, B. J. (2008) Response to intervention in secondary schools: Is it on your radar screen? *RTI Action Network*. Retrieved June 15, 2011 from http://www.rtinetwork.org/Learn/Why/ar/RadarScreen.

Foorman, B.F., Francis, D.J., Fletcher, J.M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90,* 37-55.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41,* 93-99.

Fuchs, L.S., & Fuchs, D. (2007). A model for implementing responsiveness to intervention. *Teaching Exceptional Children, 39,* 14-30.

Fuchs, L.S., Fuchs, D., & Compton, D. L. (2010). Rethinking response to intervention at middle and high school. *School Psychology Review, 39*, 22-28.

Foster, L.N., Krohn, K..R., McCleary, D.F., Aspiranti, K.B., Nalls, M.L., Quillivan, C.C., Taylor, C.M., & Williams, R.L., (2009). Increasing low-responding students' participation in class discussion. *Journal of Behavioral Education, 18,* 173-188.

Galyon, C.E., Blondin, C.A., Forbes, B.E., & Williams, R.L. *(*in press*)*. Does homework matter? Exam performance in large college classes. *Journal of Excellence in College Teaching*.

Hautau, B., Turner, H. C., Carroll, E., Jaspers, K., Krohn, K., Parker, M., & Williams, R. L. (2006). Differential daily writing conditions and performance on major multiple-choice exams. *Journal of Behavioral Education, 15*, 170-180.

Holtzman, M. (2008). Demystifying application-based multiple-choice questions. *College Teaching, 56*, 114-120.

Jimerson, S.R., Burns, M.K., & VanDerHeyden, A.M. (Eds.) (2007). Response to Intervention at school: The science and practice of assessment and intervention. In S.R. Jimerson, M.K. Burns, & A. M. VanDerHeyden (Eds.). *Handbook of Response to Intervention: The Science and Practice of Assessment and Intervention* (pp. 3-9). New York: Springer.

Keilor, B.D., Bush, R.P., & Bush, A.J. (1995). Marketing-based strategies for recruiting business students in the next century. *Marketing Education Review, 5,* 69-79.

Krohn, K.R., Foster, L.N., McClearly, D.F., Taylor, C.M, B., Nalls, M.L., Quillivan, C.C., & Williams, R.L. (2010). Effects of self-recording and contingent credit on balancing participation across students. *Journal of Behavioral Education, 19,* 134-155.

Kuh, G. D., Cruce, T.M., Shoup, R., & Kinzie, J. (2008). Unmasking the effects of student engagement on first-year college grades and persistence. *The Journal of Higher Education, 79,* 540-563.

Ma, H. H. (2006). An alternative method for quantitative synthesis of single-subject researches: Percentage of data points exceeding the median. *Behavior Modification, 30*, 598-617.

Mauldin, S., Crain, J., & Mounce, P. (2000). The accounting principles instructor's influence on students' decision to major in accounting. *Journal of Education for Business, 75,* 142-148.

Mellard, D. F., McKnight, M., & Woods, K. (2009). Response to intervention screening and progress-monitoring practices in 41 schools. *Learning Disabilities Research & Practice, 24,* 186-195.

McCleary, D. F., Aspiranti, K. B., Foster, L.N., Blondin, C. A., Galyon, C. E., Yaw, J., Forbes, B.E., & Williams, R. L. (2011). Using randomized and delayed credit for student comments to balance participation across students and time in large college courses. *The Journal of General Education, 60,* 194-214.

Oliver, R., & Williams, R. L. (2005). Direct and indirect effects of completion versus accuracy contingencies on practice-exams and actual-exam performance. *Journal of Behavioral Education, 14*, 141-152.

Parker, R.I., Vannest, K.J., & Davis, J.L. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification, 35,* 303-322.

Pithers, R.T., & Sodden, R. (2000). Critical thinking in education: A review. *Educational Research, 42,* 237-249.

Ransdell, S. (2001). Predicting college success: the importance of ability and non-cognitive variables. *International Journal of Educational Research, 35,* 357-364.

Reutebuch, C. (2008). Succeed with a response-to-intervention model. *Intervention in School and Clinic, 44*, 126-128.

Scruggs, T. E., Mastropieri, M. A., Cook, S. B., & Escobar, C. (1986). Early interventions for children with conduct disorders: A quantitative synthesis of single-subject research. *Behavioral Disorders*, *11*, 260-271.

Shinn, M.R. (2008). Implementation in Secondary Schools. In S.L. Fernley, *Answer Book on RTI* (pp. 8:1-8:17). Horsham, PA: LRP Publications.

Skinner, C.H., Pappas, D.N., & Davis, K.A. (2005). Enhancing academic engagement: Providing opportunities for responding and influencing students to choose to respond. *Psychology in the Schools, 42,* 389-403.

Skinner, C. H., & Skinner, A. L. (2008). Establishing an evidence base for a classroom management procedure with a series of studies: Evaluating the Color Wheel. *Journal of Evidence-Based Practices for Schools, 8,* 88-101.

Slavin, R.E., Karweit, N.L., & Wasik, B. (1992). *Preventing early school failure: What works*? Baltimore, MD: Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students.

Stepp, S. L., Schrock, S. A., & Coscarelli, W. C. (1996). Can a multiple-choice exam discriminate between masters and nonmasters in instructional design? *Performance Improvement Quarterly, 9*, 23-35.

Ticha, R., Espin, C. A., & Wayman, M. M. (2009). Reading progress monitoring for secondary-school students: Reliability, validity, and sensitivity to growth of reading-aloud and maze-selection measures. *Learning Disabilities Research & Practice*, 24, 132-142.

Turner, H. C., Bliss, S., Hautau, B., Carroll, E., Jaspers, K. E., & Williams, R. L. (2006). Brief daily writing activities and performance on major multiple-choice exams. *The Journal General Education, 55,* 221-246.

Turner, H., & Williams, R. L. (2007). Generic vocabulary development and performance on multiple-choice exams. *Journal of College Reading and Learning, 37,* 64-81.

U.S. Department of Education, National Center for Education Statistics (2011). *The Condition of Education 2011* (NCESS 2011-033).

U.S. Department of Education, National Center for Education Statistics (2009). *The Condition of Education 2009.*

Wallace, M., & Williams, R. L. (2003). Multiple-choice exams: Explanations for student choices. *Teaching of Psychology, 30,* 136-139.

Watson, G. B., & Glaser, E. M. (1994). *Watson-Glaser critical thinking appraisal Form S manual.* San Antonio, TX: Harcourt Brace.

Williams, R. L., & Eggert, A. (2002). Notetaking predictors of test performance. *Teaching of Psychology*, *29,* 234-237.

Williams, R.L., Oliver, R., Allin, J.L., Winnn, B., & Booher, C.S. (2003). Psychological critical thinking as a course predictor and outcome variable, *Teaching of Psychology, 30,* 220-223.

Williams, R. L., Oliver, R., & Stockdale, S. (2004). Psychological versus generic critical thinking as predictors and outcome measures in a large undergraduate human development course. *The Journal of General Education, 53,* 37-58.

Williams, R. L., Skinner, C. H., & Jaspers, K. E. (2007). Extending research on the validity of brief reading comprehension rate and level measures to college course success. *The Behavior Analyst Today*, *8,* 163-174.

Williams, R.L., & Stockdale, S.L. (2003). High-performing students with low critical thinking scores. *The Journal of General Education, 52,* 200-226.

Williams, R. L., & Worth, S. L. (2002). Thinking skills and work habits: Contributors to course performance. *The Journal of General Education, 51,* 200-227.

Zajacova, A., Lynch, S.M., & Espenshade, T.J., (2005). Self-efficacy, stress, and academic success in college. *Research in Higher Education, 46,* 667-704.

Appendices

Appendix A

Tables

Table 1

Study 1 Multivariate Analysis of Differences in Means across Pre- and Within-Course Variables of High- and Low-Exam Groups from Fall '09-Spring '11

Variable	Group	*N*	*M*	*SD*	Possible range	*F*	*df*	Sig.
Critical Thinking	Low	132	24.11	4.72	0-40	67	1	.000
	High	148	29.35	5.86				
GPA	Low	132	2.99	0.43	0-4	79	1	.000
	High	148	3.46	0.45				
Exam A	Low	132	30.94	4.04	0-50	1478	1	.000
	High	148	44.94	1.71				
Exam B	Low	132	32.23	5.80	0-50	239	1	.000
	High	148	42.16	4.94				
Exam C	Low	132	38.02	4.77	0-50	118	1	.000
	High	148	43.86	4.24				
Exam D	Low	132	36.73	5.36	0-50	157	1	.000
	High	148	43.70	3.91				
Exam E	Low	132	36.33	6.53	0-50	536	1	.000
	High	148	43.75	3.93				
Average Exam	Low	132	34.85	3.68	0-50	537	1	.000
	High	148	43.68	2.67				
Average Post-B Exam	Low	132	35.83	4.33	0-50	282	1	.000
	High	148	43.37	3.15				
Quiz	Low	132	3.70	1.02	0-5	68	1	.000
	High	148	4.52	0.61				
Practice Exam	Low	132	4.05	0.84	0-5	62	1	.000
	High	148	4.67	0.44				

Table 2

Study 1 Multivariate Analysis of Differences in Means across Pre- and Within-Course Variables of High- and Low-Exam Groups from Spring '10-Spring '11

Variable	Group	*N*	*M*	*SD*	Range	*F*	*df*	Sig.
Critical Thinking	Low	110	24.45	4.74	0-40	52	1	.000
	High	101	29.71	6.05				
GPA	Low	110	2.98	0.43	0-4	65	1	.000
	High	101	3.47	0.45				
Exam A	Low	110	30.81	4.22	0-50	985	1	.000
	High	101	45.09	1.83				
Exam B	Low	110	32.33	5.87	0-50	163	1	.000
	High	101	42.13	5.23				
Exam C	Low	110	38.15	4.86	0-50	101	1	.000
	High	101	43.93	3.25				
Exam D	Low	110	36.74	5.41	0-50	115	1	.000
	High	101	43.94	4.22				
Exam E	Low	110	36.07	6.74	0-50	97	1	.000
	High	101	43.68	4.02				
Average Exam	Low	110	34.82	3.70	0-50	386	1	.000
	High	101	43.75	2.81				
Average Post-B Exam	Low	110	35.82	4.33	0-50	201	1	.000
	High	101	43.42	3.30				
Quiz	Low	110	3.57	1.03	0-5	54	1	.000
	High	101	4.44	0.63				
Practice Exam	Low	110	4.03	0.87	0-5	42	1	.000
	High	101	4.65	0.47				
Reported Work Hours	Low	110	8.80	9.93	0-40	1.26	1	.262
	High	101	7.25	10.24				
Course Hours	Low	110	15.24	1.67	0-20	<0	1	.824
	High	101	15.37	2.26				

Table 3

Study 1 Multivariate Analysis of Differences in Means across Pre- and Within-Course Variables of High- and Low-Exam Groups for Fall '09-Spring '10 and Fall '10-Spring '11

