6-3-2016

Examination of the Impact of Students' Skill Levels on the Effectiveness of Evidence-Based Interventions for Improving Mathematics Fluency

Erica Fanning

Graduate Center, City University of New York

How does access to this work benefit you? Let us know!

Follow this and additional works at: http://academicworks.cuny.edu/gc_etds

Part of the Educational Psychology Commons

Recommended Citation

Fanning, Erica, "Examination of the Impact of Students' Skill Levels on the Effectiveness of Evidence-Based Interventions for Improving Mathematics Fluency" (2016). CUNY Academic Works.
http://academicworks.cuny.edu/gc_etds/1299

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects (2014-Present) by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
Examination of the Impact of Students' Skill Levels on the Effectiveness of Evidence-Based Interventions for Improving Mathematics Fluency

Author: Erica Fanning

Dissertation submitted to the Graduate Faculty in Educational Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York 2014

2016
This manuscript has been read and accepted for the Graduate Faculty in Educational Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Yung-Chi Chen, PH.D.

Date

Chair of Examining Committee

Bruce Homer, PH.D.

Date

Executive Officer

Georgiana Tryon, PH.D.

Erin Ax, PH.D.

Robin Codding, PH.D.

Jay Verkuilen, PH.D.

THE CITY UNIVERSITY OF NEW YORK
Abstract

EXAMINATION OF THE IMPACT OF STUDENTS' SKILL LEVELS ON THE EFFECTIVENESS OF EVIDENCE-BASED INTERVENTIONS FOR IMPROVING MATHEMATICS FLUENCY

By

Erica Fanning

Advisor: Professor Yung-Chi Chen

In recent years, many schools have attempted to address the academic difficulties of students with the use of response to intervention (RTI). RTI is a framework for matching instructional interventions with students’ academic skills in order to inform high-quality, evidence-based instruction. Numerous interventions, which have been derived from the acquisition and fluency stages of the instructional hierarchy, have been found to be effective. The present study employed a single-case multiple baseline across subjects design with an embedded alternating treatments design to examine the impact of students’ instructional level on the effectiveness of various interventions (fluency, peer tutoring, and combined, which included both performance feedback and peer tutoring) on the fluency scores for single-digit multiplication. The performance feedback intervention included feedback regarding number of correct responses within a two-minute period, which targeted fluency, while the peer tutoring intervention included a single-response repetition component that targeted accuracy. Results indicated that the performance feedback and peer tutoring interventions were not consistently effective, while the combined intervention was effective for students who began the study in the
instructional range. Increased Digits Correct per Minute (DCPM) scores positively correlated with generalization measures and students reported that they enjoyed the interventions and found them to be useful.
Table of Contents

Table of Contents ................................................................. vi
List of Tables ................................................................. xvi
List of Figures ................................................................. xvii
Chapter 1: Introduction .............................................................. 1
   The Need for Evidence-Based Interventions and Assessments in the Area of
      Mathematics ................................................................. 1
   Frameworks for Applying Academic Interventions in Schools ............ 3
   Academic Interventions: Need for Further Research ...................... 6
   A Method for Delivery of Intervention .................................... 10
   How Student Skill Level Might Predict Intervention Effectiveness ....... 11
   Research Questions ........................................................... 12
Chapter 2: Literature Review .......................................................... 13
   Current Mathematics Performance of Students in the United States .... 15
      Mathematics achievement and poverty ................................ 19
   Response to Intervention .................................................... 21
      Frameworks for decision making ....................................... 23
      Tiered interventions ....................................................... 24
      RTI and mathematics ..................................................... 25
      Need for evidenced based mathematics interventions ............... 27
   Two Instructional Frameworks for Matching Student Skills to Intervention .... 28
      Instructional hierarchy .................................................... 29
      Instructional delivery model .......................................... 34
      The relationship between the two models ............................ 37
Accuracy criteria as an expectation for student performance.............. 39
Fluency criteria as an expectation for student performance............... 41

  Historical fluency criteria.................................................. 42
  New fluency criteria.......................................................... 45

Curriculum Based Assessment: Procedures for Assessing Skill Level........ 49

  Subtypes of CBA................................................................. 51

Curriculum Based Measurement............................................... 52
Assessment Procedures.......................................................... 53

Academic Interventions.......................................................... 54

Acquisition interventions....................................................... 55

  Types of acquisition interventions....................................... 57
    Error-correction procedures............................................ 59
    Response repetition and mathematics................................. 61
    Response repetition and reading....................................... 62
    Single- verses multiple-response repetition
      procedures............................................................... 63
    Error-correction as a punishment..................................... 63
    Benefits of error-correction procedures................................ 66

Fluency interventions.......................................................... 67

  Types of fluency interventions......................................... 68
    Explicit timing.................................................................... 70

Peer Tutoring: An Efficient Method for Delivering Individualized

  Intervention............................................................................ 73
The importance of generalization .............................................. 99
Treatment integrity ............................................................ 100
Treatment acceptability ..................................................... 100
Research and Educational Implications ................................. 101
Research Questions ............................................................ 102
Hypotheses ......................................................................... 103
Chapter 3: Methods ............................................................ 105
Recruitment, Participants and Setting ................................. 105
Recruitment ......................................................................... 105
Setting .............................................................................. 106
Participants ........................................................................ 107
Measures ........................................................................... 116
AIMSweb Progress Monitoring System ............................... 116
AIMSweb Mathematics Computation (M-COMP) probes ........ 117
AIMSweb Mathematics Concepts and Applications (M-CAP) ... 119
Single-skill CBM probes ....................................................... 121
Treatment acceptability measures ........................................ 123
CBM Acceptability Scale (CMB –AS) .................................. 123
Kids Intervention Profile (KIP) ............................................ 124
Procedures ......................................................................... 124
Recruitment and training of research assistants ................... 125
Experimental Design ............................................................ 125
Baseline ............................................................................ 129
Performance feedback intervention

Peer tutoring intervention

Combined intervention

Testing for generalization

Procedural integrity

Treatment acceptability

Chapter 4: Results

Data Collection

Determination of instructional level

Treatment integrity

Treatment acceptability

Visual analysis of individual scores

Group A: Performance feedback intervention at the instructional level

Student 1

Student 2

Student 3

Student 4

Summary for students in group A performing at the instructional level

Group A (continued): Performance feedback intervention at the frustrational level

Student 5

Student 6
Student 7.................................................................176

Summary for students in group A performing at the
frustrational level....................................................178

Summary of the data for group A: Students receiving the performance
feedback intervention in isolation...........................178

Group B: Peer tutoring intervention for students at the instructional
level.................................................................182

Student 8.................................................................182

Student 9.................................................................185

Student 10...............................................................187

Student 11...............................................................190

Student 12...............................................................193

Student 13...............................................................195

Summary for students in group B performing at the instructional
level.................................................................197

Group B (continued): Peer tutoring for students at the frustrational
level.................................................................199

Student 14...............................................................199

Student 15...............................................................202

Student 16...............................................................204

Student 17...............................................................207

Student 18...............................................................209
Summary……………………………………………………………………243

Instructional level as a mediating factor………………………………….243

The impact of instructional level: Visual analysis………………………244

Instructional level: Tau-U analysis………………………………………246

T-tests for the instructional level………………………………………249

Summary……………………………………………………………………251

Mastery levels of performances…………………………………………252

Summary of Research Hypotheses…………………………………………253

Chapter 5: Discussion…………………………………………………………255

SES as a Mitigating Factor………………………………………………….256

Treatment Acceptability…………………………………………………..257

Generalizability…………………………………………………………….258

Effectiveness of the Performance Feedback and Peer Tutoring Interventions…260

Performance Feedback………………………………………………….261

Peer Tutoring…………………………………………………………….262

Summary……………………………………………………………………265

The Combined Intervention………………………………………………265

Comparison of the Tutee, Tutor, and Control Conditions………………267

Instructional Level as a Mediating Factor………………………………269

Performance Feedback Intervention……………………………………269

Peer tutoring intervention………………………………………………270

Combined intervention…………………………………………………272

Patterns across groups………………………………………………….275
Master Levels of Performance.................................................................276
Educational Implications.................................................................278
Limitations.......................................................................................279
Implications for Future Research......................................................282
Conclusion.........................................................................................284

Appendices

Appendix A. Parental Consent Form..................................................286
Appendix B. Student Assent Form.....................................................288
Appendix C. Sample Single Skill Assessment Probe............................293
Appendix D. Teacher Acceptability Scales..........................................301
  D-1. Teacher Acceptability Scale – Performance Feedback...............301
  D-2. Teacher Acceptability Scale – Peer Tutoring............................305
  D-3. Teacher Acceptability Scale – Combined....................................309
  D-4. Acceptability – Overview.........................................................313
Appendix E. Student Acceptability Scales..........................................314
  E-1. KIP – Performance Feedback Intervention.................................314
  E-2. KIP – Peer Tutoring Intervention...............................................318
  E-3. KIP – Combined Intervention....................................................324
Appendix F. Research Assistant Training Procedures..........................330
Appendix G. Procedures and Integrity Checklists.................................334
  G-1. Baseline Procedures.................................................................334
  G-2 Performance Feedback Procedures..........................................335
  G-3 Peer Tutoring Procedures........................................................337
List of Tables

Table 1. The Instructional Hierarchy .......................................................... 30
Table 2. The Instructional Delivery Model ..................................................... 36
Table 3. Student Information .................................................................. 109
Table 4. Skinner et al.’s Equated Sets of Multiplication Problems ............... 122
Table 5. Proposed Multiple Baseline Research Design .................................. 127
Table 6. Actual Multiple Baseline Research Design .................................... 128
Table 7. Mean Scores across Intervention Settings for Group A ................. 144
Table 8. Slopes Across Intervention Settings for Group A .......................... 145
Table 9. Slopes Across Intervention Settings for Group B .......................... 180
Table 10. Mean Scores across Intervention Settings for Group B ............... 181
Table 11. Summary of Findings from the Individual Analysis ...................... 217
across Groups ......................................................................................... 222
Table 13. Tau-U Effect Sizes for Intervention ............................................ 227
Table 14. Tau-U Effect Sizes for the Performance Feedback Intervention .......... 228
Table 15. Tau-U Effect Sizes for the Peer Tutoring Intervention .................. 228
Table 16. Tau-U Effect Sizes within the Combined Intervention: Group A .......... 231
Table 17. Tau-U Effect Sizes for the Combined Intervention: Group B .......... 232
Table 18. Tau-U Effect Sizes for Tutee and Tutor Conditions within the Peer Tutoring Intervention: Group B ............................................................ 236
Table 19. Tau-U Effect Sizes for tutee and Tutor conditions within the Combined Intervention: Group A ............................................................ 237
Table 20. Tau-U Effect Sizes for the Combined Intervention: Group B .......... 238
Table 21. Tau-U Effect Sizes for Instructional Groups……………………………………245
Table 22. Tau-U Effect Sizes for Control Condition across Instructional Group……..248
Table 23. Summary of Research Hypotheses and Findings…………………………254
List of Figures

Figure 1. Performance feedback intervention procedures……………………………..130
Figure 2. Peer tutoring intervention procedures…………………………………..132
Figure 3. Combined intervention procedures……………………………………..132
Figure 4. DCPM scores for each student across baseline, single intervention, and combined intervention phases………………………………………………147
Figure 5. DICPM scores for each student across baseline, single intervention, and combined intervention phases………………………………………………154
Chapter 1

Introduction

This chapter begins with an overview of the need for evidence-based interventions when teaching mathematics. It is followed by a summary of two frameworks for accurately and reliably identifying students’ needs to determine the most effective interventions. Next, specific interventions that have emerged from these frameworks will be described as well as procedures for their application in classroom settings.

The Need for Evidence-Based Interventions and Assessments in the Area of Mathematics

In spite of increased attention and urgency for students in the United States to perform well in the area of mathematics, various studies (e.g., National Assessment of Educational Progress (NAEP), 2012, 2013; Organisation for Economic Co-operation and Development (OECD), 2013b; Provasnik et al., 2012) have shown that a significant number of students consistently fail to demonstrate understanding of basic math skills. This is particularly true for students of poverty, as this subgroup consistently underperforms as compared to their higher-income peers (NAEP, 2012, 2013; OECD, 213b; Provasnik et al., 2012). Indeed, current instructional approaches are not effectively meeting the needs of all students. To date, educators do not seem to have access to the tools they need to make instructional decisions that teach even the most basic skills in mathematics to all students.

In response to the clear need for improved instruction in the area of mathematics, many schools use a Response to Intervention (RTI) framework in order to identify
students’ needs. RTI requires the use of universal screening measures to identify students’ abilities and to inform high-quality, evidence-based instruction (Fuchs & Fuchs, 2006). Further, progress monitoring is used to evaluate intervention effectiveness. Yet, for RTI to be effective, teachers need to be able to respond to the wide range of students’ needs. If educators are to apply RTI principles to the instruction of mathematics, they need a wide repertoire of evidence-based interventions from which they can choose appropriate strategies for students who present with various types of difficulties and characteristics (Codding, Hilt-Panahon, Panahon, & Benson, 2009; Lembke, Hampton, & Beyers, 2012). Thus, educators need a framework for matching instructional interventions with the particular needs of their students.

There is clearly a need for sound, empirically-based interventions in the area of mathematics that are effective for all students, particularly for those who are falling behind their peers and are coming low income homes. Despite this need, past research has been focused in the area of reading. As a result, there are relatively few empirically-supported interventions for increasing math skills (Codding et al., 2009; Gersten, Chard, et al., 2009; Lembke et al., 2012). Additionally, Gersten, Beckmann, et al. (2009) note that there is a lack of studies that have investigated cultural characteristics as mitigating factors, which means that there are relatively few effective interventions that can be identified for the subgroups of students who are falling behind their same-age peers, including students of poverty.

In addition to the need for more evidence-based interventions, it is necessary to have valid and reliable methods that can be used to match instructional interventions to the specific skill deficits that are being displayed by a particular student (Vaughn &
Fuchs, 2003). Such methods can help educators avoid the academic difficulties that frequently result from a mismatch between student skill and instructional materials (Ardoin & Daly, 2007; Gickling & Thompson, 1985; Haring, Lovitt, Eaton, & Hansen, 1978). Curriculum based assessment (CBA) has been presented as a way of addressing this need. It is defined as a procedure for determining the instructional needs of students based on the ongoing collection of data about student performance in relationship to curricular goals (Gickling & Thompson, 1985). According to Gravois and Gickling (2002), CBA allows educators to make evidence-based decisions when selecting interventions for students. Curriculum Based Measurement (CBM) is a form of CBA that involves specific standardized procedures and can be used to measure student growth over time (Deno, 2003).

Although CBM has been established as a way of making instructional decisions for students (Shapiro, 2011), there remains some debate regarding how to apply the data generated from CBA and CBM procedures in a meaningful way. As a result, Gickling (1977) derived the instructional delivery model as a way of conceptualizing students’ needs based on their accuracy scores from CBA measures. This framework will be briefly discussed in the following section.

**Frameworks for Applying Academic Interventions in Schools**

In schools, teachers are always making instructional decisions. Depending upon the culture of the school and the beliefs of individual teachers, the decisions range from adopting a pre-packaged curriculum to completely designing one’s own lesson sometimes using materials produced by others or creating one’s own materials. This means that decisions can be biased or influenced by factors that are not relevant to
educational success. However, there are frameworks that guide teachers in making objective and consistent instructional decisions.

There are two models that have been used, both independently and interdependently, for determining students’ instructional needs and for identifying the most effective instructional interventions. Gickling’s (1977) instructional delivery model is a multi-level model in which the levels are defined based on student performance on a specific task, and the focus is on defining students’ instructional level. The instructional hierarchy (Haring et al., 1978), on the other hand, is used to determine the type of instructional intervention students need.