		Fall '09-Spring'10			Fall '10- Spring '11		
Variable	Group	*N*	*M (SD)*	*F*	*N*	*M (SD)*	*F*
Critical Thinking	Low	44	24.02 (4.71)	22*	88	24.16 (4.75)	47*
	High	93	28.87 (5.98)		55	30.16 (5.60)	
GPA	Low	44	3.03 (0.40)	27*	88	2.97 (0.44)	46*
	High	93	3.44 (0.44)		55	3.49 (0.45)	
Exam A	Low	44	31.16 (5.45)	507*	88	30.83 (3.15)	900*
	High	93	45.12 (1.72)		55	44.63 (1.66)	
Exam B	Low	44	32.57 (5.99)	121*	88	32.07 (5.33)	90*
	High	93	42.67 (4.99)		55	41.31 (5.55)	
Exam C	Low	44	36.95 (4.62)	55*	88	38.56 (4.76)	71*
	High	93	43.56 (4.81)		55	44.62 (2.95)	
Exam D	Low	44	36.86 (5.07)	70*	88	36.61 (5.53)	73*
	High	93	43.81 (3.91)		55	43.93 (3.94)	
Exam E	Low	44	37.86 (6.49)	43*	88	35.56 (6.45)	71*
	High	93	43.81 (4.01)		55	43.93 (3.93)	
Average Exam	Low	44	35.10 (3.54)	257*	88	34.73 (3.76)	230*
	High	93	43.71 (2.61)		55	43.63 (2.80)	
Quiz	Low	44	4.32 (0.71)	10*	88	3.39 (1.01)	33*
	High	93	4.67 (0.54)		55	4.26 (0.66)	
Practice Exam	Low	44	4.01 (0.92)	34*	88	4.07 (0.80)	25*
	High	93	4.68 (0.42)		55	4.66 (0.47)	
Attendance	Low	44	16.18 (3.98)	20*	88	67.67 (9.22)	< 1
	High	93	18.46 (8.74)		55	68.29 (11.62)	
Homework Completion	Low	44	14.41 (1.87)	22*	88	90.22 (14.89)	1
	High	93	17.97 (1.90)		55	93.25 (13.80)	

*significant at the $p < .05$ level

Table 4

Study 1 Multivariate Analysis of Differences in Means across Pre- and Within-Course Variables of High- and Low-Exam Groups for Fall '09 (1), Spring '10 (2), Fall '10 (3), and Spring '11 (4)

		1		*2*		*3*		*4*	
		Low *N* = 29 High *N* = 49		Low *N* = 23 High *N* = 48		Low *N* = *60* High *N* = 28		Low *N* = 40 High *N* = 31	
Variable		*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*
Critical Thinking	Low	23.17	23*	25.09	7*	24.97	21*	23.00	33*
	High	28.80		29.15		30.36		29.68	
Exam A	Low	31.28	731*	30.87	186*	30.27	405*	31.08	461*
	High	44.71		45.60		44.43		44.77	
Exam B	Low	30.38	106*	33.43	54*	31.83	33*	31.85	64*
	High	42.33		43.19		40.25		42.13	
Exam C	Low	36.55	34*	36.61	44*	38.73	39*	37.33	46*
	High	43.86		43.19		44.68		44.45	
Exam D	Low	35.14	56*	37.13	34*	37.15	33*	35.58	56*
	High	43.24		44.10		43.93		44.26	
Exam E	Low	36.66	42*	38.17	17*	35.62	36*	34.38	39*
	High	44.16		43.81		42.79		44.03	
Avg. Exam	Low	34.00	169*	35.24	128*	34.72	120*	34.04	128*
	High	43.66		43.98		43.21		43.93	
Quiz	Low	4.20	10*	4.32	5*	3.22	9*	3.36	24*
	High	4.68		4.66		3.95		4.40	
Practice Exam	Low	4.04	25*	3.88	17*	3.95	21*	4.12	8*
	High	4.70		4.66		4.71		4.57	
Attend. Rate	Low	16.59	5*	16.43	9*	65.83	1	68.78	< 1
	High	17.98		18.69		68.32		67.84	
H.W Comp.	Low	16.00	8*	13.57	10*	87.80	6*	89.85	< 1
	High	17.88		17.73		95.68		90.97	
Part.	Low	12.76	18*	26.93	13*	26.75	2	19.35	4*
	High	13.52		27.57		29.86		25.58	

*significant at the $p < .05$ level

Table 5

Study 2 Participants, Year in School, Critical Thinking Percentile, and Attendance Rate

Student	Year	GPA	Critical Thinking (%)	Attendance rate
Section 1 Group				
Edith	Sophomore	3.03	25%	92%
Kylie*	Junior	2.70	5%	92%
Kathryn*	Sophomore	3.20	20%	100%
Kristin	Junior	3.30	1%	96%
Sally	Sophomore	NR	5%	100%
Section 2 Group				
Allison*	Sophomore	2.74	3%	96%
Emma	Sophomore	3.81	3%	100%
Elizabeth	Non-Degree	3.30	1%	96%
Section 3 Group				
Becky	Junior	NR	1%	96%
Stephanie	Sophomore	3.40	1%	100%

*Denotes individuals who dropped out of intervention program prematurely.
NR- Not Reported

Table 6

Tier 2 Intervention Acceptability Survey Response Averages

	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
1. I felt better prepared after meeting one-on-one or in small groups.	1	2	3	4 **4.8**	5	6
2. This intervention helped me understand why I missed exam items.	1	2	3	4	**5**	6
3. Discussing missed items helped me answer questions on the next exam.	1	2	3	4 **4.5**	5	6
4. I used exam taking strategies discussed in intervention on the next exam.	1	2	3	4	5	**5.8** 6
5. Discussing preparation strategies helped me study for the next exam.	1	2	3	4	5	**5.2** 6
6. I would recommend this intervention for those that are struggling on exams.	1	2	3	4 **4.8**	5	6
7. I will use these study and test-taking tips in my other classes.	1	2	3	4	5	**5.3** 6

Note: Shaded, inputted numbers denote the mean scores per score across participants' answers

Table 7

Tier 3 Intervention Acceptability Survey Responses

	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
1. I felt better prepared with this intervention than the last one.	1	2	3	**4**	5	6
2. This intervention helped me understand key concepts in the unit.	1	2	3	4	**5**	6
3. Discussing homework questions helped me prepare for the exam.	1	2	3	4	5	**6**
4. I changed how I completed my homework following these sessions.	1	2	3	4	5	**6**
5. Discussing preparation strategies helped me study for the next exam.	1	2	3	4	**5**	6
6. I would recommend this intervention for those that are struggling on exams.	1	2	3	4	**5**	6
7. I will use these preparation tips in completing homework in my other classes.	1	2	3	4	5	**6**

Note: Shaded numbers denote the participant's answers

Table 8

Homework Accuracy Mean, Range, and Percent of Data Points Exceeding the Median (PEM) Scores for Individuals and Groups across Baseline and Treatment Phases

Individual	Baseline	Treatment/Post-treatment	PEM
Edith	15 (29)	31 (79)	.73
Kylie*	10 (20)	21 (64)	.33
Kathryn*	10 (17)	31 (67)	.86
Kristin	15 (44)	36 (71)	1.00
Sally	7 (17)	35 (74)	.93
Section 1 Group	*12 (30)*	*34 (75)*	*.93*
Allison*	26 (64)	32 (54)	.75
Emma	27 (58)	42 (34)	1.00
Elizabeth	26 (60)	46 (36)	1.00
Section 2 Group	*27 (59)*	*44 (35)*	*1.00*
Becky	11 (34)	14 (29)	.38
Stephanie	19 (34)	32 (59)	.75
Section 3 Group	*15 (34)*	*23 (44)*	*.63*
Groups Combined	18 (41)	34 (51)	.85

*Denotes individuals who dropped out of intervention program prematurely. Their means, ranges, and PEM scores are not included in the group total or combined rows.

Table 9

Mean and Range of Percent of Homework Accuracy across Individuals and Groups for Baseline, Tier 1, Tier 2, and Maintenance Phases

Individual	Baseline	Tier 2	Tier 3	Maintenance
Edith	15 (29)	18 (56)	51 (56)	35 (18)
Kylie*	10 (20)	21 (64)	NA	NA
Kathryn*	10 (17)	31 (67)	NA	NA
Kristin	15 (44)	36 (29)	NA	44 (54)
Sally	7 (17)	32 (72)	44 (24)	NA
Section 1 Group	*12 (30)*	*29 (43)*	*48 (40)*	*40 (36)*
Allison*	26 (64)	32 (54)	NA	NA
Emma	27 (58)	36 (6)	NA	44 (34)
Elizabeth	26 (60)	37 (7)	NA	50 (39)
Section 2 Group	*27 (59)*	*37 (7)*	*NA*	*47 (37)*
Becky	11 (34)	7 (25)	21 (16)	NA
Stephanie	19 (34)	39 (53)	NA	26 (26)
Section 3 Group	*15 (34)*	*23 (39)*	*21 (16)*	*26 (26)*
Groups Combined	18 (41)	30 (30)	35 (28)	38(33)

*Denotes individuals who dropped out of intervention program prematurely. Their means and standard deviations are not included in the group total or combined rows.

Table 10

Mean Percent of Homework Accuracy and Range (in parenthesis) across Individuals and Groups for Units A-E

Individual	Unit A	Unit B	Unit C	Unit D	Unit E
Edith	15 (29)	**36 (40)**	**3 (6)**	**51 (56)**	**29 (18)**
Kylie*	13 (20)	**16 (64)**	**DROP**		
Kathryn*	14 (6)	**31 (74)**	**23 (35)**	**DROP**	
Kristin	15 (44)	**36 (29)**	**34 (41)**	**55 (46)**	**42 (46)**
Sally	9 (17)	**26 (44)**	**30 (74)**	**44 (34)**	**32 (12)**
Section 1 Group	*13 (30)*	***33 (38)***	***22 (40)***	***50 (45)***	***34 (25)***
Allison*	21 (15)	32 (64)	**32 (54)**	**DROP**	
Emma .	26 (23)	29 (58)	**36 (6)**	**49 (30)**	**40 (34)**
Elizabeth	22 (40)	29 (60)	**43 (7)**	**61 (10)**	**40 (21)**
Section 2 Group	*24 (32)*	*29 (59)*	***40 (7)***	***55 (20)***	***40 (28)***
Becky	8 (20)	6 (17)	19 (34)	**7 (25)**	**27 (16)**
Stephanie	15 (23)	25 (34)	18 (30)	**39 (53)**	**26 (36)**
Section 3 Group	*12 (22)*	*16 (56)*	*19 (32)*	***23 (55)***	***27 (26)***
Groups Combined	16 (28)	26 (51)	27 (26)	43 (40)	34 (26)

*Denotes individuals who dropped out of intervention program prematurely. Their means are not included in the group total or combined rows.
Bold denotes start of treatment phases.

Table 11

Mean Percent and Range of Homework Accuracy (in Parenthesis) across Days in Unit A

Individual	Day 1	Day 2	Day 3	Day 4
Edith	10 (20)	29 (NA)	0 (0)	20 (NA)
Kylie*	0 (0)	20 (11)	12 (33)	20 (NA)
Kathryn*	17 (33)	13 (25)	11 (24)	NA (NA)
Kristin	44 (47)	14 (NA)	0 (0)	0 (NA)
Sally	17 (33)	15 (29)	4 (14)	0 (NA)
Section 1 Group	*24 (33)*	*19 (29)*	*1 (5)*	*7 (NA)*
Allison*	17 (33)	15 (29)	30 (42)	20 (NA)
Emma	17 (33)	40 (21)	26 (38)	20 (NA)
Elizabeth	0 (0)	27 (4)	22 (57)	40 (NA)
Section 2 Group	*9 (17)*	*34 (13)*	*24 (48)*	*30 (NA)*
Becky	0 (0)	7 (14)	4 (17)	20 (NA)
Stephanie	17 (33)	0 (0)	23 (57)	20 (NA)
Section 3 Group	*9 (17)*	*4 (2)*	*14 (37)*	*20 (NA)*
Groups Combined	14 (22)	19 (15)	13 (30)	19 (NA)

*Denotes individuals who dropped out of intervention program prematurely. Their means are not included in the group total or combined rows.
NA: Not Applicable

Table 12

Mean Percent and Range of Homework Accuracy (in Parenthesis) across Days in Unit B

Individual	Day 1	Day 2	Day 3	Day 4
Edith	**17 (33)**	**44 (47)**	**56 (89)**	**16 (33)**
Kylie*	**0 (0)**	**54 (27)**	**0 (0)**	**0 (0)**
Kathryn*	**0 (0)**	**74 (13)**	**31 (39)**	**18 (29)**
Kristin	**34 (67)**	**54 (27)**	**25 (50)**	**29 (8)**
Sally	**0 (0)**	**44 (47)**	**25 (5)**	**34 (18)**
Section 1 Group	***17 (33)***	***47 (40)***	***35 (48)***	***37 (20)***
Allison*	0 (0)	64 (7)	38 (75)	27 (67)
Emma	34 (67)	64 (7)	6 (11)	11 (33)
Elizabeth	0 (0)	60 (80)	18 (14)	37 (21)
Section 2 Group	*17 (34)*	*62 (44)*	*12 (13)*	*24 (27)*
Becky	0 (0)	17 (33)	0 (0)	NA (NA)
Stephanie	34 (67)	0 (0)	31 (39)	34 (18)
Section 3 Group	*17 (34)*	*9 (17)*	*16 (20)*	*34 (18)*
Groups Combined	17 (34)	51 (36)	21 (81)	37 (21)

*Denotes individuals who dropped out of intervention program prematurely. Their means are not included in the group total or combined rows.
Bold: Individuals in treatment
NA: Not Applicable

Table 13

Mean Percent and Range of Homework Accuracy (in Parenthesis) across Days in Unit C

Individual	Day 1	Day 2	Day 3	Day 4
Edith	**0 (0)**	**0 (0)**	**4 (13)**	**6 (17)**
Kathryn*	**20 (40)**	**7 (20)**	**42 (25)**	**22 (33)**
Kristin	**17 (33)**	**36 (27)**	**58 (25)**	**25 (50)**
Sally	**74 (13)**	**0 (0)**	**21 (38)**	**22 (33)**
Section 1 Group	***30 (18)***	***12 (9)***	***28 (25)***	***18 (33)***
Allison*	**54 (27)**	**0 (0)**	**46 (75)**	**29 (50)**
Emma	**37 (7)**	**36 (27)**	**33 (75)**	**39 (17)**
Elizabeth	**40 (80)**	**35 (7)**	**33 (25)**	**39 (17)**
Section 2 Group	***26 (44)***	***36 (17)***	***33 (50)***	***39 (17)***
Becky	34 (67)	0 (0)	29 (11)	11 (33)
Stephanie	37 (7)	7 (20)	13 (25)	15 (33)
Section 3 Group	*36 (37*	*4 (10)*	*21 (18)*	*13 (33)*
Groups Combined	31 (33)	17 (12)	27 (31)	23 (28)

*Denotes individuals who dropped out of intervention program prematurely. Their means are not included in the group total or combined rows.
Bold: Individuals in treatment or maintenance phases

Table 14

Mean Percent and Range of Homework Accuracy (in Parenthesis) across Days in Unit D

Individual	Day 1	Day 2	Day 3	Day 4
Edith	23 (71)	54 (58)	79 (43)	46 (42)
Kristin	71 (57)	71 (50)	52 (46)	25 (50)
Sally	39 (67)	58 (75)	34 (18)	46 (42)
Section 1 Group	*44 (65)*	*61 (61)*	*64 (21)*	*39 (53)*
Emma	56 (80)	50 (75)	29 (57)	59 (17)
Elizabeth	64 (75)	67 (83)	59 (32)	54 (42)
Section 2 Group	*60 (78)*	*59 (79)*	*44 (45)*	*57 (30)*
Becky	4 (14)	25 (75)	0 (0)	0 (0)
Stephanie	39 (57)	38 (100)	66 (18)	13 (25)
Section 3 Group	*14 (36)*	*32 (88)*	*33 (9)*	*7 (13)*
Groups Combined	39 (60)	51 (76)	47 (25)	34 (32)

Table 15

Mean Percent and Range of Homework Accuracy (in Parenthesis) across Days in Unit E

Individual	Day 1	Day 2	Day 3	Day 4
Edith	36 (67)	36 (47)	43 (31)	25 (50)
Kristin	44 (40)	22 (67)	34 (23)	68 (15)
Sally	29 (67)	42 (47)	23 (9)	35 (30)
Section 1 Group	*36 (70)*	*33 (54)*	*33 (21)*	*43 (22)*
Emma	63 (30)	29 (67)	34 (23)	33 (15)
Elizabeth	49 (47)	36 (47)	28 (43)	45 (10)
Section 2 Group	*56 (39)*	*33 (57)*	*31 (33)*	*39 (13)*
Becky	29 (23)	18 (33)	13 (40)	23 (5)
Stephanie	32 (43)	31 (40)	7 (20)	33 (15)
Section 3 Group	*31 (33)*	*25 (37)*	*10 (30)*	*28 (10)*
Groups Combined	41 (47)	30 (49)	25 (28)	37 (15)

Table 16

Average Exam Percentages (out of 100) and Deviations from Class Mean Percentage (Dev.) across Baseline and Treatment/Post-Treatment Phases for Individuals and Groups

Individual	Baseline	Deviation	Treatment/Post-treatment	Deviation
Edith	66	-12	67	-10
Kylie*	62	-16	82	+5
Kathryn*	64	-14	67	-11
Kristin	66	-12	84	+7
Sally	62	-16	61	-17
Section 1 Group	*64*	*-14*	*72*	*-6*
Allison*	59	-11	74	0
Emma	67	-3	75	+1
Elizabeth	62	- 8	70	-3
Section 2 Group	*63*	*-7*	*74*	*0*
Becky	73	-8	69	-10
Stephanie	67	-13	87	+8
Section 3 Group	*69*	*-11*	*79*	*0*
Combined	65	-11	75	-2

*Denotes individuals who dropped out of intervention program prematurely.

Table 17

Units A-E Exam Percentages (out of 100) and Exam Percentage Deviations from Class Means (Dev.) across Individuals and Groups

Individual	A	Dev.	B	Dev.	C	Dev.	D	Dev.	E	Dev.
Edith	66	-12	**64**	**-10**	**70**	**-12**	**70**	**-6**	**64**	**-12**
Kylie*	62	-16	**72**	**-2**	**86**	**+4**	**86**	**+10**	**84**	**+8**
Kathryn*	64	-14	**66**	**-8**	**70**	**-12**	**66**	**-12**	**64**	**-12**
Kristin	66	-12	**88**	**+14**	**78**	**-4**	**88**	**+12**	**82**	**+6**
Sally	62	-16	**60**	**-14**	**76**	**-6**	**52**	**-24**	**54**	**-22**
Section 1 Group	*64*	*-14*	***70***	***-4***	***76***	***-6***	***72***	***-4***	***70***	***-6***
Allison*	52	-22	66	0	**68**	**-6**	**74**	**0**	**80**	**+6**
Emma	68	-6	66	0	**78**	**+4**	**76**	**+1**	**72**	**-2**
Elizabeth	60	-14	64	-2	**80**	**+6**	**68**	**-6**	**62**	**-10**
Section 2 Group	*60*	*-14*	*65*	*-1*	***75***	***+1***	***74***	***0***	***71***	***- 3***
Becky	66	-14	64	-10	88	0	**70**	**-8**	**68**	**-12**
Stephanie	62	-18	66	-8	74	-14	**88**	**+10**	**86**	**+6**
Section 3 Group	*64*	*-16*	*65*	*-9*	*81*	*-7*	***79***	***+1***	***77***	***-3***

*Denotes individuals who dropped out of intervention program prematurely.
Bold = Denotes post treatment implementation

Table 18

Average Exam Score (out of 50), Standard Deviation (in parenthesis), and Difference between Study 2 Participants (Part.) and Eligible Students who did participate in Study 2 (Non-Part.), and Study 2 Participants (Part.) and Low-Exam Group of Study 1 (L.E.)

Group	Part. (N = 10)	Non-Part. (N = 32)	Diff.	L.E. (N = 165)	Diff.
Unit A	31.40 (2.27)	31.59 (2.75)	-0.19	30.72 (4.02)	0.68
Unit B	33.80 (3.39)	29.13 (6.04)	4.67	31.61 (6.00)	2.19
Unit C	38.40 (3.34)	37.43 (4.80)	0.97	37.37 (5.69)	1.03
Unit D	36.90 (5.67)	33.73 (4.10)	3.17	36.34 (5.67)	0.56
Unit E	35.80 (5.45)	34.70 (5.41)	1.10	35.99 (6.78)	-0.19
Average Test Score	35.26 (3.01)	33.12 (3.12)	2.14	34.23 (4.09)	1.03
Average Test Score (Units B-E)	36.23 (3.67)	33.57 (3.69)	2.66	35.08 (4.85)	1.15
Total Test Score	176.30 (15.07)	159.06 (33.63)	17.24	166.97 (29.34)	9.33
Total Test Score (Units B-E)	144.90 (14.68)	131.58 (23.87)	13.32	136.24 (28.41)	8.66

Table 19

Units A-E Exam Percentages (out of 100) across Study 2 Participants and Exam Percentage Differences (Diff.) between Participants and Non-Participants (N = 32)

Individual	A	Diff.	B	Diff.	C	Diff.	D	Diff.	E	Diff
Edith	66	3	**64**	**6**	**70**	**-5**	**70**	**3**	**64**	**-5**
Kylie*	62	-1	**72**	**14**	**86**	**11**	**86**	**19**	**84**	**15**
Kathryn*	64	1	**66**	**8**	**70**	**-5**	**66**	**-1**	**64**	**-5**
Kristin	66	3	**88**	**30**	**78**	**3**	**88**	**21**	**82**	**13**
Sally	62	-1	**60**	**2**	**76**	**1**	**52**	**-15**	**54**	**-15**
Section 1 Group	*64*	*1*	***70***	***12***	***76***	***1***	***72***	***5***	***70***	***1***
Allison*	52	-11	66	8	**68**	**-7**	**74**	**7**	**80**	**11**
Emma	68	5	66	8	**78**	**3**	**76**	**9**	**72**	**3**
Elizabeth	60	-3	64	6	**80**	**5**	**68**	**1**	**62**	**-7**
Section 2 Group	*60*	*-3*	*65*	*7*	***75***	***0***	***73***	***6***	***71***	***2***
Becky	66	3	64	6	88	13	**70**	**3**	**68**	**-1**
Stephanie	62	-1	66	8	74	-1	**88**	**21**	**86**	**17**
Section 3 Group	*64*	*1*	*65*	*7*	*81*	*6*	***79***	***12***	***77***	***8***

*Denotes individuals who dropped out of intervention program prematurely.
Bold = Denotes post treatment implementation