The instructional delivery model defines three levels of performance: frustrational, instructional, and independent. These levels are determined based on the type of task, the number of items in the task, and how accurately the student can complete that task (Gickling & Thompson, 1985). When applied to basic skill areas, Gickling, Shane, and Croskery (1989) suggest that an instructional level for drill tasks would be represented when a student could complete a task with 70-85% accuracy. If students perform below this range, the assessed task falls in the frustrational range, and if they perform above the instructional range the task falls in the mastery range.

Within the instructional hierarchy (Haring et al., 1978), it has been suggested that students be grouped according to specific needs and then matched to research-based interventions. The instructional hierarchy contains four stages: acquisition, fluency, generalization, and adaptation; and it requires students to gain proficiency at one stage before progressing to the next one. For each stage, a different type of instructional intervention is recommended (Haring et al., 1978). For example, a student in the
acquisition stage requires modeling, feedback, and praise for effort, while a student in the fluency stage requires frequent opportunities for repetition and practice followed by targeted feedback regarding the accuracy and fluency of performance as well as praise for increased fluency (Haring et al., 1978). This paper will concentrate on the acquisition and fluency stages, as the goal of this study was to understand how to improve rates of fluent responding for basic math facts (e.g., single-digit multiplication).

The two models have been used together where the instructional delivery model serves as the framework for determining which type of intervention from the instructional hierarchy is best suited to meet student needs (Burns, Vanderheyden, & Jiban, 2006; Codding, Archer, & Connell 2010; Codding, Shiyko, Russo, Birch, Fanning, & Jaspen, 2007; Rhymer, Dittmer, Skinner, & Jackson, 2000; Rivera & Bryant, 1992). While the instructional delivery model provided an initial framework for ensuring that instructional interventions are matched to student skill levels, over the past three decades there has been much debate about the particular criteria that should be used (Burns et al., 2006). Some researchers have proposed that accuracy criteria be used to determine instructional level (e.g., Gickling & Thompson, 1985; Neef, Iwata, & Page, 1980; Robinson & Skinner, 2002), while others have advocated for the use of fluency scores instead (Burns et al., 2006; Daly & Martens, 1994; Deno & Mirkin, 1977; Vanderheyden & Burns, 2005). This paper will discuss the literature regarding these two approaches and show that fluency criteria are better indicators of student performance. Thus, Gickling’s (1977) model has been modified as fluency criteria and is now used by researchers and educators to determine instructional levels and define students’ needs.
These two models have become more relevant to current educational practice as schools are beginning to adopt RTI frameworks. Implementation of the RTI model requires that educators know how to identify specific skill deficits, use interventions to correct those skill deficits, conduct frequent measurement of the targeted skills, and evaluate student performance using single subject design methodology (Ardoin & Daly, 2007). These components of RTI reflect those of the instructional delivery model (Gickling, 1977) and the instructional hierarchy (Haring et al., 1978). Thus, it is logical to conclude that these models can be used together to provide guidance within the RTI model, and specific interventions from the instructional hierarchy have been developed in order to meet the needs of students in the various levels of the instructional delivery model.

Educators need tools for identifying and responding to students’ learning needs. Given the individual nature of each student, educators need tools that are based in research, can be easily employed, and are responsive to students’ individual needs within the larger classroom.

**Academic Interventions: Need for Further Research**

Currently, there is a dearth of research to guide educators when making decisions about effective mathematical interventions once fluency criteria have been used to assess their instructional needs. However, there is research to support the use of various types of interventions that were developed as acquisition or fluency interventions according to recommendations from the instructional hierarchy. For example, researchers have explored various types of acquisition interventions within the context of mathematics (e.g., Burns, 2005; Codding et al., 2010; Codding, Eckert, Fanning, Shiyko, & Solomon,
An important gap identified by Codding et al. (2009) is that relatively few acquisition interventions have enough support to truly be considered evidence-based. This is particularly true when considering the lack of research with regard to generalization effects and treatment integrity (Codding et al., 2009). Additionally, some types of interventions have received more attention than others. For example, error-correction procedures, in which students are provided with immediate corrective feedback and required to engage in additional practice following errors, have been found to be effective in the area of reading (e.g., Ferkis, Belifiore, & Skinner, 1997; Marvin, Rapp, Stenske, Rojas, & Bartlett, 2010; Worsdell et al., 2005); but research investigating the use of these types of procedures in the area of mathematics has been limited to only two studies (Rapp et al., 2012; Rhymer et al., 2000). Although research in the area of mathematics is limited, all studies with regard to error-correction procedures have yielded positive gains in accuracy and/or fluency rates of reading and math skills (Ferkis et al., 1997; Marvin et al., 2010; Rapp et al., 2012; Rhymer et al., 2000; Worsdell et al., 2005) for students of varying ages, which indicates that error-correction procedures are beneficial for student learning. Given the general success of these types of interventions, especially in reading, further investigation of the utility of such procedures for increasing accuracy and fluency scores for basic math tasks is necessary.
Researchers have also examined the effectiveness of several types of fluency interventions (e.g., Carson & Eckert, 2003; Cates & Skinner, 2000; Codding et al., 2007; Freeland & Noell, 2002; Hawkins, Skinner, & Oliver, 2005; Lannie & Martens, 2008; Magg, Ried, & DiGangi, 1993; McDonald & Ardoine, 2007; Monterello & Martens, 2005; Panahon & Martens, 2013; Rhymer & Morgan, 2005; Rhymer, Skinner, Jackson, McNeil, Smith, & Jackson, 2002; Skinner, 2002; Rhymer & Morgan; 2005; Van Houten & Thompson, 1976). While many studies have supported the use of fluency interventions, Burns, Codding, Boice, and Luikto (2010) report that relatively few single-case design studies exist in which researchers consider the instructional level of students before applying fluency interventions. More specifically, results from their meta-analysis found that there is not a sufficient number of studies to draw specific conclusions regarding the effectiveness of fluency interventions when applied to students in the instructional range (Burns et al., 2010). This is problematic given that fluency interventions are designed to meet the needs of students performing in the instructional range.

Explicit timing is a type of fluency intervention that has been found to be effective (e.g., Codding et al., 2007; Rhymer et al., 2002; Van Houten & Thompson, 1976) but has not been thoroughly examined when considering student skill level as a mediating factor. Although explicit timing has been found to be effective in isolation, little research exists with regard to how this simple procedure might be combined with other effective procedures (Rhymer et al., 2000). If a time constraint can motivate students to work more quickly and accurately when completing a different type of intervention, they may be able to build accuracy and fluency simultaneously, which
would allow them to move through the acquisition stage at a faster pace (Rhymer et al., 2000). Thus, this is an area in need of further investigation.

Interestingly, few studies have employed the use of multiple types of interventions as one treatment package. In fact, Burns et al. (2010) note that there has only been one study in which both acquisition and fluency interventions have been combined into one treatment package. When these two types of interventions were combined, fluency scores for participants showed significant improvement (Rhymer et al., 2000). This gap in the literature is notable because Codding, Burns, and Lukito (2011) found that students benefit more from interventions containing multiple components.

To date, research has demonstrated that what educators do makes a difference in student learning. Moreover, we know that evidence regarding students’ learning needs provides educators with much-needed information when making instructional decisions. There is a clear gap in our understanding, however, as to the types of interventions that are most effective for students at different instructional levels. There is even less information about the effectiveness of combining interventions. It is known that performance feedback is typically an integral part of both acquisition and fluency interventions (Haring et al., 1978); however, relatively few studies have examined the utility of this practice and the results have been mixed (Codding et al., 2007). In this context, research is lagging behind current practices. Explicit timing interventions allow students to see how many items they have completed in a specified time period (Codding et al., 2007), and error-correction procedures rely heavily on immediate corrective feedback following each item (Rapp et al., 2012). While initial results from research on
these practices are either mixed or promising, little is known about their effectiveness when they are combined. This highlights the need for further research assessing the utility of adding performance feedback to interventions in which there are time constraints as well as targeted feedback on performance.

**A Method for Delivery of Intervention**

Based on the potential effectiveness of providing immediate, individualized feedback within acquisition and fluency interventions, and the noted skill deficits for students in the area of mathematics, these types of interventions are a line of inquiry worth pursuing. One challenge, however, is providing the individualized feedback that is required within the context of acquisition and fluency interventions, as that would typically require more educational resources (i.e., teacher time) than is feasible in a classroom. This is particularly true for acquisition interventions where corrective feedback for individual items is required. Class-wide peer tutoring (CWPT) provides a potential solution to this challenge because it is a method for providing students with immediate, specific, and individual feedback that can minimize the educational resources necessary to provide individualized interventions for students (Cates, 2005). Class-wide reciprocal peer tutoring (CWPT) procedures are a group of instructional strategies in which all the students in a class are trained and supervised by the classroom teacher to instruct other students, in which they take turns serving as the tutor and the tutee (Kunsch, Jitendra, & Sood, 2007). While many of the error-correction procedures previously discussed have utilized procedures in which an adult and a student are working one-on-one, CWPT is a promising structure for expanding the impact of these procedures.
While significant support for the use of CWPT exists (e.g., Hawkins, Musti-Rao, Hughes, Berry, & McGuire 2009; Kunsch et al., 2007 Meneses & Gresham, 2009; Robinson, Schofield, & Steers-Wentzell, 2005), there is a lack of literature regarding the implications of student skill level in relationship to the effectiveness of these types of interventions. Additionally, although several studies have included minority students and at-risk youth, Robinson et al. (2005) note that there is a need for additional research regarding students who receive free and reduced lunch.

**How Student Skill Level Might Predict Intervention Effectiveness**

Researchers report that an instructional match between skill level and intervention is inherent to successful academic instruction (Burns et al., 2006; Coddington et al., 2007; 2010; Gickling, 1977; Haring et al., 1978; Rhymer et al., 2000; Rivera & Bryant, 1992; Vanderheyden & Burns, 2005). Researchers have applied fluency criteria to data generated from CBM measures to identify the level of the instructional delivery model in which students are performing (Burns et al., 2006; Coddington et al., 2007; 2010; Rhymer et al., 2000; Rivera & Bryant, 1992; Vanderheyden & Burns, 2005). They then choose an intervention from the appropriate stage of the instructional hierarchy (Haring et al., 1978) to address students’ needs.

Guided by these procedures, this study sought to expand the current literature by exploring the effectiveness of performance feedback, combined interventions, and the use of a reciprocal CWPT procedure as a method for delivering individual interventions to students. Students’ skill levels during baseline procedures and socio-economic status were considered as potential factors that mediate or moderate the effectiveness of the interventions. Further, the different roles within the peer tutoring procedure (i.e., tutor,
tutee) were examined to see how they may differ in their levels of effectiveness for students. In sum, based on the above discussion, the present study sought to answer the following questions:

**Research Questions**

The present study sought to answer the following questions:

**Question 1:** Is students’ SES, as indicated by free lunch status, related to the effectiveness of the performance feedback, peer tutoring, and combined interventions?

**Question 2:** Will increased fluency scores correlate with a discernible increase in students’ scores on generalization measures?

**Question 3:** Are the performance feedback and peer tutoring interventions in isolation effective for increasing fluency of performance with regard to single-digit multiplication?

**Question 4:** Are the performance feedback and peer tutoring interventions more effective when they are combined into one multicomponent treatment?

**Question 5:** Is the tutee role of the peer tutoring intervention more effective than the tutor role?

**Question 6:** Does instructional level affect the effectiveness of these interventions?

**Question 7:** Does instructional level affect the amount of time necessary for students to achieve mastery levels of learning?
Chapter 2
Literature Review

This chapter contains 12 sections. The first section begins with the findings of three major studies that suggest too many students in the United States are failing to develop basic understanding of mathematics. The first section then highlights research that suggests students of poverty are less likely to meet grade level standards in the area of mathematics, indicating that this is a particular sub-group in need of intervention. The second section begins with an explanation of how Response to Intervention (RTI) can be used in schools to address student skill deficits by using regular assessment to make informed instructional decisions. Next, this section explains the differences between two decision making models (i.e., the standard protocol framework and the problem-solving approach) that can be applied to RTI procedures, followed by a detailed description of the tiers of RTI. Finally, this section details how RTI is applied to the area of mathematics and the lack of evidence-based interventions in this academic area is discussed.

The third section of this chapter includes descriptions of the instructional hierarchy and the instructional delivery model; both of which serve as conceptual frameworks guiding the application of appropriate academic interventions based on students’ skill level in area of mathematics. It also includes a review of existing literature regarding the different accuracy and fluency criteria that have been applied to these models. Additionally, this section provides evidence to support the use of specific fluency criteria for making instructional decisions. The fourth section provides an explanation of Curriculum Based Assessment (CBA) and Curriculum Based Measurement (CBM) and explains how its standardized assessment procedures can be
used in conjunction with fluency criteria to make informed instructional decisions within the frameworks of the instructional hierarchy and the instructional delivery model.

The fifth section provides a detailed explanation of acquisition and fluency interventions within the instructional hierarchy. This section then includes a review of research studies that have provided support for specific acquisition and fluency interventions. A focus is placed on the acquisition intervention known as “error-correction” and the fluency intervention known as “explicit timing” because these interventions were used in the present study. The sixth section includes a discussion of the literature regarding peer tutoring as a method for delivering academic interventions. The discussion centers on the utility of reciprocal peer tutoring and demonstrates how class-wide peer tutoring (CWPT) can be used to decrease the demand of educational resources when implementing individualized interventions. The seventh section highlights the importance of various types of performance feedback across academic interventions. It also includes a review of existing literature regarding the possibility that performance feedback may be inherently useful in isolation or as an additional component to existing interventions.

The eighth section includes a detailed review of the only intervention that combined acquisition and fluency interventions into one treatment package in an attempt to improve student skills in the area of mathematics. It includes a rationale for further research and provides the foundation for the present study. Section nine synthesizes the research in this chapter on current student performance in the area of mathematics, the importance of fluency, the application of RTI, the use of CBM in conjunction with fluency criteria, and the combined use of the instructional hierarchy and the instructional delivery model.
delivery model. This synthesis focuses on the utility of over-correction interventions and explicit timing for the area of mathematics and provides an argument for the combined use of these procedures within the framework of CWPT. The inherent importance of performance feedback is discussed and the rationale for exploring it in isolation is presented. Additionally, the importance of treatment integrity and acceptance are discussed. Finally, a rationale for the present study is also provided in this section. Research and educational implications are provided in section eight, while research questions and hypotheses of this study are described in sections nine and ten, respectively.

**Current Mathematics Performance of Students in the United States**

Numerous studies (e.g., NAEP, 2012, 2013; OECD, 2013b; Provasnik et al., 2012) demonstrate that there is cause for concern regarding student performance in mathematics in the United States. More specifically, a significant number of students are consistently failing to demonstrate understanding of basic math skills and struggle to apply these skills in higher level reasoning and applications assessments. This is particularly true for female, black, and Hispanic students, as well as students of poverty, as these subgroups consistently underperform on these assessments. According to the National Assessment of Educational Progress (NAEP) and Trends in International Mathematics and Science Study (TIMSS), student performance in mathematics has increased over the past 20 years; however, the improvement has been minimal in recent years (NAEP, 2013; Provasnik et al., 2012).

The NAEP, TIMSS, and Program for International Study Assessment (PISA) are three ongoing wide-scale studies that utilize periodic criterion-referenced academic
assessments over time to track educational trends (NAEP, 2013; OECD, 2013b; Provasnik et al., 2012). While NAEP is a national study, TIMSS and PISA are two significant international studies that provide useful information about student performance in mathematics in the United States. While the first NAEP assessments were administered in 1969, current data are reported based on assessments that have been administered every two years, beginning in 1990. The TIMSS began in 1995 and has been administering assessments every four years, while the PISA began in 2000 and has been administering assessments every three years. Over the years, all three studies demonstrate that many students in the United States consistently underachieve on these mathematics assessments indicating a need to improve mathematics instruction in this country.

The abovementioned studies have consistently yielded concerning results in the mathematics assessments for students in the United States. In 2013, the scores on the NAEP assessment indicate that only 42% of fourth-grade students and 36% of eighth-grade students were able to demonstrate proficient levels of performance (NAEP, 2013). This indicates that less than half of the participants were able to consistently apply integrated procedural knowledge and conceptual understanding to mathematical problem solving (NAEP). Perhaps even more concerning is that 17% of fourth-grade students and 26% of eighth-grade students performed below basic levels of performance, suggesting that they are likely to be performing well below grade level in the area of mathematics as they have not acquired basic knowledge of essential math concepts. The TIMSS study continues to highlight this trend, as 19% of fourth-grade students and 32% of eighth-grade students failed to apply basic mathematic knowledge to straightforward situations
(Provasnik et al., 2012). Similarly, the most recent PISA results indicated that 26% of students in the United States failed to meet basic baseline standards (OECD, 2013b), which means that a quarter of 15-year-olds did not achieve minimum outcomes of mathematics instruction. When taken together, these results indicate that a large proportion of students in the U.S. are failing to achieve basic levels of understanding in the area of mathematics. Furthermore, the results indicate that an even larger number of students are unable to apply basic skills to applications and problem solving concepts. This indicates a need for educators to gain better understanding of how to address these clear deficits.

Moreover, it is distressing to note that these studies provide evidence that low-performing students continue to suffer over time. Data from the TIMSS study indicate that there was only a slight decrease in the proportion of underperforming fourth graders between the 2009 and 2011 test administrations, and no improvement for eighth-grade students during this time frame (Provasnik et al., 2012). Similarly, data from the NAEP 2009 and 2013 test administrations indicate that the proportion of students who fell at or below the tenth percentile did not show improvement during that time period (NAEP, 2013). Finally, data from the 2012 PISA study show that there has been no decrease in the proportion of underperforming students in the United States since 2003 (OECD, 2013). In sum, these data indicate that the number of students who are struggling to attain basic understanding of mathematics is not diminishing over time. It could be inferred that current instructional practices are not meeting the needs of students who are struggling the most. It is necessary for educators to develop and utilize evidenced based practices to better meet the needs of these students.
While the proportion of the lowest-performing students in mathematics has not increased in the past decade, the data from the PISA study also show that students from the United States are not competitive with their counterparts in other countries. They yield a national average that consistently falls below the overall international average and below the national average of 25 of the 34 OECD countries (OECD, 2013b). The TIMSS study is slightly more encouraging. Students in the United States performed above the international average in both the fourth and eighth grades. In addition, the United States was ranked within the top 15 out of 57 participating educational systems in fourth grade and within the top 24 of the 56 participating systems in eighth grade (Provasnik et al., 2012). When considering the differences in the types of problems and purposes of the PISA and TIMSS it may be that students in the United States are better able to perform straightforward mathematics tasks, but struggle when asked to apply these skills to real world situations, as required in the PISA assessment.

To summarize, data indicate that students in the United States consistently underperform on national and international assessments (NAEP, 2012; 2013; OECD, 2013b; Provasnik et al., 2012). Furthermore, the number of students in the United States who are failing to meet basic standards on these assessments is remaining consistent over time. Thus, it can be inferred that current educational practices are not meeting the needs of the students who are showing the greatest deficits in the classrooms. This means it is necessary to assess current educational practices to ensure that all students are able to achieve greater levels of success in the area of mathematics. Beyond these results, there is a particular relationship between mathematics achievement and poverty in the United States, which we will explore in the next section.
Mathematics achievement and poverty. While all students in the United States have access to education, there is a difference in mathematics performance between those students in poverty and those who are not. We define students in poverty as those who qualify for free- and reduced lunch program. This federally-funded program provides free and reduced lunches for students based on income level and family size (Department of Agriculture, 2013). NAEP (2012) indicated that although math scores improved for students across all income levels, a persistent gap remains between students who pay for lunch, and those who qualify for free and reduced lunches. More specifically, in 2011, 74% of all fourth-grade students who scored below the twenty-fifth percentile qualified for free or reduced lunch (NAEP, 2012). This percentage increased slightly to 76% in 2013 (NAEP, 2013). This statistic is not yet available for 2013 for eighth-grade students, but in 2011, 68% of eighth-grade students who performed at or below the 25th percentile qualified for free and reduced lunch. Scores of this nature indicate that these students failed to demonstrate basic levels of performance, suggesting that they have not acquired basic knowledge of essential math concepts. When taken together, these percentages indicate that a disproportionate number of students who qualify for free and reduced lunch underachieve in mathematics. Given current economic trends, it is clear that there is an unfulfilled need for improving this subgroup’s performance in mathematics.

Student achievement seems to directly correlate with income level. NAEP (2013) indicated that students who qualify for free lunch yielded the lowest test scores, while students pay for lunch earned the highest test scores. Students who qualify for reduced lunch earned test scores that fell in the middle of these two groups. In 2013, 28% of fourth-grade students who qualified for free lunch failed to meet basic standards. Even
worse, 41% of eighth-grade students who qualified for free lunch failed to meet basic levels of performance. These numbers are abysmal when considering that only 7% of students who were not eligible for free or reduced lunch failed to meet basic standards. A similar trend is seen when considering the students who qualify for reduced lunch. Within this sub-group, 15% of fourth graders and 29% of eighth-graders failed to meet basic standards (NAEP, 2013). Based on these data, it seems evident that there is a correlation between poverty and poor performance in the NAEP mathematics assessment.

The results of TIMSS and PISA studies also supported the notion that students from low-income families continue to fall behind their higher income peers in the United States (OECD, 2013b; Provasnik et al., 2012). The PISA study found that 15% of the variance in student performance in the United States can be accounted for by socio-economic status. Similarly, the TIMSS study reported a pattern in which schools that have a lower percentage of students who receive free/reduced lunch typically yielded a higher average score on the mathematics assessment than schools that have higher percentages. Additionally, for fourth-grade students, schools with at least 50% of students who receive free/reduced lunch yielded average scores that fell significantly below the scores of schools with lower proportions of low-income students (Provasnik et al., 2012). Here again, there is a link between poverty and poor performance on internationally respected mathematics assessments.

The PISA findings expand the discussion regarding the poor performance of economically disadvantaged students. Their data indicate that only 5% of economically disadvantaged students scored above expectations based on their economic means (OECD, 2013b), which is significantly below the international OECD average of 7%,
indicating that fewer students of poverty in the United States are likely to overcome their difficulties than are students in many other countries. The report further shows that 5% is approximately one third of the proportion of economically disadvantaged students in Hong Kong-China, Macao-China, Shanghai-China, and Viet Nam who surpass income-based expectations. The OECD (2013b) reported that socio-economically disadvantaged students may underperform because they tend to show less engagement, drive, motivation, and self-beliefs than advantaged students. These may be areas to consider when developing educational interventions and instruction.

Based on the results of NAEP (2012), TIMSS (Provasnik et al., 2012), and PISA (OECD, 2013b), it is clear that students in the United States continue to struggle in the area of mathematics; yet what would be most effective in terms of intervention is not clear. There is need for sound, empirically-based interventions in the area of mathematics that can work for all students, specifically for those who are falling behind due to economic circumstances. The next sections will include a discussion of Response to Intervention (RTI), which may be a possible framework for conceptualizing improved mathematics instruction and intervention.

Response to Intervention

In recent years, many schools are attempting to address the academic difficulties of students through a system called Response to Intervention (RTI). RTI is an early detection, prevention, and support system that allows for the identification of struggling students and provides support as necessary so that they do not fall behind (Gersten, Beckman, et al., 2009). It is a multi-tiered system that includes universal screening,
progress monitoring, high quality, evidenced based instruction, and evidenced based decision-making (Fuchs & Fuchs, 2006).

With the reauthorization of Individuals with Disabilities Education Act (IDEA) in 2004, schools were given an alternative to relying on the discrepancy between cognitive ability and achievement for identifying students as eligible for special education under the category of Specific Learning Disability (SLD). Specifically, reauthorized IDEA stated that a student can be classified as having an SLD if he/she does not respond to scientific, research-based intervention as a part of the evaluation procedures (IDEA, 2004). When this law was signed into action, it allowed for more educational flexibility that caused more schools to begin the implementation of systems that fall under the RTI category (Hoover & Love, 2011).

Research and practice has led researchers to draw two general conclusions regarding the use of RTI procedures. First, RTI models have been successfully implemented in schools to meet the needs of struggling learners. Second, RTI models look different and are presented in different formats across schools and districts (Lembke et al., 2012). This means that procedures can be applied in various ways to meet the needs of individual schools and the diverse populations they serve.

Vaughn and Fuchs (2003) reported that RTI models screen all children for academic and behavioral problems, monitor the progress of children at risk for difficulties in these areas, and provide increasingly intense interventions based on the response to progress monitoring assessments. These procedures allow for school professionals to identify problems earlier, determine if students are responding to the current programming, and provide services to all students who are in need. A benefit of
this type of procedure is that all struggling students are allowed an opportunity to receive intervention, which means that additional services are not only provided to students who qualify for special education.

Bramlett, Cates, Savina, and Lauinger (2010) stated that the purpose of RTI is to provide solid evidenced based interventions within the student’s current placement before considering a change in placement. Thus, it is essential that interventions include specific evidence-based strategies for students who are not achieving similarly to their same grade peers. The central elements of RTI include the use of scientific, research-based interventions in general education; measurement of student response to the intervention; and use of data to modify the type, frequency, and intensity of intervention (Kavale & Spaulding, 2008).

**Frameworks for decision making.** Currently there are two frameworks under which the RTI process can be conceptualized. These frameworks provide different procedures for determining which types of interventions to use and when to use them. When service providers use the same evidence-based interventions for all children with similar problems, the standard protocol framework is being utilized (Fuchs, Mock, Morgan, & Young, 2003). Lembke, McMaster, and Stecker (2010) explained that, within this framework, manualized interventions are purchased and followed with fidelity in conjunction with the core curriculum instruction. It has been noted that these manualized types of interventions can be highly effective, but they are typically designed to improve the skills of a particular kind of student. Because these treatments require the educator to follow a specific protocol, interventions cannot be tailored to meet individual needs.
The second and more commonly practiced framework involves a problem-solving approach in which a shared decision-making team identifies a behavior or academic problem and proposes strategies and interventions to remediate the problem. Ongoing assessment is used to evaluate the effectiveness of the interventions, which allows for decisions to be made regarding whether the intervention should continue or a change in instruction should be made (Tilly, 2002). This process can be explained through a four phase procedure that includes: (a) a problem identification phase; (b) a problem analysis phase; (c) an intervention implementation phase; and (d) a program evaluation phase (Fuchs et al., 2005). Tilly (2002) notes that this is a more individualized procedure, as it allows for better understanding of the unique needs of each student when implementing interventions. When using this framework, teachers must determine the unique needs of each student through assessment and then apply a relevant evidenced based intervention. This process not only allows teachers to choose individualized interventions for students, but it can also be relatively more cost effective as the evidenced based interventions are often available for free, as opposed to purchasing an expensive intervention package (Tilly, 2002).

**Tiered interventions.** While there are two frameworks under which RTI can be conceptualized, a defining characteristic of all forms of RTI is that they must include a multi-tiered approach. Typical models include a three-tiered approach, although some argue that a fourth tier should also be included. In the three tiered approach, tier 1 includes high quality instruction provided for all students in general education, where students receive positive reinforcement for appropriate social and academic behavior. This tier is typically thought of as the point at which primary prevention begins (Lembke,
et al., 2012). Tier 2 typically includes small-group intervention for students whose performance and rate of progress continues to lag behind same-grade peers, and more individualized behavior supports when necessary. Tier 3 includes intensive individualized interventions, which most likely include special education services and/or the initiation of processes to determine special education eligibility. Researchers suggest that typically tier one meets the needs of eighty percent of the student population, while tier two interventions are needed for approximately twenty percent and approximately five percent of students require intervention at the tier 3 level (Fletcher & Vaughn, 2007; National Research Center on Learning Disabilities [NRCLD], 2007).

Reschly (2005) argued that individualized intervention should be attempted before determination for special education takes place, therefore creating a fourth tier. However, because a four-tiered model is not typically used, the next section will focus on how to implement appropriate mathematics interventions within the framework of a three-tiered model. Although RTI is a tiered model, Shinn (2007) noted that services are provided along a continuum in which all students receive whole group instruction or support while students who are exhibiting difficulties are identified to receive interventions that are provided at the necessary level of intensity.

**RTI and mathematics.** Although the focus has recently shifted to include the area of mathematics, it has been noted that the majority of the research regarding the utility of RTI has been conducted in the area of reading (Lembke et al., 2012). Although Lembke et al. (2012) hypothesize that the shift has occurred because educators have realized that many students who struggle with reading also struggle with math, perhaps there are more fundamental reasons for enhancing the instructional practices in the area
of mathematics. For example, an understanding of mathematics is essential for many basic life activities, including purchasing goods and services, balancing a budget, and meeting typical work demands (Minskoff & Allsopp, 2003). Additionally, high stakes tests such as the SAT’s, which are used for college acceptance decisions, and statewide assessments that are used for No Child Left Behind (NCLB) include mathematics sections (Minskoff & Allsopp, 2003). These tests can be used for teacher, school, and student evaluation, which necessitates that there is a heavy emphasis on helping students achieve higher scores on these assessments. Finally, an understanding of mathematics is necessary for students to succeed in other subjects such as science, economics, and computer literacy (Minskoff & Allsopp, 2003). Thus, without a basic knowledge of mathematics, students may struggle to pass classes and standardized tests, which could lead to academic failure. When taking these factors into account, it seems logical that educators would want to ensure that sound instructional practice is developed for the area of mathematics.

As the focus shifted to the area of mathematics, several recommendations have been put forth with regard to the types of interventions that should be implemented across tiers. The National Center on RtI (2010) stated that tier 1 must include high quality instruction that implements the core curriculum with fidelity. Although specifics regarding what qualifies as “high quality” have not been agreed upon in the literature (Gersten, Beckmann, et al., 2009), most researchers believe that differentiated instruction in conjunction with flexible groupings and peer tutoring is essential to student learning at this tier (Lembke et al., 2012). Tier 2 interventions typically take place in a small group setting, four to five times a week and range from 20-40 minutes per session (Lembke et
al., 2012) for approximately 10-15 weeks (The National Center for RtI, 2010). Gersten, Beckmann, et al. (2009) noted that instruction at this level must be explicit, systematic, and supplemental to the core curriculum. More specifically, it was suggested that students need to be provided with repeated opportunities for modeling and guided practice in which they receive corrective feedback throughout the intervention. Gersten, Beckmann, et al. reported that instruction at tier 3 should include one-on-one tutoring with intensive instruction. It was reported that this instruction should include explicit instruction, opportunities for verbalization of mathematical reasoning, various visuals, a logical sequence of examples, and frequent performance feedback.

In addition to the specific guidelines for tiered intervention, Gersten, Beckmann, et al. (2009) also provided general recommendations to be used for any students who are receiving interventions within tier 2 or 3 in the area of mathematics. First, they noted that the intervention should be explicit and systematic and should provide models of proficient problem solving and frequent cumulative review. They noted that interventions should be focusing intensely on working with whole numbers from kindergarten through grade 5 and on rational numbers from grades 4 through 8, and they should also receive specific instruction with regard to solving word problems. Gersten, Beckmann, et al. also noted that it was important for students at all grade levels to devote approximately 10 minutes in each intervention for building fluent retrieval of basic arithmetic facts.