Appendix B

Figures

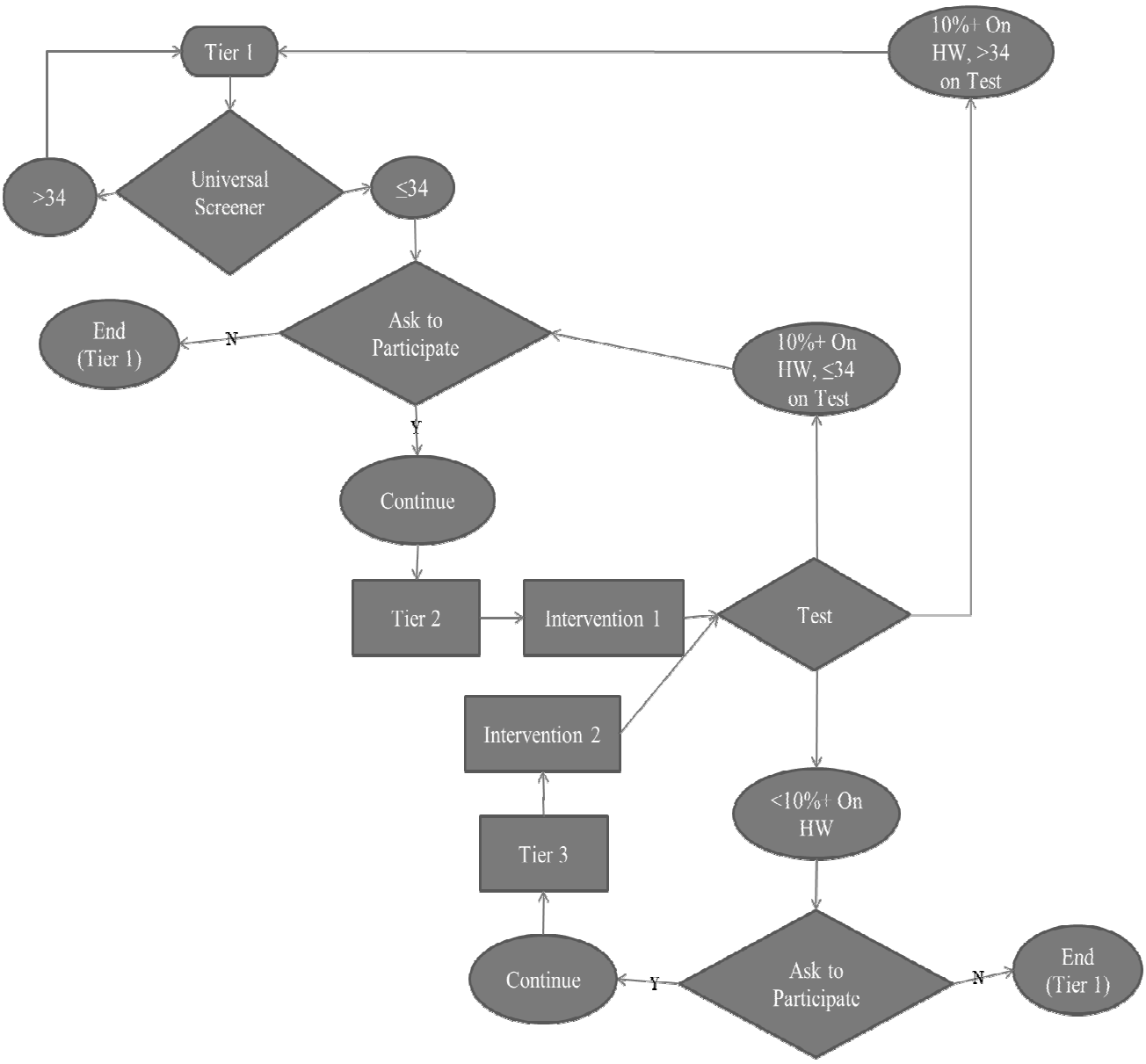


Figure 1. Flow of intervention procedures for Study 2.

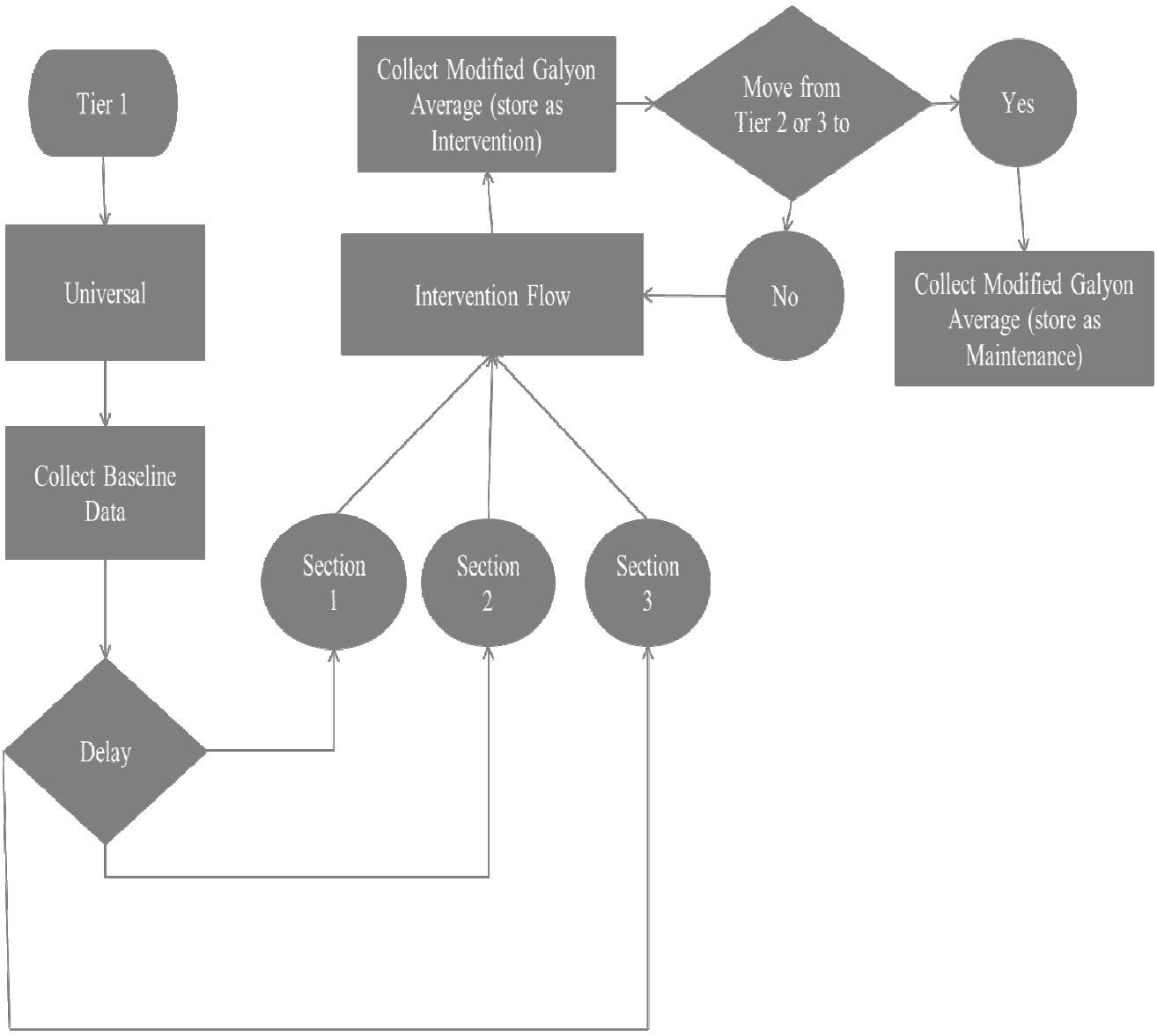


Figure 2. Flow of research design procedures for Study 2.

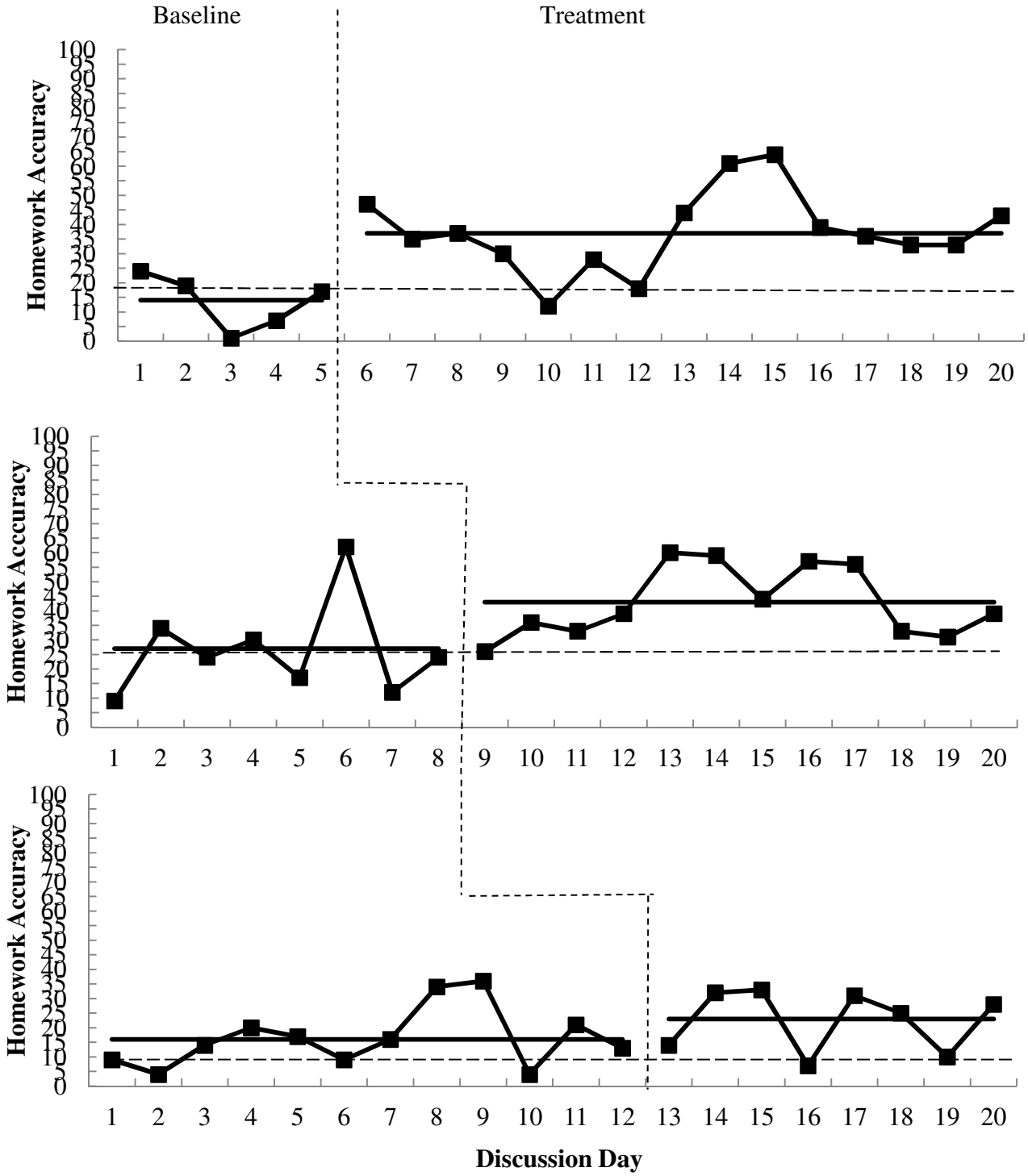


Figure 3. Daily Homework-Accuracy Scores in Baseline and Treatment Phases between Section 1, Section 2, and Section 3. Phase Means are Denoted by a Solid Line and the Baseline Median Level is Denoted by a Broken Line.