**Need for evidenced based mathematics interventions.** It has been noted that the ratio of reading to mathematics interventions has grown from 16:1 in 1986-1995 to 5:1 in 1996-2005 (Gersten, Chard, et al., 2009). While this shows a significant increase in the
number of research studies attributed to the area of mathematics, there remains a need to
close the gap. Furthermore, it is necessary for educators to consider cultural and
linguistic characteristics so that they can choose interventions that have been shown to be
effective for the particular students with whom they are working (Gersten, Beckmann, et
al., 2009). Thus, if educators are to apply RTI principles to the instruction of
mathematics, it is necessary to have a wide repertoire of evidenced based procedures
from which they can choose appropriate strategies for students who present with various
types of difficulties and characteristics. As previously noted, students who receive free
and reduced lunch are one subgroup who are underperforming in the area of mathematics.
Therefore, it is necessary to ensure that the research based interventions have also been
validated for use with this population.

One of the key components of RTI is matching instructional interventions to the
specific skill deficits that are being displayed by a student (Vaughn & Fuchs, 2003).
Thus, it is necessary to have a consistent and evidenced based process for identifying
student needs as academic difficulties can result from a mismatch between student skill
and instructional materials (Ardoin & Daly, 2007; Gickling & Thompson, 1985; Haring,
et al., 1978). The following section includes a discussion of research based models that
may guide educators in making informed decisions regarding appropriate interventions
for student need.

Two Instructional Frameworks for Matching Student Skills to Intervention

There are two primary models for matching student skills to instructional
interventions: the instructional hierarchy and the instructional delivery model. Both of
these models have been relied upon over the past 30 years to determine the type of
mathematics instruction that would be most effective by seeking to match instruction with particular student needs. While the goal is the same, there are distinct differences between the two models.

**Instructional hierarchy.** The instructional hierarchy, a heuristic proposed by Haring et al. (1978), is a guide for identifying effective instructional interventions by grouping students according to their specific needs and then matching them to research-based interventions. The instructional hierarchy is a four-stage developmental approach (i.e., acquisition, fluency, generalization, and adaptation) requiring students to master a specific academic skill before progressing to the next one (see Table 1 for an overview of stages, academic skills, and suggested interventions). It delineates a developmental trajectory describing student performance from novice to fluent stages. In the first stage, acquisition, students perform slowly and inaccurately while working toward understanding the skill. In the second stage, fluency, students are performing slowly and accurately albeit with faster response rates than in the acquisition stage. In the third stage, generalization, students work toward applying the targeted skill to a new setting, while in the fourth stage, adaptation, they are learning to use the skill to solve problems (Haring et al., 1978). The stages will be described in more detail and paired with specific suggestions for intervention throughout this section.
Table 1

*The Instructional Hierarchy*

<table>
<thead>
<tr>
<th>Stage</th>
<th>Student Skills</th>
<th>Suggested Interventions</th>
</tr>
</thead>
</table>
| Acquisition    | ● Student is beginning to learn an academic skill.  
● Student cannot accurately complete an academic task.                                                                                         | ● Modeling  
● Immediate response evaluation and corrective feedback.  
● Praise for effort.                                                                                           |
| Fluency Stage  | ● Student can complete a task slowly and accurately.                                                                                                                                       | ● Repeated drill and practice.  
● Feedback regarding accuracy and fluency of performance.  
● Praise for increased fluency rates.                                                                 |                                           |
| Generalization | ● Student can accurately and fluently complete an academic task.  
● Student can generalize the apply skill to new situations, but may confuse target skill with other, similar skills.                   | ● Discrimination and generalization training.  
● Praise for using skill in new settings or situations                                                                                               |
| Adaptation     | ● Student can accurately and fluent complete an academic task.  
● Student can apply the skill in novel situations without prompting.                                                                                  | ● Provide instruction on the core elements or big ideas of the target skill so that the student can access it when facing unfamiliar tasks and situations.  
● Student must practice using the target skill with some modifications in new situations and settings.  
● Praise improvement on ability to adapt the skill.                                                                                                 |
Each stage of the instructional hierarchy describes where students fall on the developmental trajectory of learning a specific academic skill (Haring et al., 1978). For example, a student who is in the acquisition stage is in the beginning of acquiring a particular academic skill, such as subtraction, and would struggle to perform any academic task that requires that particular skill set, such as calculating change. At each stage, students need different types of instructional interventions and feedback.

According to Haring et al. (1978), a student in the acquisition stage would require modeling, feedback, and praise for effort. In the fluency stage, students are accurate but are not fast; they require frequent opportunities for repetition and practice followed by targeted feedback regarding the accuracy and fluency of performance as well as praise for increased fluency (Haring et al., 1978). Students in the generalization stage of the instructional hierarchy are able to respond accurately and fluently while also applying the skill in new situations and settings. Students in this stage, however, may confuse the target skill with similar skills (e.g., confusing number operation signs); therefore, teachers need to provide students with opportunities to use the target skill regularly and for different tasks. If the student confuses the target skill with other similar skills, the teacher could provide discrimination training. Discrimination training is a metacognitive process that requires the student to ascertain when to use each specific skill. That is, the teacher may train for generalization in order to clearly illustrate ways in which the academic skill can be used in different contexts for the student (Haring et al., 1978).

Generalization training provides the student with opportunities to use newly learned skills in different contexts (Ardoin, McCall, & Klubnik, 2007). In the adaptation stage of learning, the student is fluent and accurate and can apply the skill in novel situations.
without prompting (Haring et al., 1978). It is recommended that teachers of students who are in this stage of learning focus on teaching the student to articulate the core elements of the target skill so that the student can access it when facing unfamiliar tasks and situations. At the same time, students need opportunities to practice the target skill with some modifications in new situations and settings (Haring et al., 1978). Although this model is over 30 years old, it is still considered an effective intervention heuristic (Ardoin & Daly, 2007). Ardoin and Daly (2007) note that the concept of the instructional hierarchy was developed in a way that likens the instruction of academic skills to that of behavioral shaping. Typically, behavioral researchers examine changes in the frequency of responses as a target behavior is brought under stimulus control. In the context of academics, specific academic skills are identified as target behavior and an academic skill is brought under stimulus control through instruction and intervention. Students will show an increase in the frequency of accurate responses as the academic skill is remediated. Once the academic skill is acquired, students are explicitly trained to generalize the newly acquired behavior or skill to other appropriate stimuli. The instructional hierarchy provides guidance regarding how and when to use prompts that target specific responses. It also provides guidelines regarding how to structure contingencies for accurate and inaccurate responding (Ardoin & Daly, 2007).

Ardoin and Daly (2007) note that the instructional hierarchy allows one to use behavioral principles to guide academic instruction in two fundamental ways. First, it provides a framework that asks educators to focus on students’ responses and the way these responses change as skills are developed. In this case, the rate of accurate responding increases as the skill deficit is mediated through proper instruction. Secondly,
the instructional hierarchy, with its developmental trajectory, provides a guideline on how to react to changes in the frequency of accurate student responding. If educators can learn to demonstrate appropriate reactions to the changes in student responses, stronger response repertoires that are more generalizable can be developed. This means that teachers can use these guidelines to modify instruction based on individual need. As a result, students are better able to develop the academic skills they need. Students are explicitly taught how to use skills in response to different types of stimuli; in this case, the different types of stimuli are different types of academic tasks that require the same academic skill.

The instructional hierarchy has become more relevant to current educational practice as schools are beginning to adopt RTI models. Implementation of the RTI model requires that school-based evaluation teams learn how to evaluate specific skill deficits, develop interventions to correct those skill deficits, conduct frequent measurement of the targeted skills, and evaluate student performance using single subject design methodology (Ardoin & Daly, 2007). These components of RTI clearly reflect those that are included in the instructional hierarchy that was proposed by Haring et al. (1978). Thus, it seems clear that the instruction hierarchy provides a framework that can be used within the RTI model. Although the instructional hierarchy provides a framework for matching student skills to appropriate academic interventions, it does not provide criteria or procedures for assessing where a student falls on the hierarchy. This important piece is addressed in the instructional delivery model, proposed by Gickling (1977). In this model, the accuracy with which students perform a specific academic task is important in
determining their instructional needs. This model will be presented in the following section.

**Instructional delivery model.** Gickling’s (1977) instructional delivery model is different from the instructional hierarchy put forth by Haring et al. (1978), as his model focuses more on discrete tasks and independent seatwork rather than on instructional recommendations. The instructional delivery model includes three levels: frustrational, instructional, and independent levels (see Table 2 for an overview of the levels, proposed criteria, and recommended tasks). In later research the term “independent” becomes synonymous with “mastery” and the two terms are used interchangeably throughout the literature. For the purposes of this study, we will refer to this level as the mastery level of performance. In Gickling’s model, levels of performance are determined based on the type of task, the number of items in the task, and how accurately the student can complete that task (Gickling & Thompson, 1985).

Gickling et al. (1989) discuss two types of mathematics tasks: drill and application. The type of task dictates the level of accuracy with which a student must be able to complete a task to fall in the instructional range. For example, Gickling et al. propose that an instructional level for drill tasks would be represented when a student could complete a task with 70-85% accuracy, while the instructional level for application task is represented by an accuracy rate of 85-100%. If a student performs below these ranges, the assessed task falls in the frustrational range, and if they perform above the instructional ranges the task falls in the mastery range. As it is impossible to complete a task with more than 100% accuracy, students are in the mastery range when they approach 100% accuracy on application tasks (Gickling et al., 1989). These accuracy
ranges differ because the demands of the two types of tasks require different levels of understanding in order for students to be successful.

Gickling and Thompson (1985) suggest that drill tasks represent a practice stage of acquiring new information and should not be too challenging for students. Items should contain new facts, concepts, and problems only when students fall within the instructional range (Gickling et al., 1989). Once a student has progressed to the point where rates of accuracy are above 85%, they have reached the mastery level for drill tasks. Within this model, drill activities need to be related to curriculum content and, ideally, taught within a real-world context; they should also be presented prior to content instruction (Gickling & Thompson, 1985). Because problem-solving and application problems require a conceptual understanding of the mathematics being studied, it is recommended students begin to complete these types of tasks once they have mastered the basic skills included in the drill tasks (Gickling et al., 1989).

For both types of tasks, it is important they are carried to their logical conclusion, meaning that once instruction of a skill begins, it should continue until mastery is achieved. Mastery is achieved when students are able to comprehend the task and can perform at independent levels.

Gickling and Thompson (1989) report that application of the instructional delivery model can be utilized with the application of a few basic rules. Their instructional recommendations use students’ performance to guide the level of tasks required of students rather than focusing on the nature of instructional procedures. For example, they state that the percentage of “knowns,” which is a term that refers to items that are able to be recognized and completed by the student, should remain high. This
means that when completing a task, students should know how to complete most items. They also recommend that new material should be confined to the margin of challenge: As teachers progress through curriculum, new information should not cause accuracy rates to drop below the recommended instructional ranges for the type of task being presented. If teachers are unsure of whether a student can complete an item, it should be considered unknown.

Table 2

The Instructional Delivery Model

<table>
<thead>
<tr>
<th>Stage</th>
<th>Grades</th>
<th>Original Accuracy Criteria</th>
<th>Historical Fluency Criteria</th>
<th>New Fluency Criteria</th>
<th>Suggested Interventions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frustrational Level</td>
<td>1-3</td>
<td>&lt;70% Accurate</td>
<td>&lt;10 DCPM</td>
<td>&lt;14 DCPM</td>
<td>Scaffolding and direct instruction.</td>
</tr>
<tr>
<td></td>
<td>4-5</td>
<td></td>
<td>&lt;20 DCPM</td>
<td>&lt;24 DCPM</td>
<td></td>
</tr>
<tr>
<td>Fluency Level</td>
<td>1-3</td>
<td>70-85% Accurate</td>
<td>10-19 DCPM</td>
<td>14-31 DCPM</td>
<td>Drill Tasks</td>
</tr>
<tr>
<td></td>
<td>4-5</td>
<td></td>
<td>20-39 DCPM</td>
<td>24-49 DCPM</td>
<td></td>
</tr>
<tr>
<td>Mastery Level</td>
<td>1-3</td>
<td>&gt;85% Accurate</td>
<td>&gt;19 DCPM</td>
<td>&gt;31 DCPM</td>
<td>Problem-solving and conceptual tasks.</td>
</tr>
<tr>
<td></td>
<td>4-5</td>
<td></td>
<td>&gt;39 DCPM</td>
<td>&gt;49 DCPM</td>
<td></td>
</tr>
</tbody>
</table>

Note: a developed by Gickling (1977); b developed by Deno and Mirkin, 1977; c developed by Burns et al. (2006); DCPM = Digits Correct per Minute.

Gickling and Armstrong (1978) show support for the use of the instructional delivery model when they found that when students were completing independent tasks at their instructional level, improvements were seen with regard to comprehension, on-task behavior, and task-completion as compared to when students were asked to complete
tasks at their frustrational or mastery levels. Students who were asked to perform at their frustrational level tended to avoid the task, while students who were completing tasks at the mastery level finished assignments quickly and engaged in high levels of off-task behavior. As these results suggest, this study uses task difficulty and independent work as the criteria for determining instructional interventions, as opposed to specific instructional recommendations.

_The relationship between the two models._ Although there are substantial differences between the instructional hierarchy and the instructional delivery model, some parallels can be drawn. For example, both suggest that academic skills are hierarchical in nature in that students must master basic skills before attempting to apply them to more difficult tasks. Specifically, Haring et al. (1978) suggest that students must be able to demonstrate the ability to accurately and fluently demonstrate knowledge of academic skills before attempting to generalize them to other situations or apply them to problem-solving tasks. Similarly, Gickling (1977) recommends that students display high levels of accuracy when demonstrating academic skills before attempting to apply those skills to problem-solving tasks.

The instructional delivery model and the instructional hierarchy continue to guide research, as the two models are frequently used in combination to determine the instructional needs of students (e.g., Burns et al., 2006; Codding et al., 2007; Codding et al., 2010; Rhymer et al., 2000; Rivera & Bryant, 1992; Vanderheyden & Burns, 2005), as the levels of the instructional delivery model are applied to the instructional hierarchy to identify appropriate interventions for students. Although the instructional delivery model (Gickling, 1977) largely focuses on independent work, researchers have inferred that
explicit instruction and scaffolding are necessary when students are not performing at the instructional level for a particular task (Codding et al., 2007). Researchers hypothesize that if students are performing a task in the frustrational range, they are likely to require an intervention associated with the acquisition stage of the instructional hierarchy in order to learn the academic skills associated with that task (e.g., Burns et al., 2006; Codding et al., 2007; Codding et al., 2010; Rhymer, et al., 2000; Rivera & Bryant, 1992; Vanderheyden & Burns, 2005). Similarly, a student who is asked to complete a task in the instructional range might require an intervention associated with the fluency stage of the instructional hierarchy, and a student who is completing applications tasks in the mastery range might benefit from interventions described in the generalization and application stages of the instructional hierarchy.

Although the instructional delivery model provided an initial framework for ensuring instructional interventions are matched to student skill levels, there has been much debate about the particular criteria that should be used to do this over the past three decades (Burns et al., 2006). Some researchers have proposed that accuracy criteria be used to determine instructional level (e.g., Gickling & Thompson, 1985; Neef et al., 1980; Robinson & Skinner, 2002), while others have advocated for the use of fluency scores instead (Burns et al., 2006; Daly & Martens, 1994; Deno & Mirkin, 1977; Vanderheyden & Burns, 2005). The following sections will discuss the literature regarding these two approaches and show that fluency criteria are a better indicator of student performance. Thus, Gickling’s (1977) model has been modified, as fluency criteria are now used to determine instructional levels. However, his concepts are still used to assess student performance and identify instructional needs.
Accuracy criteria as an expectation for student performance. Much of the research regarding accuracy criteria has been based upon drill activities (e.g., Gickling & Thompson, 1985; Neef et al., 1980; Robinson & Skinner, 2002). Burns (2004) cites several studies that have shown that teaching academic skills through drill tasks has led to increases in student performance with regard to more advanced skills, such as advanced computational skills and mathematical reasoning (e.g., Dehaene & Akhavein, 1995; Jones & Christensen, 1999; Tzelgove, Porat, & Henik, 1997; as cited by Burns, 2004) and hypothesizes that research has focused primarily on drill activities because they are easy to assess and because increased opportunities for practice are inherent to student learning. As previously noted, Gickling and Thompson (1985) suggest that students should be able to complete 70-85% of a given academic task in order to be in the instructional range. When considering these criteria, it is important to remember that Gickling (1977) proposes that tasks falling in the instructional range are at the optimal difficulty level to promote student learning because students show higher levels task completion, on-task behavior, and comprehension when completing tasks that fall within this range.

Neef et al. (1980) conducted a study that suggests a lower level of accuracy is required for student success. Their study demonstrated that students progressed through academic tasks in which they could accurately complete only 50% of items in a given task. Similarly, Robinson and Skinner (2002) found that students were able to show progress on tasks in which less than 50% of the items were considered to be known to the participants, but found that results differed depending on the type of mathematics tasks students were completing. In this study, improvement was seen on a timed mental math task, but not on an untimed, written task. Cook, Guzaukas, Pressley, and Kerr (1993)
state that proposed ranges of the ratios of known items to new items for optimal student learning range from 10% to 50% in the literature. They suggest that universal accuracy criteria may not be the best approach, and instead they recommend that educators meet the individual needs of students by determining optimal ratios for each learner within a particular subject area or task. While it is always beneficial to meet the needs of individual students, creating individualized criteria for each individual would not result in consistent decision-making procedures across students. Based on these studies, the field does not have a reliable gauge for determining the ratio of known to unknown items for most learners.

In 2004, Burns conducted a meta-analysis comparing effect sizes of studies that used the percentage of correct items on a given task to identify student ability levels in an attempt to find an optimal ratio of known versus unknown items. Burns wanted to see what percentage of known items led to the greatest effect sizes. He compared 17 studies that used percentages of <50% known, 50%-69% known, 70%-85% known, and 90% known as groupings for identifying the instructional level of students. Burns found that all groups resulted in large effect sizes, with the exception of the <50% known grouping, which yielded an effect size that represented a small to medium effect. There was no discernable difference between any of the other groups, which indicated that no one proposed range is better than another based on this study. The only conclusion that could be drawn was that it is necessary to ensure that students can complete at least 50% of a task in order to ensure success.

It is evident that the research regarding accuracy ratios is mixed. Research has clearly demonstrated that no particular criteria ensures successful student learning. The
next section discusses the fluency approach to identifying the instructional needs of students and provides evidence to suggest that it may be a better indicator when attempting to match instruction to skill level.

**Fluency criteria as an expectation for student performance.** As previously noted, the instructional hierarchy highlights fluency of academic skills as an essential component for student learning (Haring et al., 1978). The second stage of the model is called the fluency stage and it requires that students are able to quickly and accurately complete academic tasks. Additionally, Gersten, Beckmann, et al. (2009) specifically note that most effective mathematics interventions within the RTI model include practice and instruction geared towards building fluent retrieval of basic arithmetic facts. In more general terms, it has also been noted that fluent computation is a goal for mathematics instruction (National Council of Teachers of Mathematics, 2000). Additionally, a goal within the new *Common Core Standards* specifically states that students who are in third grade should be able to fluently multiply and divide within 100 (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010). Based on these goals, fluent rates of accurate responding are an inherent part of mathematics instruction. As a result, it would be logical to use fluency scores to determine skill level as this measure coincides with academic goals within the frameworks of the instructional hierarchy, the Common Core, and the implementation of RTI.

Although accuracy scores provide information regarding student ability to complete a task, fluency scores provide a more comprehensive view of student ability as they provide information regarding both accuracy and speed of performance (Burns et al.,
2006). More specifically, accuracy is assessed to some degree when fluency scores are used, while speed of performance is not assessed when using accuracy scores. Furthermore, research has supported the use of fluency criteria over accuracy criteria for identifying the instructional level of students (Burns et al., 2006). When the two foci of assessment were compared, fluency data were found to be more reliable and they correlated better with criterion measures than did accuracy data. Therefore, it is important to understand and consider the role of fluency in assessing student performance. The following sections highlight research that has focused on fluency criteria for identification of students’ instructional levels.

**Historical fluency criteria.** Deno and Mirkin (1977) proposed a fluency approach in which both speed and accuracy were assessed to determine students’ place within the instructional delivery model. They developed specific criteria that suggested a certain number of digits correct per minute (DCPM) would determine the instructional level of each student. Specifically, they proposed that students in first through third grades fell in the instructional range if they were able to complete between 10-19 DCPM. Students in the fourth and fifth grades fell in the instructional range if they were able to complete between 20-39 DCPM. Students who yielded DCPM scores that fell below these ranges fell in the frustrational range, while students who yielded DCPM scores that fell above these ranges fell in the mastery range.

Over the years, studies have utilized Deno and Merkin’s (1977) fluency criteria for identifying a student’s instructional level. Using fluency data as a diagnostic tool for instructional interventions has been found to improve student outcomes (e.g., Burns et al., 2010; Codding et al., 2007; Vanderheyden & Burns, 2005). Codding et al. (2007) used
the fluency approach developed by Deno and Mirkin (1977) to identify students who
were performing in the frustrational and instructional ranges and compared the
effectiveness of different interventions (i.e., no intervention, acquisition intervention, and
fluency intervention) for students in different instructional ranges. Acquisition
interventions are those that take place when a student is in the first stage of the
instructional hierarchy and they include modeling, feedback, and praise for effort (Haring
et al., 1978). Fluency interventions are those that take place in the second stage of the
instructional hierarchy and they seek to increase speed with which items are completed
through opportunities to engage in repeated practice and/or performance contingencies
(Burns et al., 2010). These types of interventions be discussed at length in future
sections. Codding et al. found that the fluency intervention was most effective for
students who were performing in the instructional range and it was least effective for
students who fell in the frustrational range. They also found that there was no significant
difference between the three groups until the initial fluency scores were considered. The
results of this study show that the use of the fluency criteria may be able to guide
instructional decision making. Once the instructional level of students was considered,
Codding et al. were able to demonstrate that students who fell in the instructional range
benefited most from the intervention that was designed to best meet the needs of students
who were performing within that range.

Vanderheyden and Burns (2005) also used Deno and Mirkin’s (1978) fluency
criteria to match instruction with student need as part of a school-wide mathematics
intervention. They conducted a yearlong study that targeted fluency building within
basic computational skills with students in the third through fifth grades. They used
Curriculum Based Assessment (CBA) and Curriculum Based Measurement (CBM) procedures throughout the year in order to guide interventions for students. Results indicated that the use of these procedures resulted in substantial growth with regard to fluency scores and led to significantly improved scores on a standardized norm-referenced group test. The researchers noted that the use of CBA allowed the school to identify academic deficits across grade level and provided data for sound instructional decision making.

Shapiro (2011) notes that there are no clear criteria for setting frustrational, instructional, or mastery levels in math. He states that there are several ways to approach assessment in that local or national norms can be used. Shapiro cited the criteria put forth by Deno and Mirkin (1977) and noted that it was considered to be historical in nature. He then compared the historical criteria to two local norms for levels of fluency that were differentiated by socio-economic status. In one district with a moderate socio-economic base, second and third grade students who completed 10 DCPM and fourth and fifth grade students who completed 13 DCPM fell in the instructional range. However, in a district with a high socio-economic base, third graders who completed 10 DCPM, fourth graders who completed 17 DCPM, and fifth graders who completed 21 DCPM fell in the instructional range. The variability in criteria put forth by Shapiro indicates that there are no universally accepted criteria in the literature. Shapiro notes that the suggested criteria can best be described as “valuable best guesses” (p. 157) due to the lack of consensus.

Although both fluency and accuracy criteria have been used by researchers and practitioners to guide their research and interventions, they are not based on empirical evidence. There are no data with regard to how the fluency criteria (Deno & Mirkin,
1977) or the accuracy criteria (Gickling & Thompson, 1985) were derived (Burns et al., 2006). Additionally, it was reported that very little data exist to validate the use of either set of criteria. While there is little psychometric data with regard to the two approaches, Burns et al. note that, given that fluent computation is typically a goal for mathematics instruction, choosing fluency over accuracy measures is a logical choice. This notion was reinforced in an earlier section when we considered that fluent computation highlighted in the instructional hierarchy, best practices for RTI, and the Common Core Curriculum. The next section details an attempt by Burns et al. (2006) to empirically derive fluency criteria that can be used to identify student skill levels and subsequently match instruction to needs.

**New fluency criteria.** Burns et al. (2006) conducted a study that sought to provide additional information regarding the use of the accuracy and fluency approaches. First, they attempted to determine if either the accuracy or the fluency approach led to stronger alternate-form reliability coefficients. They found that fourth- and fifth-grade students demonstrated higher reliability coefficients when compared to second and third graders. They also found that aggregated fluency data were more reliable than accuracy data. Next, they attempted to determine which approach correlated better with scores from a commercially available, standardized mathematics test. They found that fluency scores yielded higher correlations when compared with student performance on commercial tests, than did accuracy scores. Finally, they wanted to see if data regarding student growth that were obtained during a protocol-based math intervention could be used to suggest potential criteria for instructional level. They found that only the categorical fluency data (Deno & Mirkin, 1977) appeared to be adequately reliable to
make screening and instructional decisions for individual students in fourth and fifth grades. Scores from second and third grade, however, were not strong enough to be confidently used to make decisions about instructional levels for individual students. Therefore, we can say that, while not necessarily applicable to all grade levels, fluency data are the more reliable tool for making instructional decisions.

As part of this study, Burns et al. (2006) empirically derived new fluency criteria for determining the instructional levels of students in grades 2-5. The new criteria proposed by Burns et al. (2006) utilize higher scores to determine whether students are placed in an instructional or mastery level than those used by Deno and Mirkin (1977). Burns et al. found that students in second through third grades were performing in the instructional range if they completed between 14-31 DCPM, while students in fourth and fifth grades were performing at the instructional range if they were able to complete 24-49 DCPM. This greatly impacted the level of intervention students received.

In order to create the new criteria, Burns et al. (2006) computed the mean and one SD range of the fluency score from the initial mathematics single-skill probes for children whose growth slopes met or exceeded the 66th percentile. An instructional level criterion was suggested for a second- and third-grade group and a fourth- and fifth-grade group by computing the starting fluency score that fell 1 SD above and below the mean for each group. Fluency scores that fell below this new range fell in the frustrational range, while scores that fell above this range were within the mastery or independent level. Next, researchers sorted students into groups based on scores from a mixed skill probe according to the new criteria. They compared these groups to those formed by the Deno and Mirkin (1977) criteria using a Wilcoxon signed rank test and found significant
changes in classifications. The ramifications of these changes are that students had to achieve higher levels of fluency before progressing to the next stage of learning.

Burns et al. (2006) also conducted several reliability and validity checks. Delayed alternate-form reliability was estimated by using Kendall’s tau to correlate the categories from the two probes and categorical data. Correlations were also generated using Spearman’s rho with the continuous data from the SAT-9 mathematics standard scores to estimate criterion-related validity. The reliability coefficients were .51 for the total sample of 434 students and .35 and .63 for the two grade groups. Percent agreement was 69.0% for the second- and third-grade group and 92.4% for the fourth- and fifth-grade group. Validity coefficients were .08 and .49 for the grade groups and .27 for the total sample. Thus, the coefficients were more promising for the fourth- and fifth-grade group.

Next, Burns et al. (2006) used the mixed probe to classify children in the frustrational, instructional, or mastery ranges based on the new criteria. Growth rates were calculated to determine if students in the instructional range showed greater improvement than students in the other categories. They did this to evaluate the validity of the new criteria by applying it to a different set of data (mixed skilled probes versus single skill probes from which the criteria were derived). The researchers found that for the second- and third-grade group, the average growth per week was 1.77 DCPM per week for the frustrational range, 2.01 DCPM per week for the instructional range, and 1.55 DCPM per week for mastery. Students in the fourth- and fifth-grade group showed average growth per week of 1.16 DCPM in the frustrational range, 1.44 DCPM for the instructional range, and 1.25 DCPM for the mastery range. Although a one-way analysis of variance did not reveal significant effects, the Cohen’s $d$ indicated a small average
effect of .22 between instructional and frustrational and .24 between instructional and mastery. Researchers conducted the same analyses using the historical ranges proposed by Deno and Mirkin (1977) and found similar effect sizes. One notable difference was that a larger effect size was seen between the frustrational and mastery ranges for the historical criteria than was seen between the two ranges using the new criteria.

Burns et al. (2006) note that the new criteria yielded similar psychometric properties to the historical criteria. They also note, however, that there are several limitations to this study. First, the data were collected within one school district and it is not known the degree to which these scores are applicable to students in other districts. They also report that their fluency scores data were slightly skewed and leptokurtic which may increase the risk for a Type 1 error. The researchers also noted that there were some effect sizes between grades within each group. They derived their criteria based on groupings to match them with the historical criteria, but recommended that further research be done to determine if grade-specific criteria are more appropriate. Burns et al. noted that the main advantage of their data is that they are empirically derived. As psychometric properties were similar to the historical criteria put forth by Deno and Mirkin (1977), it may be advantageous to use the new criteria moving forward.

In sum, there is much disagreement in the literature regarding how to identify student skill levels in order to determine instructional need. Although research has supported the use of accuracy data (Gickling & Thompson, 1985; Neef et al., 1980; Robinson & Skinner, 2002), more recent research suggests that the use of fluency data is more reliable and valid (Burns et al., 2006). Additionally, as previously reported, fluent computation is inherent to the instructional hierarchy (Haring et al., 1978) and a specific
goal of RTI (Gersten, Beckmann, et al., 2009) and the Common Core Standards (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010). Thus, the use of fluency criteria is more closely linked to current academic goals. Although neither the historical fluency criteria (Deno & Mirkin, 1977) nor the new fluency criteria (Burns et al., 2006) provide strong psychometric properties, the new criteria provide information regarding their empirical origin. Additionally, because the new criteria rely on higher cut-off scores for determining students’ levels, they set a higher standard for students. By setting higher standards for students it helps ensure that students are receiving more intensive interventions for a longer period of time, as they have to achieve higher levels of performance before attaining the next level of instruction. This is preferred over providing students with less intervention when they need it. Thus, for the purposes of this study, the new fluency criteria were used for identification of the instructional levels of students.

The goal of instruction should be student learning. Yet determining just what students need is not straightforward. Therefore, frameworks for matching instructional interventions with students’ needs are crucial. Two models have been developed within the field of mathematics instruction and recent findings suggest that using fluency criteria for diagnosing student level in order to target instruction and choose those tasks that will best help students along their developmental trajectory.

**Curriculum Based Assessment: Procedures for Assessing Skill Level**

The instructional delivery model was initially conceived as a way of conceptualizing the newly developed method of assessment that became known as CBA (Gickling, 1977). Gickling and Thompson (1985) defined CBA as a procedure for
determining the instructional needs of students based on ongoing student performance and existing curriculum. They noted that CBA allows teachers to focus on an individual’s skills relative to the curriculum. Once teachers have this information it becomes easier to alter the difficulty level of assigned tasks according to the student’s academic ability. According to Gravois and Gickling (2002), CBA allows us to make evidence-based decisions when selecting interventions for students.