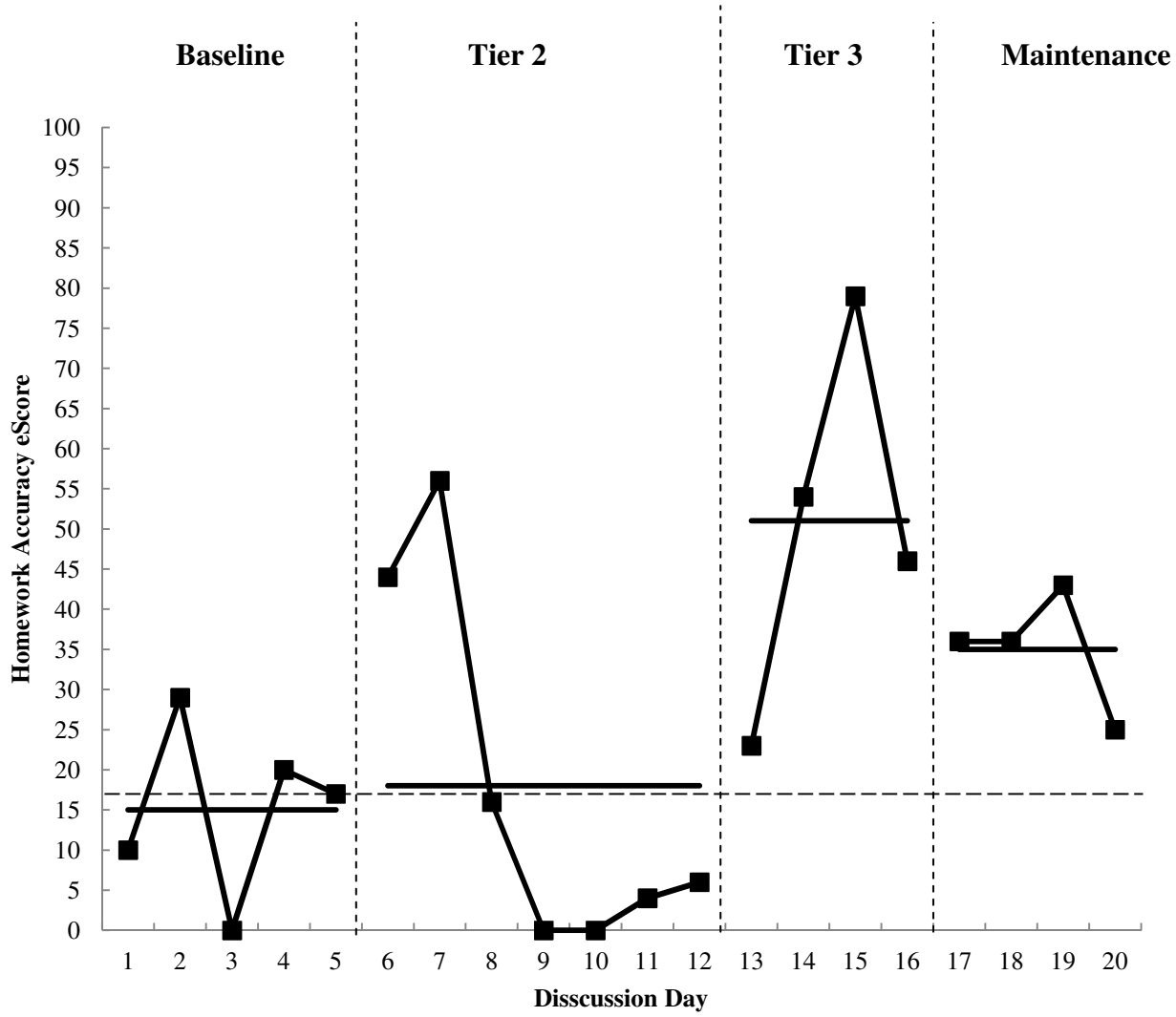


Figure 4. Edith's (Section 1) Daily Homework-Accuracy Scores across Baseline, Tier 2, Tier 3, and Maintenance Phases. Phase Means are Denoted by a Solid Line and the Baseline Median Level is Denoted by a Broken Line.

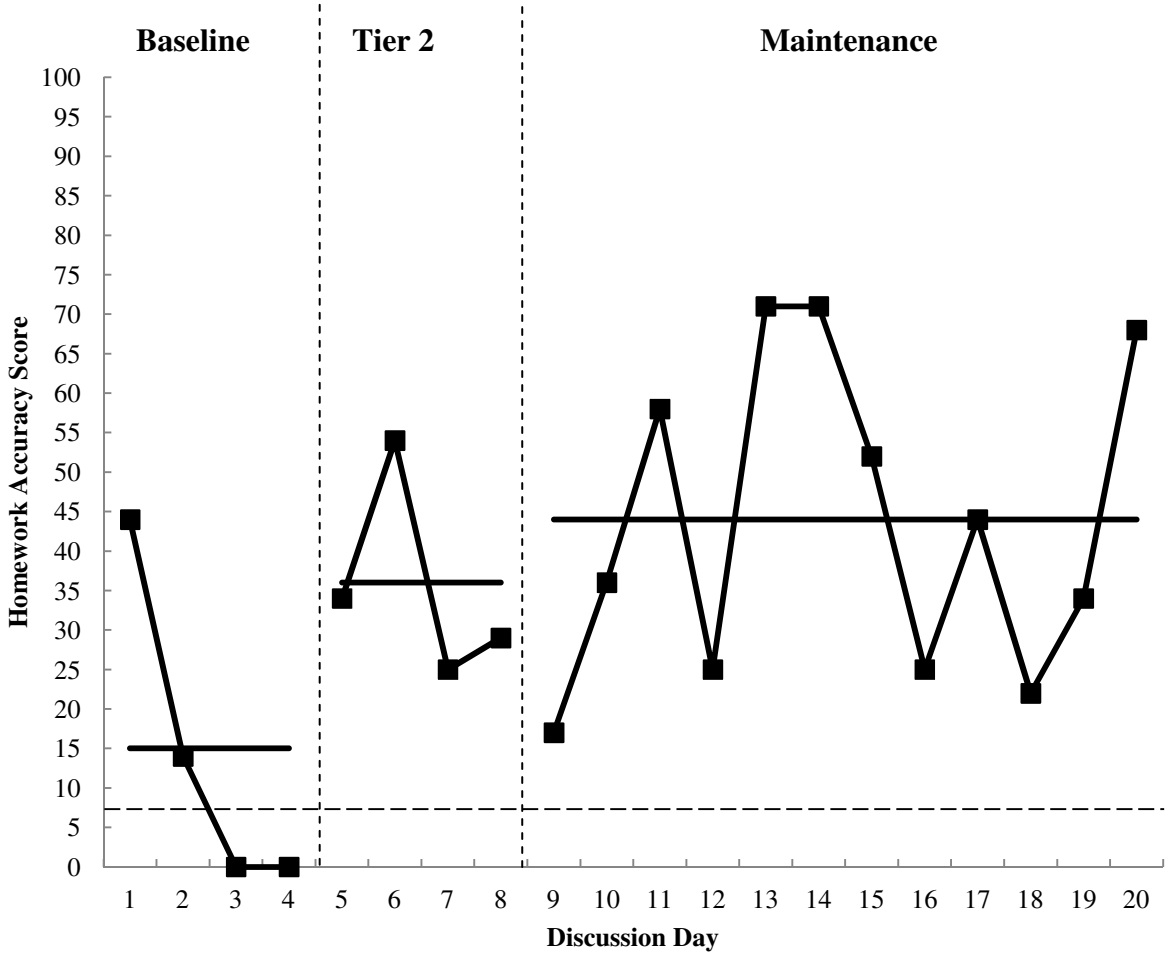


Figure 5. Kristin's (Section 1) Daily Homework-Accuracy Scores across Baseline, Tier 2, and Maintenance Phases. Phase Means are Denoted by a Solid Line and the Baseline Median Level is Denoted by a Broken Line.

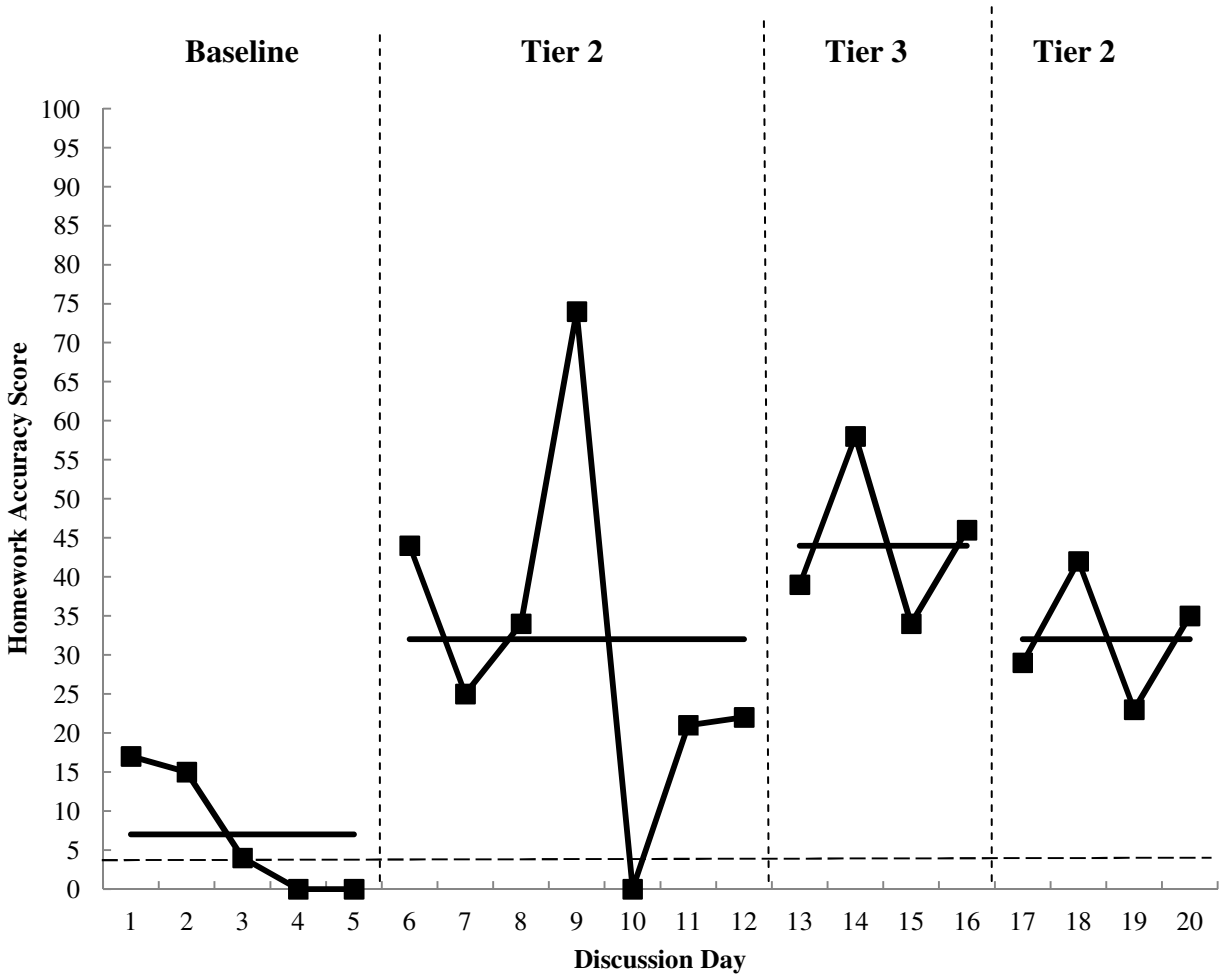


Figure 6. Sally's (Section 1) Daily Homework-Accuracy Scores across Baseline, Tier 2, and Tier 3 Phases. Phase Means are Denoted by a Solid Line and the Baseline Median Level is Denoted by a Broken Line.