Although current mathematics curriculum focuses instruction on concepts and applications and has de-emphasized the direct instruction of computational competency, CBA continues to focus on computational objectives because this is the foundation upon which success in other aspects of mathematics is built (Shapiro, 2011). This idea coincides with those put forth in the instructional hierarchy and instructional delivery model, where it is inherent that students master basic skills before moving on to more complex tasks (Gickling, 1978; Haring et al., 1978).

CBA involves sampling student behavior within the instructional materials and comparing student skill and task demand in a standardized way (Gravois & Gickling, 2002). For example, when a student is working on a single skill such as multiplication, the target behaviors would be answering correctly. The task would be a worksheet with single digit multiplication problems. The frequency of accurate responding would be assessed, which is also known as the rate or speed at which a student can accurately complete single digit multiplication problems. Additionally, one might consider error rates by assessing the number of mistakes the student makes while completing the assessment probe. Further, it may be important to look at patterns of responding. The types of errors a student makes may help guide specific areas to target for intervention.
As previously noted, if a student is completing this task with a fluency rate that falls in the instructional range, then a fluency intervention would be most applicable. A student who is responding in the frustrational range may respond better to an acquisition intervention (Burns et al., 2006; Burns et al., 2010; Codding et al., 2007; Deno & Mirkin, 1977).

**Subtypes of CBA.** Fuchs and Deno (1991) classify all forms of CBA as general outcome measures or specific sub-skill mastery models. General outcome measure models use grade-level, standardized measures that have established reliability and validity; they are curriculum independent and have been shown to be related to overall student achievement (Fuchs & Deno, 1991). Measures are samples of performance that are brief and timed and use fluency rates as scores. The purpose of these types of models is to measure long-term student progress across a wide range of skills. While this type of information may tell us when an instructional change needs to happen, it does not provide specific information regarding what types of changes should be made (Shapiro, 2011).

For specific sub skills, criterion-referenced measures are used and they are typically based on the development of a skills hierarchy (Shapiro, 2011). These measures are used to determine if students are meeting the short-term instructional objectives of a specific curriculum. Moreover, they may provide information regarding the types of instructional modifications that may be necessary.

The measures used in the specific sub-skill mastery models sometimes have a direct relationship with the long-term goals of a specific curriculum, although this is not required (Shapiro, 2011). These types of models require a different measurement procedure with the teaching of each new academic skill. Thus, measures are not
standardized from one academic skill set to the next. Often, these probes are created by teachers and scoring procedures can vary widely and have included accuracy ratios, fluency rates, and analysis of error patterns.

**Curriculum Based Measurement.** Curriculum based measurement (CBM) is a type of CBA meets all the requirements of CBA (e.g., measure is related to curriculum, assessments are conducted frequently, data is used to inform instructional decisions) and it also has two additional properties (Fuchs & Fuchs, n. d.). First, each CBM probe assesses performance on identical, or parallel, measures over time. This means that each probe samples the year-long curriculum in exactly the same way using prescriptive methods for creating the probes. Fuchs and Fuchs (n. d.) further note that CBM is usually conducted with universal tests that have been designed to assess common curricula. This differs greatly from CBA which requires teachers to design their own assessment procedures.

The second defining feature of CBM is that it is standardized, which enhances reliable and valid scores (Fuchs & Fuchs, n. d.). CBM provides a standardized set of materials that has been developed by researchers to produce meaningful and accurate information. This is significant because the adequacy of teacher-developed and commercially produced CBA probes is relatively unknown. There is no research regarding the reliability or validity of CBA procedures, while CBM has been widely researched. Shapiro (2011) reports that CBM has become known as a highly reliable and valid method of conducting academic assessments. For example, the test-retest reliability of CBM over a one-week interval has been found to be high ($r = .93$) and parallel forms of reliability have been reported to be moderate to high ($.48 < r < .72$) (Shinn, 1989). Additionally, the criterion validity of mathematics probes has also been examined by comparing the CBM mathematics probes with the Operations Subtest of the Metropolitan Achievement Test and Problem-Solving Subtest. Mean correlations were $.54$, and $.42$, respectively (Shinn, 1989).
Although there are several universally available CBM assessments (e.g. DIBELS, AIMSweb), researchers have also used standardized procedures to create CBM measures for specific skills in order to assess the effectiveness of academic interventions (e.g. Codding et al., 2006; Codding et al., 2007; Shapiro, 2011). Shapiro (2011) provides several websites that provide templates for creating CBM probes along with specified administration and scoring procedures. The proposed study will use the procedures described by Shapiro and modified by Codding et al. (2006) to create single-skill probes. These procedures are detailed in Chapter 3.

**Assessment procedures.** Shapiro (2011) suggests that CBA can be used in order to determine the instructional level of a student. Because CBM procedures have been found to have better psychometric properties than other CBA approaches, CBM probes will be used in the proposed study. As noted above, specific sub-skill mastery models are often used to place a student in a particular instructional level in the instructional delivery model. In order to do this effectively, the evaluator should first determine the specific types of math problems that are in need of assessment and intervention. This can be achieved by obtaining a sequence of instruction for computational skills within the district. Next, the evaluator works with the teacher to determine the specific skill areas that may fall in the acquisition or fluency phases on the hierarchy. In other words, the evaluator should ask the teacher what skills the students are able to complete within those two ranges (Shapiro, 2011). This can help evaluators to narrow the scope of skills to be assessed.

Next, it is important to determine which type of CBM probe to use. Shapiro (2011) reports that single-skill probe sheets are useful for providing specific recommendations regarding skill deficits and strengths. These are typically used when evaluators are attempting to determine the instructional placement of a student. Additionally, these types of sheets can be used to monitor student growth with regard to
newly taught skills as instruction is being modified (Shapiro, 2011). Shapiro notes that multiple-skills probes are more useful when the evaluator wishes to assess a broader range of skills at one time. He stated that these are typically used when conducting progress monitoring of student performance or for screening purposes.

For the purposes of this study, single-skill probes were used to assess student progress throughout the intervention, while multiple-skills probes were used to test for generalization and were administered before and after the intervention. Furthermore, Shapiro’s recommendations for practice were utilized, as the PI was familiar with the grade-level curriculum and also conferred with the classroom teacher in order to gain a sense of student skill levels. As we now have assessment procedures and instructional frameworks to guide the current study, it is now necessary to further explore the aspects of acquisition and fluency interventions. The following sections will highlight key features of these interventions.

**Academic Interventions**

Fluency criteria can be applied to both the instructional hierarchy and the instructional delivery model to assess students’ levels with the goals of targeting instruction and choosing tasks that will best help students along their developmental trajectory. As previously noted, the instructional hierarchy includes four stages: acquisition, fluency, generalization, and adaptation (Haring et al., 1978) and provides suggestions for interventions at each stage. The instructional delivery model (Gickling, 1977) has been utilized as a framework for determining the stage of the instructional hierarchy at which a student is performing. For example, students performing in the frustrational and fluency levels of the instructional delivery model would require the
academic interventions described in the acquisition and fluency stages of the instructional hierarchy, respectively (Codding et al., 2007). This paper will concentrate on the first two levels of both models (i.e., the acquisition and instructional stages, and frustrational and fluency levels), as the goal of this study was to understand how to improve rates of fluent responding for basic math facts (e.g., single digit subtraction, addition, multiplication, or division). According to Haring et al. (1978), students in the acquisition stage require modeling, opportunities for practice, feedback, and praise for effort. Interventions that contain these components are known as acquisition interventions.

Likewise, students are in the fluency stage require frequent opportunities for repetition and practice, followed by targeted feedback regarding the accuracy and fluency of performance, as well as praise for increased fluency. Interventions that contain these features are known as fluency interventions. The following sections will review the research associated with various types of interventions that fall within the acquisition and fluency stages of instruction.

**Acquisition interventions.** Acquisition interventions are those that take place when a student is in the first stage of the instructional hierarchy (Haring et al., 1978). In this stage, students struggle to perform the targeted skill and they benefit from practice, feedback, modeling, and praise for effort (Haring et al., 1978). In terms of mathematics and for the purposes of this study, this means that students in this stage are struggling to complete basic computational problems. Through the lens of the instructional delivery model, students performing in the frustrational range are unable to complete a newly presented mathematical operation with a high level of accuracy because they lack the specific academic skills required to complete the mathematical operation (Gickling,
1977). Like those students at the acquisition stage of the instructional hierarchy, these students are likely to require additional support when completing a new operation in order for the activity to be a positive learning experience. Thus, many researchers have utilized acquisition interventions to improve accuracy and fluency rates when student performance falls in the frustrational range (Burns et al., 2010; Codding et al., 2007; 2011; Codding, et al., 2009). We can therefore say that the principles put forth in the instructional delivery model have been used to identify students who would benefit from acquisition interventions described in the instructional hierarchy.

Acquisition interventions are important because they lay the foundation for students’ future work developing fluency. Indeed, students need to be able to respond to math facts accurately prior to being able to develop fluency with those same math facts. Feedback, modeling, and praise for effort are the three prominent features of these interventions. Within the context of acquisition interventions, feedback includes immediate response evaluation and error-correction. This means that individual student responses are evaluated and errors are immediately corrected. This is not done following the completion of a whole task or assessment, but instead takes place following the completion of each individual item within the task. This type of feedback is provided to students until they are able to respond accurately. The purpose is to ensure that students are not engaging in high rates of incorrect responding (Skinner & Smith, 1992). Essentially, frequent feedback ensures that students are not practicing errors (Rivera & Bryant, 1992). Once feedback is provided regarding an incorrect response, it is necessary for students to have the skill modeled, or demonstrated appropriately (Haring et al., 1978). In its simplest form, students are shown or told the math fact and the correct
answer. Haring et al. (1978) also note that acquisition interventions should include a component of praise that is specific to student effort. Behavior-specific praise has been found to increase time on-task, responding, and correct answers (Duchaine, Jolivette, & Fredrick, 2011). When students are unable to present with high rates of accurate responding, it is necessary to focus on the effort put into acquiring the new academic skill (Haring et al., 1978).

Types of acquisition interventions. Over the past four decades, researchers have explored various types of acquisition interventions. Several studies provide support for the use of incremental rehearsal (e.g., Burns, 2005; Codding et al., 2010), self-instruction (e.g., Meichenbaum & Goodman, 1971; Wood et al., 1993), cover-copy-compare (Codding et al., 2006; Codding et al., 2007; Codding et al., 2009; Skinner et al., 1997), and error-correction procedures (e.g., Ferkis et al., 1997; Marvin et al., 2010; Rapp et al., 2012; Sulzer-Azaroff & Mayer, 1986; Worsdell et al., 2005). In a review of the literature synthesizing the research of simple and moderate interventions in mathematics, it was found that the use of incremental rehearsal, self-instruction, and cover-copy-compare interventions yielded large effect sizes with regard to improvements in fluency rates (Codding et al., 2009). Error-correction procedures were not included in this literature review, possibly because there have been few studies that have utilized these procedures in the area of mathematics. While each of these procedures has unique characteristics, they all have similar attributes that provide opportunities for modeling, practice, and immediate feedback.

Although acquisition interventions have been shown to be effective for improving basic math skills, Codding et al. (2009) note that there are several gaps in the literature.
First, they report that when considering students’ individual computational skills, only self-instruction meets the evidence-based criteria presented by Horner, Carr, Halle, McGee, Odom, and Woolery (2005). Although this literature review did not include error-correction interventions, it is clear that these types of procedures would not meet the standards put forth by Horner et al. because there are only two studies that utilize this type of intervention in the area of mathematics. Horner et al. (2005) state that in order for a mathematical intervention to meet the standard to be considered evidence-based practice, several criteria must be met: (a) there must be at least five single-case studies that meet minimally acceptable methodological criteria with documented experimental control found in peer-reviewed journals; (b) of the five studies, at least three must be carried out by different researchers across three different locations; and (c) there must be a cumulative sum of at least 20 participants across the five studies. Codding et al. note that if these criteria are applied, there remains a significant need for additional research, as there are few interventions in the area of mathematics that can be considered to be evidenced based. Additionally there is a lack of research regarding generalization effects and treatment integrity (Codding et al., 2009). Lastly, they have called for more information regarding the dosage of interventions. That is, how much time and at what frequency must the intervention be implemented for a student to show improvement.

Given the existing gaps in literature, further research is needed in this area. The current study sought to expand the literature base regarding acquisition interventions by focusing on error-correction procedures, whose practices have been found to be effective, but whose use has been limited in the area of mathematics. A specific error-correction procedure, known as positive practice overcorrection or *correct response repetition*, was
used as part of the current study because it has been found to be effective in increasing accuracy and fluency scores for students in the area of reading (e.g., Ferkis et al., 1997; Marvin et al., 2010; Worsdell et al., 2005); however, research investigating the use of these types of procedures in the area of mathematics has been limited to only two studies (Rapp et al., 2012; Rhymer et al., 2000). The following sections detail how error-correction procedures include the components (e.g., feedback, modeling, and praise) put forth in the acquisition stage of the instructional hierarchy (Haring et al., 1978) to meet student needs.

*Error-correction procedures.* Error-correction procedures involve an instructor prompting a learner to elicit a correct response following the occurrence of an error (Worsdell et al., 2005) through punishing errors, providing multiple opportunities to practice accurate responses, and ensuring that the last response made was accurate (Sulzer-Azaroff & Mayer, 1986). These procedures require that students practice a specific skill. When an error is made, the instructor immediately provides feedback to the student and, when necessary, models the correct response. The student is then given more opportunities to practice the modeled response. Praise is often provided following correct responses (Worsdell et al., 2005). As previously noted, within the acquisition stage, students experience modeling, have opportunities for practice, and receive immediate feedback as well as praise (Haring et al., 1978). It should be evident that error-correction procedures are acquisition interventions as they obviously include these essential components.

Response repetition (RR) is a specific type of error-correction procedure in which students are typically required to repeat the correct response, a related response, or an
irrelevant response three to five times if an incorrect answer is given (Rapp et al., 2012); however, it has also been reported that RR has been used in studies where only single responses were required (Worsdell et al., 2005).

When implementing a “correct RR” procedure, students are required to repeat the correct response following an error. A correct RR procedure may sometimes be referred to as positive practice overcorrection if students are required to repeat the correct response multiple times. Positive practice overcorrection is an explicit approach to error-correction that has been shown to increase the probability that a student will provide accurate responses (Rhymer et al., 2002). It requires students to repeatedly emit the correct response following an error. Unlike some RR procedures, positive practice overcorrection always requires multiple repetitions and it never requires students to repeat a related or irrelevant response.

Another type of RR procedure that has been found to be effective is known as a “related RR” procedure. In a related RR procedure, a similar response is repeated following the occurrence of an error, rather than a correct response. For example, a student may be working to acquire a specific skill, such as subtraction with borrowing. When an error is made on an item, the student is guided to use the same skill on a different task. Unlike correct RR procedures, related RR procedures require students to repeat and practice a specific concept or skill rather than memorize answers to specific problems (Rapp et al., 2012).

The last type of RR procedure is the “irrelevant RR” procedure in which students are required to provide a completely irrelevant response following an error. The response
has nothing to do with the correct answer, but is used solely as a punishment for the error that has occurred.

Response repetition and mathematics. To date, there are two studies that have investigated the impact of error-correction procedures on improving rates of accurate responding to basic math facts. Rapp et al. (2012) demonstrated that correct RR and related RR procedures yielded improvements of varying degrees across four participants. More specifically, consistent improvements were seen in three of the four students, while the fourth student demonstrated inconsistent progress. In this study, modeling occurred in two different ways. In the correct RR procedure, students were given a visual model of the problem, while in the related RR procedure the student followed the verbal model of the instructor and was required to repeat the strategy on a new task. The instructor provided immediate feedback after each item in both RR procedures.