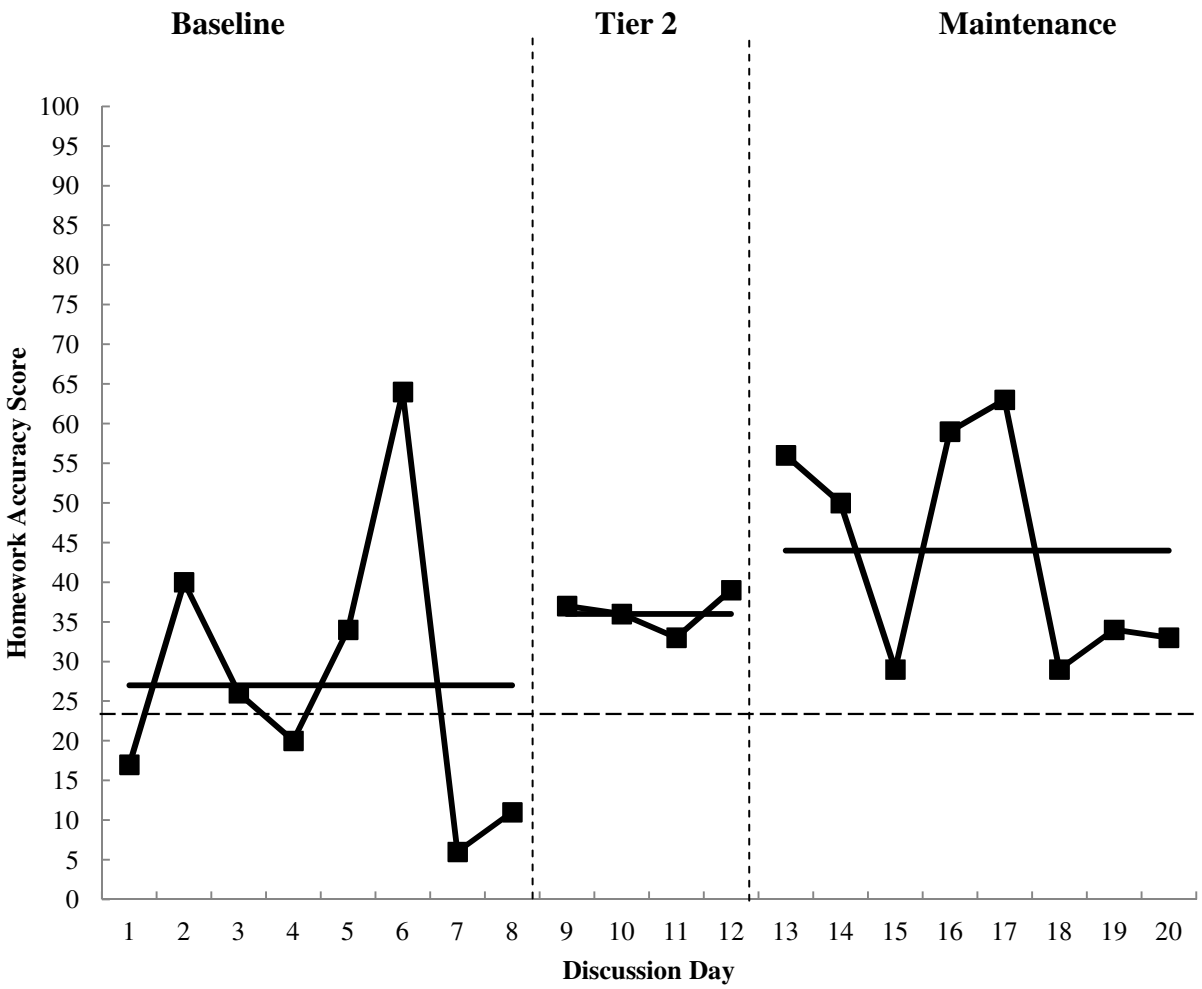


Figure 7. Emma's (Section 2) Daily Homework-Accuracy Scores across Baseline, Tier 2, and Maintenance Phases. Phase Means are Denoted by a Solid Line and the Baseline Median Level is Denoted by a Broken Line.

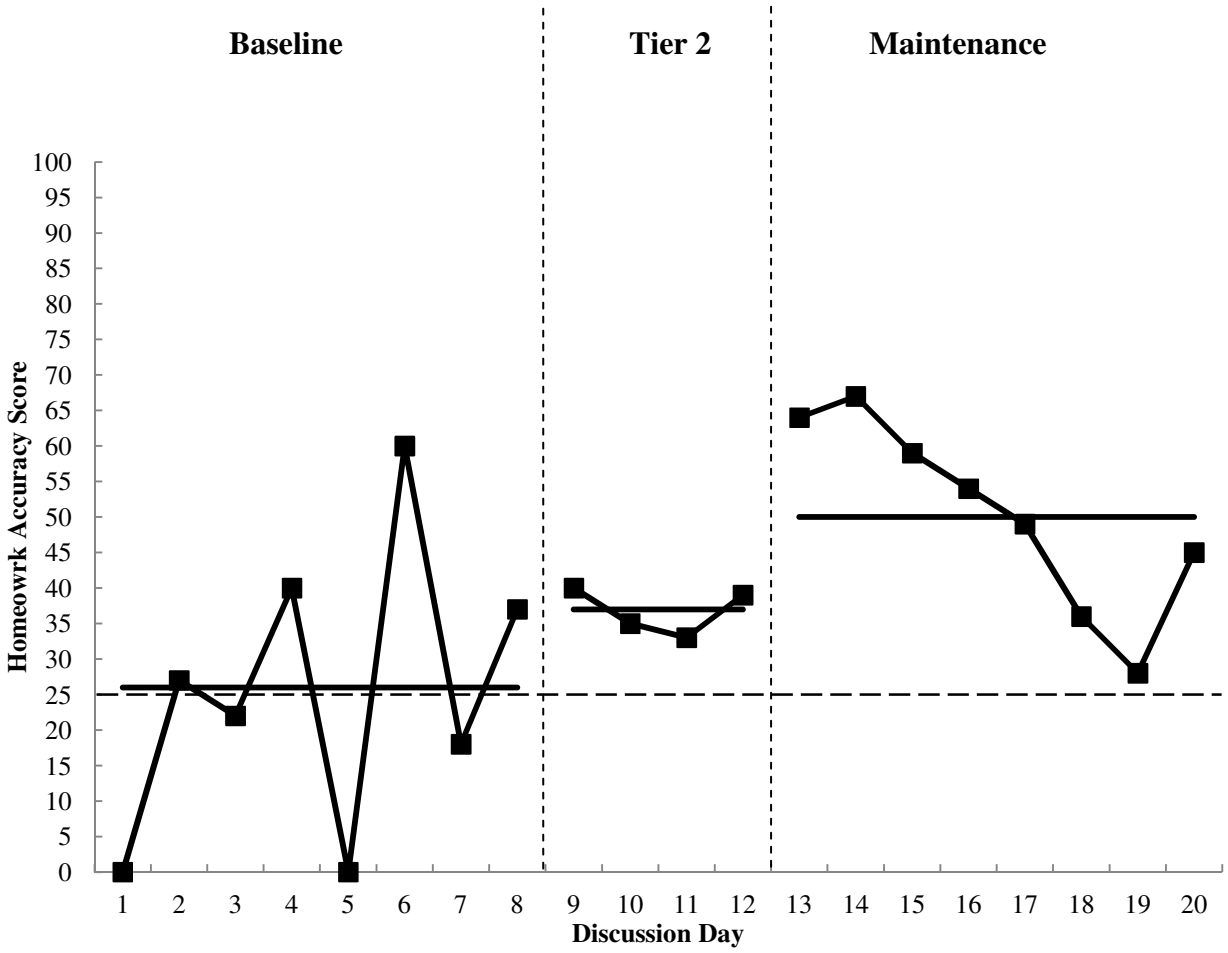


Figure 8. Elizabeth's (Section 2) Daily Homework-Accuracy Scores across Baseline, Tier 2, and Maintenance Phases. Phase Means are Denoted by a Solid Line and the Baseline Median Level is Denoted by a Broken Line.

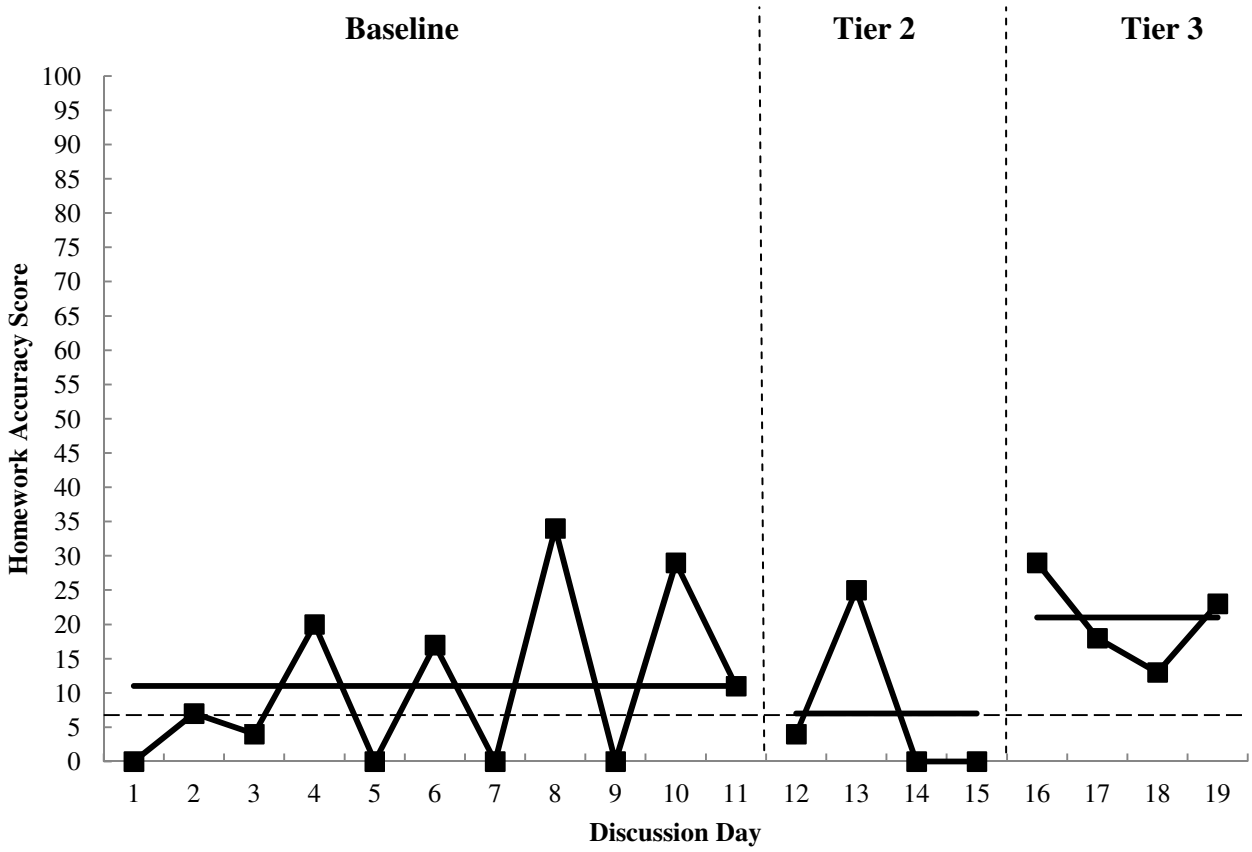


Figure 9. Becky's (Section 3) Daily Homework-Accuracy Scores across Baseline, Tier 2, and Tier 3 Phases. Phase Means are Denoted by a Solid Line and the Baseline Median Level is Denoted by a Broken Line.

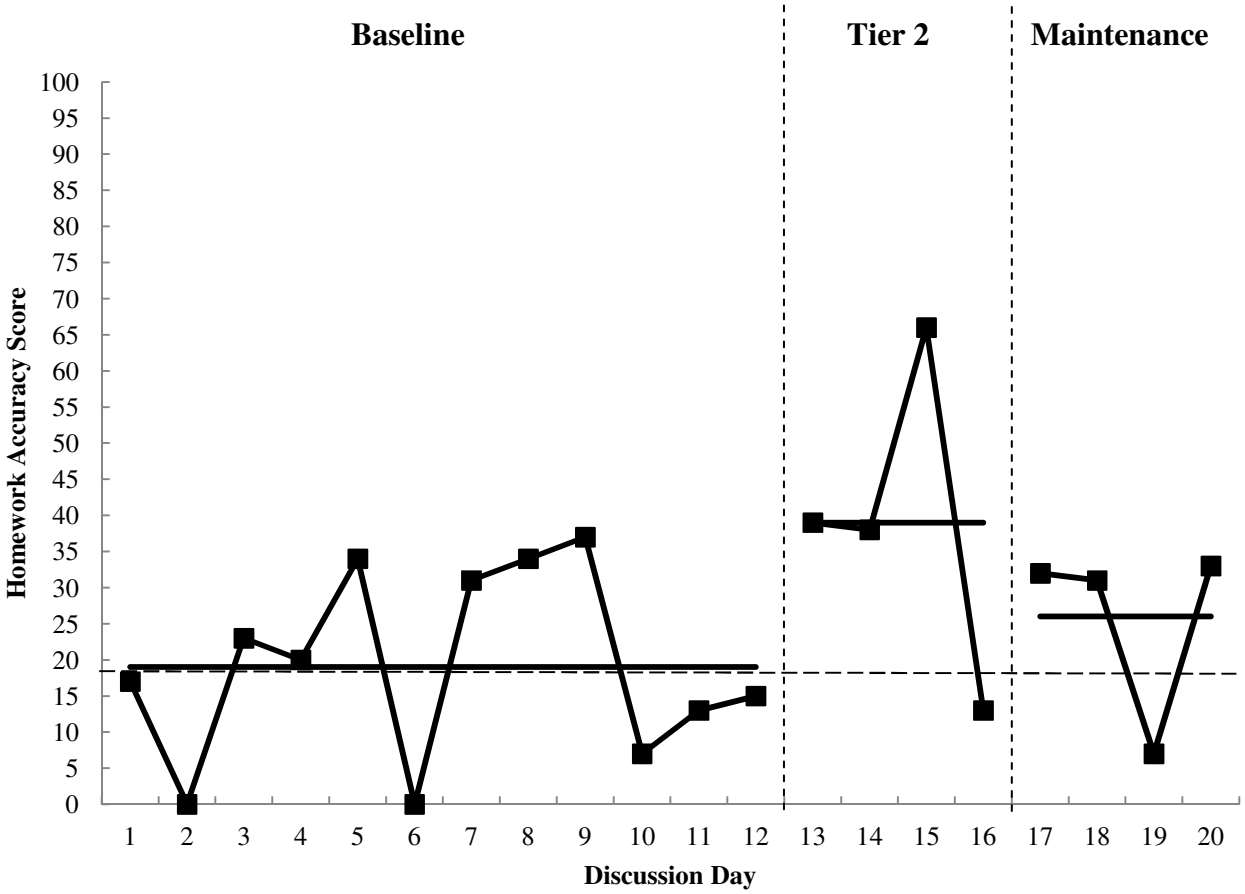


Figure 10. Stephanie's (Section 3) Daily Homework-Accuracy Scores across Baseline, Tier 2, and Maintenance Phases. Phase Means are Denoted by a Solid Line and the Baseline Median Level is Denoted by a Broken Line.

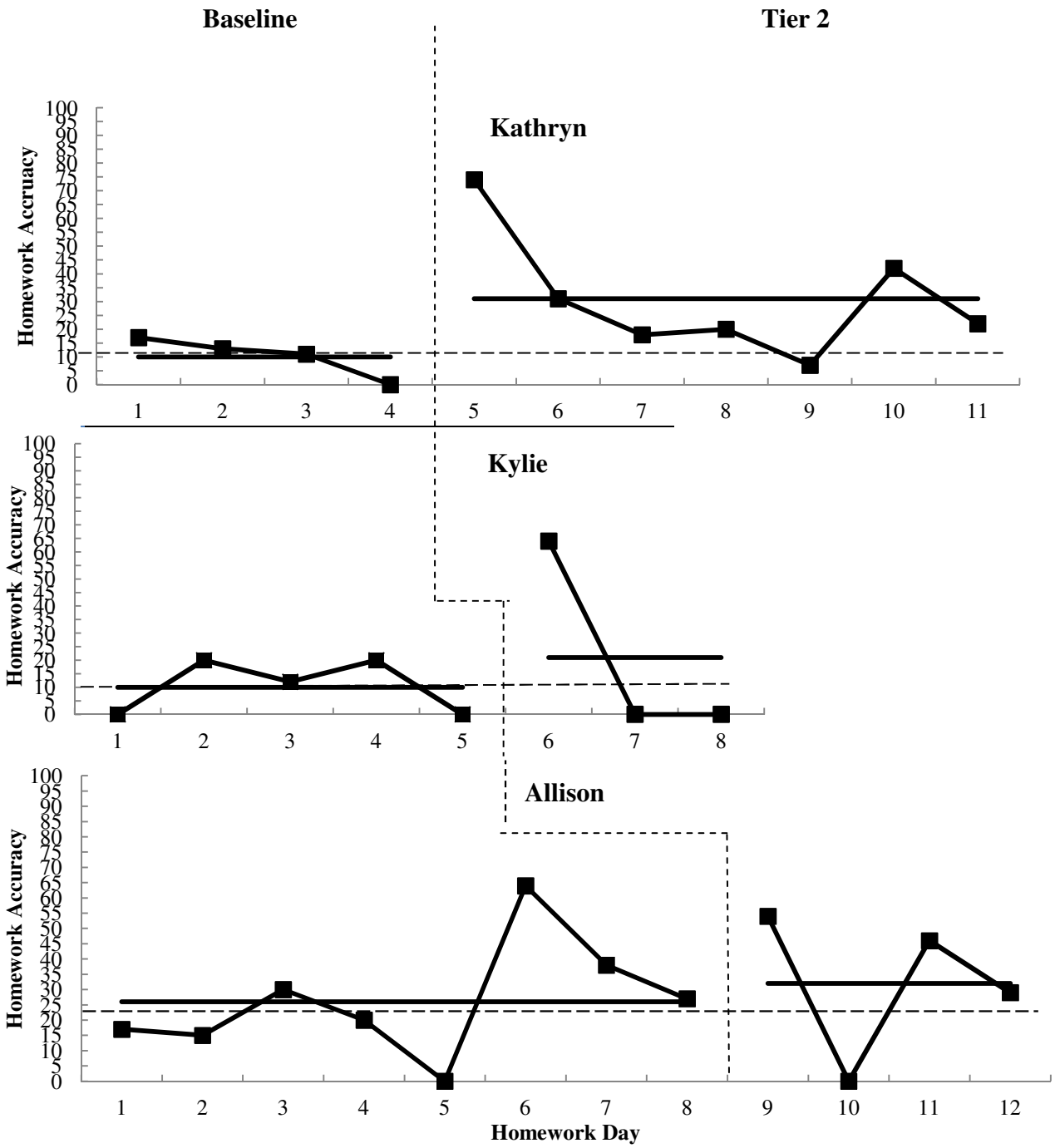


Figure 11. Daily Homework-Accuracy Scores across Baseline and Tier 2 Phases for Kathryn, Kylie, and Allison, Participants who Discontinued Intervention Services. Phase Means are Denoted by a Solid Line and the Baseline Median Level is Denoted by a Broken Line.

Appendix C

Scripts of Invitation to Participate in Tier 2 and Tier 3 Interventions

1. Invitation to begin Tier 2 Intervention

Section 1, sent September 9th by lead GTA

> "Based on your Exam A performance, I would like to invite you to participate in one 30-minute small group study session this upcoming unit. The purpose of the session is to review test-taking strategies and help you develop a clearer idea of how information in each unit is linked to exam questions. This session should be really beneficial, as it aims to better acquaint you with the test material and help you more effectively study for future exams. Please let me know by e-mail no later than Sunday evening on September 11th whether you are interested. The group sessions will be scheduled and conducted by Carolyn Blondin, who has 3 years of experience teaching this course. "

Section 2, sent September 28th by lead GTA

> "Based on your past exam performance, I would like to invite you to participate in one 30-minute small group study session this upcoming unit. The purpose of the session is to review test-taking strategies and help you develop a clearer idea of how information in each unit is linked to exam questions. This session should be really beneficial, as it aims to better acquaint you with the test material and help you more effectively study for future exams. Please let me know by e-mail no later than Sunday evening on October 2nd whether you are interested. The group sessions will be scheduled and conducted by Carolyn Blondin, who has 3 years of experience teaching this course. "

Section 3, sent October 19th by 2:30 by lead GTA

> "Based on your past exam performance, I would like to invite you to participate in one 30-minute small group study session this upcoming unit. The purpose of the session is to review test-taking strategies and help you develop a clearer idea of how information in each unit is linked to exam questions. This session should be really beneficial, as it aims to better acquaint you with the test material and help you more effectively study for future exams. Please let me know by e-mail no later than Sunday evening on October 20th whether you are interested. The group sessions will be scheduled and conducted by Carolyn Blondin, who has 3 years of experience teaching this course. "

2. Invitation to Continue Tier 2, message sent by primary researcher

a. Section 1, sent September 28th
b. Section 2 or Section 1, sent by October 19th
c. Section 3, Section 2, or Section 1, sent by November 7th

"Based on your recent exam performance, I would like to continue working with you again this unit in the same manner as last. Again, we will meet in small groups once prior to the exam and the purpose of the session is to review test-taking strategies and help you develop a clearer idea of how information in each unit is linked to exam questions. These sessions should continue to be beneficial, as they aim to better acquaint you with the test material and help you more effectively study for future exams. Please let me know by e-mail no later than Sunday evening on October 2nd (Thursday evening on October 20th; Tuesday evening on November 9th)."

3. Invitation to begin Tier 3 Intervention, message sent by primary researcher

a. Section 1, sent September 28th;
b. Section 2 or Section 1 sent by October 19th
c. Section 3, Section 2, or Section 1 sent by November 7th

"Based on your recent exam performance, I would like to increase the intensity of our review during Unit B (C, D, E) by offering two 45-minute private study sessions prior to the next exam. The purpose of these sessions will be to help you master key concepts discussed in class and covered in your daily homework assignments. One session will be provided following the first discussion day of this unit and the other will be provided following the last discussion day of the unit. These sessions should be really beneficial, as they aim to enhance your exam preparation strategies and ability to critically apply concepts discussed in the unit. In order to set up our first appointment in a timely manner, please let me know by e-mail no later than Sunday evening on October 2nd (Thursday evening on October 20th; Tuesday evening on November 9th) whether you are interested."

4. Invitation to continue Tier 3Interventions, message sent by primary researcher

a. Section 1, sent October 19th;
b. Section 2 or Section 1 sent by November 7th

"Based on your recent exam performance, I would like to continue working with you again this unit in the same manner as last, holding two 45-minute private study sessions prior to the next exam. As with the case before, the purpose of these sessions will be to help you master key concepts discussed in class and covered in your daily homework assignments. One session will be provided following the first discussion day of this unit and the other will be provided following the last discussion day of the unit. These sessions should be really beneficial, as they aim to enhance your exam preparation strategies and ability to critically apply concepts discussed in the unit. In order to set up our first appointment in a timely manner, please let me know by e-mail no later than Thursday evening on October 20th (Tuesday evening on November 9th) whether you are interested."

Appendix D

Treatment Integrity Checklist for Tier 2 Intervention

Date _____

Time Started _____ Time Ended _____

____ 1. Discuss what difficulties the student experienced during the last exam:

Notes:_____

_____

____ 2. Discuss methods the student used for preparation:

Notes:_____

_____

____ 3. Review select missed exam items on the recent exam: determine why inaccurate answer was chosen for each item and explain the correct answer

Items Discussed:_____

_____

____ 4. Review how select missed exam items corresponds to homework questions using the Exam Sources Document

____ 5. Provide general strategies for answering multiple-choice questions:

1. Understand what the question is asking
2. Eliminate unsupportable options
3. Minimize choice of options with extreme terminology

____ 6. Encourage general strategies for completing homework

____ 7. Remind the student of references available

Appendix E

Treatment Integrity Checklist for Tier 3 Intervention

Date _____ Session Number_____

Time Started _____ Time Ended _____

Homework Questions to Discuss:_____

____ 1. Select a homework question and ask the student to explain the answer

Notes:_____

_____

____ 2. Identify and discuss missing or inaccurate concepts in homework question

Notes: _____

_____

____ 3. Check for understanding of concepts discussed related to homework question

____ 4. Repeat steps 1-3 for each homework question selected for discussion

____ 5. Review any additional questions requested for review

Notes: _____

_____

Appendix F

List of Discussion Questions Selected for Accuracy Evaluation

Unit A

Day 1

What factors likely account for the differences in exercise patterns for boys and girls?

Given that girls experience more stress reduction from exercise than do boys, why don't more girls exercise regularly?

Day 2

Evaluate the validity of the claim that schools could be the single most influential institution in society in promoting healthy and productive living.