This study contributes to the literature in several ways. First, Rapp et al. used a multiple-baseline single-case design in an attempt to address the need highlighted by Horner et al. (2005) for more research in this format. Second, they demonstrated that RR procedures could be used to improve fluency scores across a variety of mathematical operations (e.g., addition, subtraction, and multiplication), as intervention took place in all three areas throughout the course of the study. Rapp et al. also presented math facts on flashcards and worksheets, which indicates that RR procedures can be implemented using a variety of materials. These findings indicate that different types of RR procedures can be used to improve fluency and accuracy for basic math skills.

Rhymer et al. (2000) utilized positive practice overcorrection procedures, also known as correct RR procedures, as part of a multi-component treatment for increasing
mathematics fluency in fourth-grade students. Their study included a peer tutoring procedure in which students were required to copy a problem with the correct answer three times if they answered incorrectly when presented with a flashcard by another student. A single subject, alternating treatments design with four students was utilized. Results of the study indicated that three of the four students displayed increased fluency rates for single-digit multiplication as a result of a treatment package that utilized this procedure in conjunction with additional treatment components that will be discussed in detail later in this paper.

Rapp et al. (2012) reported that Rhymer et al.’s (2000) study appears to be the only prior study that ever utilized a positive practice overcorrection method in an attempt to improve mathematical skills. Although there is relatively little evidence for the use of error-correction procedures in the area of mathematics, Rapp et al. reported that several studies have validated this type of procedure for increasing sight-word reading by students (Ferkis et al., 1997; Marvin et al., 2010; Worsdell et al., 2005). Because there is a lack of research in the area of mathematics, the following sections highlight the application of these procedures within the academic discipline of reading.

Response repetition and reading. Marvin et al. (2010) conducted a study in which students’ accuracy when reading sight words increased in response to the implementation of a correct RR procedure. In this study, students who read a word incorrectly were required to repeat the word five times while looking at the word. Marvin et al. also assessed for generalization and found that correct responding generalized to untrained words and contexts for the participants. This finding indicates that these types of procedures may lead to generalization of skills, which provides further
support for the notion that teaching academic skills through drill tasks leads to increases in student performance in other related skills (Burns, 2004).

*Single- verses multiple-response repetition procedures.* Thus far, this review has focused on multiple-response (MR) repetition procedures, which are more closely aligned with positive practice overcorrection procedures. However, there have been some studies in the field of reading that have also sought to determine if MR repetition procedures are more beneficial than single-response (SR) repetition procedures (Ferkis et al., 1997; Worsdell et al., 2005). Ferkis et al. (1997) compared an SR response condition in which students were required to repeat the correct word once following an error to an MR response condition in which students were required to repeat the correct word five times following an error, and no difference was found.

In contrast, Worsdell et al. (2005) conducted multiple studies and found that increased opportunities for error-correction produce better results. In their first study, students were required to engage in one or five response repetitions for each incorrectly read word. In the second study, response repetition was required following every incorrect response or every third incorrect response. In both conditions, increased opportunity for practice was more effective in increasing word-reading skills. Given the direct contrast in findings, more research is needed to determine the ideal number of repetitions for these procedures.

*Error-correction as a punishment.* One assumption regarding the practice of RR and positive practice overcorrection is that performance is improved due to the increased opportunities for a correct response to occur when a student is presented with the relevant stimulus (e.g., a math problem) and direct feedback (Worsdell et al., 2005). However,
Skinner et al. (1997) have cautioned that these types of procedures may be perceived as a punishment for incorrect responding. As such, one must consider possible negative side effects when taking into account the potential consequences of using a procedure that involves a punishment component. For example, if the procedure is too tedious or aversive, students may avoid the overcorrection practice and pretend they have not responded incorrectly. This could be done in several ways. For example, the teacher is not always present to give immediate feedback. If peer tutoring is being utilized, students may agree not to tell the teacher about errors so they do not have to engage in an error-correction procedure.

Treatment acceptability has been found to be an inherent part of successful interventions (Skinner et al., 1997). If students perceive an intervention as helpful or beneficial, they may apply it to other aspects of their academic experience. For example, it was reported that during a math study that assessed the acquisition intervention of cover-copy-compare, a student utilized that intervention procedure to study for a science test without prompting from experimenters (Skinner, Turco, Beatty, & Rasavage, 1989). Although the combination of positive practice overcorrection and another form of intervention may yield higher rates of accurate performing, Skinner et al. (1997) noted that if the error-correction procedures are perceived as too aversive, they may cause students to avoid or escape the overall intervention. This is particularly concerning given the general acceptance of some of the other interventions when they are used in isolation. Thus, Skinner et al. recommend that the number of accurate responses required for a positive practice overcorrection procedure be limited so as not to discourage student engagement.
Worsdell et al. (2005) believed it was important to determine the extent to which the negative reinforcement of the RR procedure may improve student performance. They conducted a study in which participants were using RR procedures to improve their ability to read sight-words. In one intervention phase, a correct RR procedure was utilized in which an incorrect response required a student to repeat back the correct word five times while looking at that word on a flashcard. In the other phase, an irrelevant RR procedure was implemented in which students were required to follow the same procedure, but they were only required to repeat back an irrelevant word.

Worsdell et al. (2005) found that only three of the nine participants showed greater increases during the relevant phase when compared to the irrelevant phase, while five of the nine participants showed no discernible difference between the two phases. One participant showed a greater increase during the irrelevant condition. These results indicate that benefits of the RR procedures were not solely linked to the repetition of the correct response, meaning that repetition of any item, even if it was unrelated to the incorrect item led to positive results. Worsdell et al. suggest these results indicate that the negative reinforcement aspect of the RR procedure largely contributes to improved accuracy.

Although mixed results were found with regards to the comparison of relevant and irrelevant RR procedures, improvements in reading ability were found in both conditions (Worsdell et al., 2005). Despite the fact that Relevant RR procedures were not significantly better than irrelevant RR procedures, it seems as though the use of relevant procedures makes logical sense because they are more tightly aligned with instructional goals. Worsdell et al. (2005) report that while only modest differences were found
between the two procedures, the Relevant RR procedure was slightly superior to the irrelevant RR procedure. It is possible that these small differences may have become more significant if the study was conducted over a longer time period. Additionally, Relevant RR procedures provide two possible benefits; they combine opportunities for practice and add a component of negative reinforcement (Worsdell et al., 2005).

Benefits of error-correction procedures. An interesting aspect of the positive practice overcorrection and RR procedures is that all studies have yielded positive gains in accuracy and/or fluency rates of academic skills. The studies have assessed reading (Ferkis et al., 1997; Marvin et al., 2010; Worsdell et al., 2005) and math skills (Rapp et al., 2012; Rhymer et al., 2000) in students of varying ages. The procedures have included various ways of presenting relevant stimuli (e.g., flash cards, word lists, worksheets) and the quantity and accuracy of required responses following errors has differed across studies. Some studies have asked students to provide correct responses, while others have asked students to provide related responses or irrelevant responses. Despite all of these differences, positive effects were found. The implication is that error-correction procedures involving opportunities for response repetition in the wake of errors are beneficial for student learning.

As previously noted only two studies have applied error-correction procedures to the area of mathematics (Rapp et al., 2012; Rhymer et al., 2000). Given the general success of these types of interventions, especially in reading, further investigation of the utility of such procedures for increasing accuracy and fluency scores for basic math tasks is a fruitful line of inquiry. The current study sought to do so by utilizing a positive practice overcorrection procedure in conjunction with peer tutoring and performance
feedback procedures. Peer tutoring will be described in detail later in this paper, while performance feedback will be discussed as a fluency intervention in the next section.

**Fluency interventions.** As discussed in the previous section, accuracy is the goal of the first stage of learning a new task. Accuracy, however, is not enough. Students who have progressed through the acquisition stage of the instructional hierarchy are able to respond accurately; they continue, however, to respond slowly and with some difficulty (Haring et al., 1978). Students who accurately and slowly complete academic tasks are in the second stage of the instructional hierarchy, known as the fluency stage (Haring et al., 1978). Fluency interventions, therefore, seek to increase the speed with which items are completed. In order to attain this goal, students require frequent opportunities for repetition and practice, feedback regarding fluency and accuracy of performance, as well as encouragement and praise for increased rates of responding (Haring et al., 1978). Within the context of the instructional delivery model, the needs of the students within the instructional range parallel those of students in the fluency stage of the instructional hierarchy: students need to increase the speed at which they can accurately perform an academic task (Gickling, 1977). Thus, it can be inferred that students performing in the instructional range of the instructional delivery model would benefit from fluency interventions as well.

Fluency interventions differ from acquisition interventions in specific ways. Although both include opportunities for repetition and practice, the nature of feedback and praise are different. During acquisition interventions students receive feedback on their performance on individual items, while during fluency interventions feedback is given after the completion of the academic task in the form of an overall fluency score.
In the area of mathematics, this means that students are told how many Digits Correct per Minute (DCPM) they were able to complete. Sometimes, students are also told how many Digits Incorrect per Minute (DICPM) were present in the work sample, meaning that they were told how many errors were present in their work. Students and instructors may look at individual mistakes, but the focus is now on overall speed of performance. While praise is focused on the amount of effort put forth by the student during acquisition interventions, it shifts to increased rates of accurate responding during fluency interventions (Haring et al., 1978).

**Types of fluency interventions.** Over the years, researchers have examined the effectiveness of several types of fluency interventions. Contingent reinforcement (Carson & Eckert, 2003; Freeland & Noell, 2002; Panahon & Martens, 2013), self-monitoring (Lannie & Martens, 2008; Magg et al., 1993), interspersal (Cates & Skinner, 2000; Hawkins et al., 2005; McDonald & Ardoin, 2007; Monterello & Martens, 2005; Skinner, 2002), and explicit timing (Codding et al., 2007; Rhymer et al., 2002; Rhymer & Morgan, 2005; Van Houten & Thompson, 1976) have been identified as effective fluency interventions. While each intervention has unique features, they all provide frequent opportunities for repetition and practice, feedback regarding fluency and accuracy of performance, as well as encouragement and praise for increased rates of responding. As previously noted, these are the defining components of the fluency interventions described by Haring et al. (1978).

With any intervention, the role of the student is crucial. According to Carson and Eckert (2003), student effort and motivation are other factors that can play a role in low accuracy and fluency scores. Haring et al. (1978) suggest that a key component of
fluency interventions involves praising students for increased scores. Further, once students have acquired the skill in the acquisition phase, they now must be motivated to practice and enhance that skill. Thus, many of the fluency interventions incorporate basic behavioral principles in order to increase student engagement and effort.

Although many studies have supported the use of fluency interventions in general, Burns et al. (2010) report that relatively few single-case design studies exist in which researchers consider the instructional level of students before applying fluency interventions. The results of Burns et al.’s meta-analysis indicate that fluency interventions produce small to moderate effect sizes when they are applied to students in the frustrational range of performance; however, they were unable to draw any specific conclusions regarding the effectiveness of fluency interventions when applied to students in the instructional range. This raises serious questions given that fluency interventions are designed to meet the needs of students performing in the instructional range. As previously noted, Burns et al. also found that acquisition interventions yield small to moderate effect sizes when they are applied to students who were performing in the instructional range, suggesting that acquisition interventions are better suited for students performing in the frustrational range. Thus, because similar effect sizes were seen when fluency interventions were applied to students in the frustrational range, they propose that this could be evidence to support the notion that fluency interventions are not best suited for students in the frustrational range of performance. This conclusion clearly highlights the need for further research in this area.

The current study sought to expand the limited research regarding fluency interventions by focusing on explicit timing procedures, whose practices have been found
to be effective (e.g., Codding et al., 2007; Rhymer et al., 2002; Van Houten & Thompson, 1976) but have not been thoroughly examined when considering student skill level as a mediating factor. The following section details how explicit timing incorporates the inherent features (e.g., practice, feedback, praise) put forth in the fluency stage of the instructional hierarchy (Haring et al., 1978) to meet student needs.

**Explicit timing.** Explicit timing is a fluency intervention in which students are given a specific time limit to complete an academic task (Rhymer et al., 2002). Moreover, students receive feedback on their progress during periodic intervals as they complete a task. Codding et al. (2007) note that this allows students to receive feedback regarding how many problems they have completed at predetermined intervals, which may naturally motivate them to perform quicker as each minute passes. This procedure has been shown to be effective in numerous studies (e.g., Codding et al., 2007; Rhymer et al., 2002; Van Houten & Thompson, 1976).

In its most common form, students are given a worksheet and told that they will have a certain amount of time to complete as many problems as possible. Typically, the time frame is broken into smaller increments after which students mark progress. For example, Codding et al. (2007) asked students to practice a math worksheet for five minutes. After each minute they were asked to stop and circle the problem they have just completed. They then continued working until the next minute is up. This process was repeated until the student had worked for five consecutive minutes. Other researchers (e.g., Rhymer et al., 2002) have used explicit timing practices to set specific limits regarding the length of time students were given to practice with flash cards. For example, at the end of two minutes, students could be asked to count how many
flashcards have been answered correctly, rather than looking at a work sheet to see how many problems have been completed. Thus, the inherent feature of explicit timing procedures is that they involve a brief and specific time period after which feedback regarding rate of performance is assessed and shared with the student.

Performance feedback is inherent in this type of procedure as students can see the number of items they have completed in each minute based on the circled problems, or the number of flashcards they have completed when they finish the task (Codding et al., 2007). However, feedback is only given regarding to the number of items completed and does not provide information regarding accuracy. Thus, it is important that students are able to accurately respond before engaging in this intervention. Similar procedures are used in all explicit timing interventions, although the materials used (e.g., flashcards, worksheets) and length of the intervention may vary.

Research has supported the use of these types of interventions. Early research by VanHouten and Thompson (1976) found that the use of explicit timing procedures led to increased fluency rates for the completion of basic math facts for second-grade students, while maintaining a high level of accuracy. Similarly, Rhymer and Morgan’s (2005) study found that explicit timing increased third-grade students’ fluency rates and maintained acceptable levels of accuracy on two-digit by two-digit subtraction problems.

Other studies have sought to further investigate the effectiveness of explicit timing procedures. For example, Rhymer et al. (2002) found that explicit timing improved fluency scores for the completion of basic math facts in sixth-grade students. However, when assessing for increased fluency rates in more complex tasks, such as multi-digit multiplication, Rhymer et al. found that the intervention was not effective.
This may indicate that explicit timing is a procedure that should be reserved for basic skill development. Furthermore, Coddington et al. (2007) found that explicit timing was more effective than a cover-copy-compare procedure when attempting to improve subtraction fluency in third-grade students who were performing in the instructional range (Deno & Mirkin, 1977). They suggest that this supports the theory that fluency interventions are superior to acquisition interventions when students are performing in the instructional range. Overall, research appears to suggest that explicit timing procedures are most effective when applied to the learning of basic skills within the instructional range.

Little research exists, however, with regards to how explicit timing might be combined with other effective procedures. Because it has been hypothesized that students are naturally motivated to work quickly and accurately when presented immediate feedback following a time constraint (Coddington et al., 2007), this type of procedure might keep students working at a fast pace on other types of interventions. For example, if an acquisition procedure were timed, the student may progress more quickly and be able to build accuracy and fluency simultaneously. This study sought to explore this hypothesis as it investigated the use of explicit timing procedures in two ways. First, it was used as part of a multi-component intervention. Second, it was used in isolation in order to investigate its level of effectiveness when considering a student’s instructional level.

Thus far, this paper has focused on academic interventions that have been developed based on the instructional hierarchy. It has been clear that acquisition and fluency interventions require a significant amount of individual and direct feedback. This
can be problematic if educators are required to constantly provide that feedback, as this requires a significant amount of educational resources in a time when budget cuts and staff reductions are common place. Thus, it has been suggested that these interventions can be implemented within the framework of peer tutoring interventions. As the present study involved the use of peer tutoring, the following section will explore the research regarding its use.