To what extent does Slide 14 indicate that our nation is on the right track in reducing drug use among high school seniors?

Day 3

Why is self-directed quitting typically a more successful way of giving up smoking than cold-turkey quitting? What assumptions underlie the two approaches?

What are the major similarities and differences between the dietary plans highlighted in this unit (original food pyramid, redesigned food pyramid, Atkins food pyramid, and my pyramid)?

What criteria should be used in evaluating the efficacy of a special dieting plan?

What are the principal differences in the recommended approaches to quitting smoking and altering one's food intake to lose weight?

Day 4

What are the most important pros and cons of abstinence-only versus abstinence-plus sexuality education?

Unit B

Day 1

How should the distinction between concrete and formal operational reasoning influence the types of learning experiences that teachers provide students?

According to Piaget, what experiences facilitate the natural development of conservation? How is conservation related to schemes and operations?

Day 2
What is the principal distinction between what IQ tests and achievement tests measure? Which provides more useful information about a child's cognitive development in school?

To what extent are IQ tests helpful to educators in serving the intellectual needs of children? How could curriculum-based assessment (CBA) be more useful than IQ tests in determining how to promote children's academic development?

Day 3
Which has greater academic value—determining students' IQ scores or their creativity scores? Why?

Which has the greatest potential for fairly and effectively assessing a child's cognitive potential, IQ tests, creativity tests, or critical thinking tests?

Day 4
Why is direct instruction among the most recommended approaches for remediating the deficits associated with identified learning disabilities?

What are the major differences between the whole language and phonics approaches for promoting reading skills?

Explain the respective roles of task analysis, curriculum-based assessment, drill and practice, criterion-referenced evaluation, and feedback in direct instruction?

Unit C

Day 1
What are the principal similarities and differences between Slavin's and Kohn's models of cooperative learning?

In what ways could cooperative learning be beneficial or detrimental to the academic development of high-achieving students?

Day 2
Explain how the combination of individual and group-reward contingencies would facilitate performance more than either individual or group contingencies separately.

What would be the pros and cons in CWPT of dividing the class on a random or ranking basis?

Do the logistics of CWPT legitimately qualify as peer tutoring or is there a more accurate label for this process?

Day 3
What appear to be the social characteristics of students who commit violent acts at school? How can schools prevent these social characteristics from turning into violent actions?

Why is classroom observation considered the most accurate method of assessing student relationships and social skills?

Day 4
Why would structured controversy be more effective in reaching conclusions about controversial issues than simply having students defend their own opinions?

How does the notion of perspective taking factor into structured controversy?

What are some likely contributors to an authoritarian parenting style?

Explain how parenting styles differentially affect students' school performance?

Unit D

Day 1
Compare the nature and predictive potential of the different self-concept models.

How are the notions of locus of control and self-efficacy alike and how are they different?

Contrast the ways high and low achievers account for success and failure experiences.

Describe transitions in causal attributions from kindergarten to the high school years.

Day 2
Contrast behavioristic and humanistic analyses of the relationship between behavior and feelings? Which model offers the greater potential for enhancing both behavior and feelings?

What are the different ways that extrinsic reinforcement can affect intrinsic reinforcement?

Explain the similarities and differences between punishment, extinction, and response cost.

Compare the behavioral and humanistic positions on educational goals.

Day 3
Explain how the behavioral approach can be used to achieve humanistic goals.

Compare optimists' assumptions of good and bad events with the explanations pessimists would likely advance for good and bad events.

Day 4
Evaluate the effectiveness and efficiency of using brainwave biofeedback to alter cognitive and behavioral characteristics associated with ADHD.

Explain the difference between rate of suicide as a comparative cause of death for adolescents/young adults versus older adults.

Unit E

Day 1
Explain the relationship between moral reasoning and moral conduct.

Explain how an incentive for higher scores and perceived surveillance affect the probability of cheating on the Circles Test. How are cheating and lying distinguished on the Circles Test?

How are Kohlberg's clinical interview and Rest's Defining Issues Test alike and different?

Day 2
Why have the instances of cheating in school nearly doubled in the last 30 years?

Why would high-GPA students be less likely to observe cheating than low-GPA students but more likely than low-GPA students to confront cheaters?

What are the likely effects of service-learning activities on both helper and the helpee?

Day 3
At international, national, and personal levels, how can humankind satisfy current needs for natural resources without undermining the habitability of the earth for future generations?

What is the current status of global warming and what are the prospects for global warming in the 21st Century? What do you see as the most compelling arguments for or against the reality of global warming?

Explain how early childhood personality tendencies could predict adult political ideology. How could genetic predispositions and environmental influences factor into your explanation?

Day 4
Identify ideological and psychological characteristics shared by religious fundamentalists (e.g., Christian, Jewish, Muslim, Hindu) worldwide. What do these characteristics suggest about the possibility of peace across cultures heavily dominated by fundamentalist ideology?

What are the similarities and differences between blind and constructive patriotism? When one politician attacks another politician's patriotism, what is the likely form of patriotism embraced by the attacker and by the attackee?

Appendix G

Sample Discussion Question Grading Rubric, Adapted from Galyon and Colleagues (in press)

1. **What factors likely account for the differences in exercise patterns for boys and girls? (p. 2)**

Exam Questions 44, 46

3
-The influence of social norms -It is more socially acceptable for boys to be aggressive and competitive -Girls prefer non-competitive physical activity

2. **Given that girls experience more stress reduction from exercise than do boys, why don't more girls exercise regularly? (p. 3)**

Exam Question 44

5
-Girls may have fewer role models who exercise -Females may receive approval for other activities -Other activities diminish time for exercising -In order to experience stress reduction girls must actually exercise first -Girls may not attribute stress reduction to exercise

3. **Evaluate the validity of the claim that schools could be the single most influential institution in society in promoting healthy and productive living. (p. 4, Slide 5)**

Exam Question 2

7
-A large number of children attend school -Health and physical education can be effective - Children spend more time per day at school than any other institution -Some parents may have more influence over their children -Many parents spend little time with their children or have poor health practices -If teachers model and teach good health practices, there could be far reaching effects -Many educators have not bought into the notion of improving societal health through what they teach and model.

4. **To what extent does Slide 14 indicate that our nation is on the right track in reducing drug use among high school seniors? (Slide 14)**

Exam Questions 36, 38

4
-High school seniors use of all three drug categories has decreased for all three patterns of use consistently from 1999 to 2008 -The tread has been especially favorable with respect to smoking rates -This pattern suggests that our nation is on the right track to reducing drug use -Because nicotine is a gateway drug for using other drugs, reduced smoking is likely to lead to a

reduction in the use of other drugs

5. **Why is self-directed quitting typically a more successful way of giving up smoking than cold-turkey quitting? What assumptions underlie the two approaches? (pp. 10, 11, Slide 20)**

Exam Question 33

7
-Self-Directed involves systematically developing a plan to stop smoking -Self-Directed assumes quitting should be supported by environmental changes -Some examples of environmental changes (include removing smoking cues, spending less time in smoking situations, spending more time with non-smokers, asking smoking friends to refrain from offering you cigarettes) -Environmental changes are followed by a target date in which smoking is no longer permitted -Cold turkey operates on the assumptions that quitting can occur smoker has enough will power -Both approaches involve total cessation from smoking -Cold-turkey involves no environmental changes

6. **What are the major similarities and differences between the dietary plans highlighted in this unit (original food pyramid, redesigned food pyramid, Atkins food pyramid, and my pyramid)? (Slides 22-25)**

Exam Questions 7, 8

12
-The original food pyramid and the redesigned food pyramid emphasize whole grains -The original food pyramid and the redesigned food pyramid emphasize high levels of vegetables and fruit -Original food pyramid does not distinguish between complex and refined carbs -The redesigned food pyramid minimizes white and refined carbs -Redesigned food pyramid minimizes red meat -Redesigned food pyramid emphasizes vegetable oils -The Atkins food pyramid emphasizes increasing proteins -Atkins food pyramid minimizes carbs -The redesigned food pyramid appears to be the most efficacious in facilitating long-term health - The redesigned pyramid allows for alcohol use in moderation - The redesigned pyramid includes exercise - My pyramid is individualized

7. **What criteria should be used in evaluating the efficacy of a special dieting plan? (p. 12)**

Exam Questions 4, 6

6
-Criteria should include whether the plan incorporates essential nutrients -Criteria should include whether the plan is balanced in regard to whole grains, vegetables, omega-3 fatty acids, fruit, protein, dairy products, and complex carbs (need 1) -Criteria should include whether the plan has low levels of saturated and trans fat -Criteria should include whether the plan has high levels of fiber

-Whether it can be maintained over the long-run -Drastic changes in diet could be detrimental to health

8. What are the principal differences in the recommended approaches to quitting smoking and altering one's food intake to lose weight? (pp. 10, 12, 13)

Exam Question 12

5
-Most successful prognosis for stopping smoking is to keep trying to quit -May need to use a variety of strategies to finally quit smoking -Trying a variety of diets to lose weight may result in decreased metabolism -Decreased metabolism makes it harder to lose weight in the future -Quitting smoking and losing weight are similar in that both are easier to accomplish when one exercises

9. What are the most important pros and cons of abstinence-only versus abstinence-plus sexuality education? (p. 15)

Exam Questions 47, 48, 50

5
-Abstinence-only pro: if followed, students are guaranteed to be safe from unhealthy sexual behaviors -Abstinence-only con: The total-abstinence expectation of abstinence-only sexuality education may be unrealistic for many students -Abstinence-only con: provides little or no information regarding safe sex -Abstinence-plus pro: if students do engage in sexual activity, they will obtain the knowledge of contraceptives to keep them safe -There is little evidence that discussing contraceptives encourages sexual activity

10. What moral issues, if any, should be considered in sexuality education? (p. 15)

Exam Questions 47, 50

4
-Could discuss whether premarital sex violates the well-being of your partner -Results of unwanted pregnancy -Increases the risk of sexually transmitted disease -Violates one's personal/religiously-based ethics about premarital or extra-marital sex

Appendix H

Tier 2 Intervention Acceptability Survey

	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
1. I felt better prepared after meeting one-on-one or in small groups.	1	2	3	4	5	6
2. This intervention helped me understand why I missed exam items.	1	2	3	4	5	6
3. Discussing missed items helped me answer questions on the next exam.	1	2	3	4	5	6
4. I used exam taking strategies discussed in intervention on the next exam.	1	2	3	4	5	6
5. Discussing preparation strategies helped me study for the next exam.	1	2	3	4	5	6
6. I would recommend this intervention for those that are struggling on exams.	1	2	3	4	5	6
7. I will use these study and test-taking tips in my other classes.	1	2	3	4	5	6

Appendix I

Tier 3 Intervention Acceptability Survey

	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
1. I felt better prepared with this intervention than the last one.	1	2	3	4	5	6
2. This intervention helped me understand key concepts in the unit.	1	2	3	4	5	6
3. Discussing homework questions helped me prepare for the exam.	1	2	3	4	5	6
4. I changed how I completed my homework following these sessions.	1	2	3	4	5	6
5. Discussing preparation strategies helped me study for the next exam.	1	2	3	4	5	6
6. I would recommend this intervention for those that are struggling on exams.	1	2	3	4	5	6
7. I will use these preparation tips in completing homework in my other classes.	1	2	3	4	5	6

Vita

Carolyn Anne Blondin was born in Normal, Illinois. She moved at age 4 to Chagrin Falls, Ohio, where she spent her childhood and adolescence. She obtained a B.A. in Psychology at Elon University in Elon, North Carolina in 2006. After graduating, she worked as a psychological assistant in Gainesville, Florida. Starting in 2008, Carolyn attended The University of Tennessee's School Psychology Ph.D. program. In December of 2010, she received a M.S. in Applied Educational Psychology. Carolyn will receive her Ph.D. in August 2013 upon completion of a year-long pre-doctoral internship with the Tennessee Internship Consortium in Knoxville, TN.