**Peer Tutoring: An Efficient Method for Delivering Individualized Intervention**

Peer tutoring is an instructional method in which students serve as tutors for one another (Cates, 2005). Class-wide peer tutoring (CWPT) procedures are a group of instructional strategies in which all the students in a class are trained and supervised by the classroom teacher to instruct other students (Kunsch et al., 2007). Often times, reciprocal peer tutoring is used in which students take turns and serve as both the tutor and as the tutee (Kunsch et al., 2007). One of the benefits of CWPT is that it allows students to provide the necessary feedback and reinforcement to one another, thus allowing individualized instruction or intervention to take place simultaneously with multiple students. While many of the error-correction procedures previously discussed have utilized procedures in which an adult and a student are working one-on-one, this type of intervention can also be implemented within the structure of CWPT. CWPT are preferred by educators because more students can respond in the same amount of time in which only one student would be able to respond during teacher-led intervention (Kunsch et al., 2007). Thus, the use of CWPT is extremely efficient because fewer educational resources are necessary as one teacher can utilize individualized interventions for several students at one time (Cates, 2005).
Robinson et al. (2005) conducted a review of literature that explored many different aspects of peer tutoring procedures. They found that peer tutoring was beneficial for both tutors and tutees at all levels of academic achievement. That is, when students who are performing in the average or below average range on the targeted skill act as tutors, positive academic outcomes have been noted for both the tutee and tutor. This means that students can serve both the tutor and tutee role regardless of their current level of academic performance. These findings suggest that reciprocal peer tutoring is a highly effective educational practice.

**Reciprocal peer tutoring.** Cates (2005) implemented a reciprocal peer tutoring procedure in which students used flashcards to drill one another for three minutes. The tutor would show the tutee a flashcard. If the tutee answered correctly, the tutor said “good” and put the flashcard in the correct pile. If tutees answered incorrectly, they would have to try again. After the three minutes, students counted the number of flashcards in the correct pile. Although Cates did not describe any particular type of intervention, this procedure clearly combined explicit timing, with a performance feedback component as students were given a specific time within which they were to complete the task and they were provided with immediate feedback regarding accuracy and fluency. Results of Cates’s study indicated that the peer tutoring intervention showed improved accuracy and fluency scores for all four participants. However, it was found that the younger, 8-year old dyad showed greater gains than did the older, 10- and 11-year old dyad. These findings are consistent with the results of the Kunsch et al. (2007) meta-analysis in which larger effect sizes were found for elementary-age students where
the effect size was 0.57, as compared to secondary-school students where the effect size was 0.18.

**Reciprocal verses non-reciprocal class-wide peer tutoring (CWPT).** As previously noted, CWPT procedures maximize opportunities for learning, as they allow multiple students to receive individualized feedback without one-on-one attention from a teacher (Cates, 2005). Menesses and Gresham (2009) conducted a study with second, third, and fourth graders who were performing below grade level on CBA probes. These students were randomly assigned to a reciprocal CWPT condition, a nonreciprocal CWPT condition, and a baseline condition. They used a flashcard procedure similar to the one used by Gardner et al. (2001) in that a single-response procedure was used if the student was unable to answer correctly or within three seconds. The students used 10 flash cards within a 3-minute time frame. Once they went through the 10 cards, they were shuffled and re-administered. Following the three-minute interval, the tutor presented the 10 cards a final time and did not provide any verbal feedback. Correct cards were posted on a green square and incorrect cards were posted on a red square. Once the student was able to answer all 10 items correctly on two consecutive occasions, 10 new cards were provided for the tutoring procedure. Although a specific time limit was set, this is not an example of explicit timing as students were not attempting to complete as many items as possible in a given time frame. Nevertheless, the results highlight the effectiveness of peer tutoring.

Results of Menesses and Gresham’s (2009) study indicate that both reciprocal and nonreciprocal procedures yield comparable improvements for tutees. Moreover, the authors note that reciprocal peer tutoring is more efficient than the nonreciprocal
approach because twice as many students are working to develop a particular skill in the same time frame. Those students who only served as tutors did not show as much of an improvement in performance in pre- and post-test scores. Additionally, students who participated in the CWPT conditions yielded greater gains than students who were in the baseline condition. There was not marked improvement in the scores of the students in the baseline condition between the pre- and post-test scores. Meneses and Gresham also note that this study provides evidence that students who are performing below average in mathematics can serve as effective tutors, which further supports that idea that a peer tutor does not have to be a highly-skilled student.

**Reciprocal class wide peer tutoring (CWPT).** Hawkins et al. (2009) used a similar reciprocal CWPT strategy in which students tutored each other for five minutes each using single-digit multiplication facts. They used a procedure identical to the one described by Meneses and Gresham (2009), except students worked with 12 cards within a 5-minute period instead of 10 cards in a 3-minute period. Results indicate that these procedures lead to significant increases in fluency rates of accurate responding with the introduction of the peer tutoring intervention. While Hawkins et al. utilized Deno and Mirkin’s (1977) fluency criteria, raw data can be applied to the criteria developed by Burns et al. (2006). When using the more recent criteria it can be seen that seven students yielded average baseline scores in the instructional range of performance, while three students yielded average scores in the frustrational range. Results indicate that mastery levels of performance were achieved by five students for the first problem set, by six students for the second set, and by four students in the third problem set. Perhaps the lower levels of achievement in the third set were due to the fact that the intervention was
utilized across fewer sessions as the third set was the last to be included in the intervention phase. Additionally, all students displayed a substantial increase in fluency scores upon the implementation of the peer tutoring condition.

Gardner et al. (2001) used a reciprocal CWPT tutoring procedure as part of an after-school program in which the main purpose was to increase academic skills in at-risk youth. Participants had a history of behavior difficulties, presented at one or more years below grade level in both reading and math, and all were from low-socioeconomic, African American families. In this study, students used flashcards to practice multiplication problems. The tutor would show the tutee a flash card and if the student answered correctly, the tutor would say “good” and move on to the next card. If the tutee answered incorrectly, the tutor would say “try again”. If the student answered incorrectly again or took longer than three seconds to respond, the tutor would state the problem with the correct answer and ask the tutee to repeat the math fact and answer. Although Gardner et al. did not explicitly discuss acquisition interventions or the specific components of the peer tutoring intervention, this repetition of the correct response is clearly an example the single-response (SR) repetition procedure described by Worsell et al. (2005).

Results of Gardner et al.’s (2001) study indicated that all students improved both accuracy and fluency rates on multiplication facts. Although Gardner et al. (2001) did not discuss instructional levels, they did use fluency scores as indicators of success. Thus, it is possible to apply the fluency criteria developed by Burns et al. (2006) when considering the results of this study. Data showed that eight of the 10 participants began this study in the frustrational range of performance, while one of the students was
performing in the mastery range and another was performing in the instructional range. Post-test results indicated that only two of the students remained in the frustrational range of performance. However, it is important to note that one of these students was not able to complete any problems correctly at the start of the intervention. Impressively, seven of the participants scored in the mastery range of performance according to the criteria put forth by Burns et al. (2006) during the post-test. One of the students remained in the instructional range; however, this student still improved a total of 25 DCPM over the course of the study. Additionally, accuracy rates showed a substantial improvement. Eight of the students began the intervention performing with less than 30% accuracy, and seven of these students improved to perform with over 92% accuracy. Four participants improved to 100% accuracy at the post-test. These results indicate that the class-wide reciprocal peer tutoring method was successful in increasing fluency and accuracy rates for at-risk youth. The studies reviewed in this section provide evidence that class-wide peer tutoring is effective in helping students improve basic skills in mathematics; yet there is more to understand.

**Need for further research.** The research regarding peer tutoring is extensive and has demonstrated that peer tutoring is an effective tool for improving mathematics skills. The abovementioned procedures have included opportunities for practice, immediate feedback, and some have included a component of explicit timing (Cates, 2005; Hawkins et al., 2009; Menesses & Gresham, 2009). These studies have included participants of different ages, grades, race, gender, and ability levels. Moreover, various procedures have been employed to address different types of basic math facts. Some studies have demonstrated that filling the role of the tutor has positive academic outcomes and that the
selection of a student as a tutor does not need to be based on the academic ability of that student (Robinson et al., 2005). Thus, peer-tutoring programs in which students serve both the role of the tutee and the tutor may be largely beneficial (Kunsch et al., 2007).

While the research regarding peer tutoring is extensive, there remains a need for further investigation. For example, there is a lack of literature regarding the effectiveness of particular interventions for students of different skill levels. Although Hawkins et al. (2009) note that students were performing at the low end of the instructional range according to the Deno and Mirkin (1977) criteria, they do not discuss the implications of this, nor do they discuss student outcomes in terms of progression on the learning hierarchy. Other studies have included fluency and accuracy data for pre- and post-test comparisons, but have not determined the extent to which these types of procedures may be effective for students who are performing in different ranges (e.g., Gardner et al., 2001; Menesses & Gresham, 2009). Additionally, although several studies have included minority students and at-risk youth, Robinson et al. (2005) note that there is a need for additional research regarding students who receive free and reduced lunch. This is particularly relevant as national and international studies have identified the free and reduced lunch population as a group that does not make the same progress in mathematics as those students who do not receive free and reduced lunch (NAEP, 2012, OECD, 2013b; Provasnik et al., 2012).

Performance feedback has been an integral element of all of the interventions discussed thus far. Again, the main benefit of peer tutoring is that offers an opportunity for students to provide feedback to one another, which allows for more students to receive individualized feedback at one time (Cates, 2005; Kunsch et al., 2007).
Additionally, a predominant feature of both acquisition and fluency interventions is performance feedback (Haring et al., 1978). While various types of interventions require very different types of feedback, it is clear that performance feedback allows for students to make academic gains. The following section will discuss the important role of performance feedback in different types of intervention procedures.

**The Importance of Performance Feedback**

Performance feedback has been found to be an important component of many different types of interventions (Brosvic, Dihoff, Epstein, & Cook, 2006; Codding et al., 2006; Codding et al., 2007; Codding, Lewandowski, & Eckert, 2005; Dihoff, Brosvic, Epstein, & Cook, 2005); it provides an opportunity for students to receive information regarding specific performance on an academic task. Performance feedback seeks to motivate students because it may cause students to be intrinsically motivated to exceed previous performances (Shapiro, 2011). This is particularly useful because natural reinforcers are preferred, as they are more effective than external reinforcers (Alberto & Troutman, 2009). Additionally, behaviors that are maintained through natural reinforcers have an increased chance of being maintained over time (Alberto & Troutman, 2009).

Performance feedback has already been discussed within the context of various interventions. For example, Haring et al. (1978) report that students must receive information regarding their performance in order to make meaningful corrections and show growth throughout the instructional hierarchy. However, the type of feedback varies depending on the instructional level of the student. As previously noted, acquisition interventions require immediate response evaluation and error correction, while fluency interventions require that only feedback regarding a student’s fluency score...
be given after the completion of an academic task. In addition to these two very different forms of feedback, different types of praise, which is a positive form of feedback, are recommended for each stage of the instructional hierarchy (Haring et al., 1978). It is recommended that acquisition interventions include a component of praise that is specific to student effort, while the focus shifts to increased rates of accurate responding during fluency interventions (Haring et al., 1978). The praise aspect of the instructional hierarchy is important as behavior-specific praise has been found to increase time on-task, responding, and correct answers (Duchaine et al., 2011). What has yet to be explored is the timing of the feedback; this will be addressed in the next section.

**Methods of delivering feedback.** The provision of immediate, affirming, and corrective feedback facilitates the acquisition and retention of basic mathematic skills. Brosvic et al. (2006) compared the effects of immediate verses delayed feedback on accurate student responding to basic math facts. They found that delayed feedback showed no improvement over the control condition, but immediate performance feedback showed significant growth in the accuracy of student performance. Brosvic et al. also compared the way in which feedback was delivered. In one condition, the educator provided the feedback, while in another the answer was concealed on the page for students to see after providing an answer. They found no difference in the method of delivery when assessing student outcomes.

In a similar study, Dihoff et al. (2005) found that students with slight cognitive delays also responded well to immediate performance feedback. In contrast with the findings of Brosvic et al. (2006), these students showed increased accuracy when feedback was provided by the teacher as opposed to independent methods, which
suggests that students who have greater cognitive and academic deficits experience increased benefit with one-to-one instruction. However, Dihoff et al. note that the students showed significant improvement in both conditions. Therefore, if resources and teacher time is limited, there is evidence that an independent procedure can also be successful. Unfortunately, both of the Dihoff et al. and Brosvic et al. studies reported accuracy percentages and no information is available with regard to fluency of performance. Thus, it is impossible to determine the instructional range in which these students were performing.

**Performance feedback within interventions.** Performance feedback is frequently an integral part of treatment packages. For example, error-correction procedures are a group of acquisition interventions that rely heavily on immediate corrective feedback within the procedure, as students are provided with immediate feedback following each item and required to correct mistakes (Rapp et al., 2012). Additionally, performance feedback is an intrinsic part of explicit timing interventions, which are a group of fluency interventions, as students are able to see how many items they have been able to complete in the specified time period (Codding et al., 2007). As performance feedback is a key component of both of the interventions that were used in this study, it is necessary to understand its scope and utility.

**Performance feedback as an isolated component.** Codding et al. (2007) suggested that performance feedback can also be used as an added component to existing academic interventions. However, relatively little research exists with regard to the utility of performance feedback as the main component of an intervention within the area of mathematics. Codding et al. (2005) examined the utility of performance feedback and
goal setting as an intervention for improving basic math fluency in students with attention deficit hyperactivity disorder (ADHD). They found that students gained more fluency when they were able to set their own goals based on the performance feedback they received. The study provided evidence that performance feedback interventions can be used to help improve math fluency scores. However, because the study focused on students with ADHD, more research is necessary for the results to be generalized to other populations.

Few studies have focused on performance feedback as an isolated component within treatment packages, although several studies have combined performance feedback with other types of interventions to enhance their effectiveness (Codding et al., 2007). In one study conducted by Codding et al. (2006), a performance feedback component was isolated and used in conjunction with the acquisition intervention known as cover-copy-compare. In this study, students were given information regarding their performance on the assessment that directly followed the cover-copy-compare procedure. This was in order to determine if adding an assessment feedback component regarding correct or incorrect digits would make the cover-copy-compare intervention more effective. Results of Codding et al.’s (2006) study indicated that fluency rates did not show significant improvement when either type of feedback was provided as compared to when cover-copy-compare was used in isolation. Codding et al. (2006) hypothesized that differences may not have been accounted for because performance generalized across conditions in their alternating treatments design. That is, as students acquired the skill they were able to perform it regardless of which type of feedback was given.
Another study that was conducted by Rhymer et al. (2000) also supported the use of performance feedback as a specific addition to a multi-component intervention. In this research, performance feedback was used as the final phase of a multi-component intervention in which students were provided information regarding their previous highest score and encouraged to beat it (Rhymer et al., 2000). Three out of four participants showed an increase in fluency scores with the addition of this component, indicating that performance feedback increased the effectiveness of the intervention.

The results are mixed when performance feedback is combined with other interventions, further supporting the need for more research.

Need for further research. Currently, we know that fluency when performing basic problems in mathematics is important. In addition, the delivery of instruction or interventions matters, and reciprocal CWPT is an efficient method that has had a positive impact on increasing students’ fluency scores. As students are building accuracy and fluency, the feedback they receive impacts their progress. However, the relationship between particular interventions and types of feedback with instructional levels is less clear. Thus, we need to know more.

Based on the pervasive presence of performance feedback throughout the various interventions, it seems clear that additional research is needed to better understand the role of performance feedback in increasing math fluency scores. Furthermore, because research has yielded mixed results regarding the addition of performance feedback to existing interventions (e.g., Codding et al., 2006; Rhymer et al., 2000), it is necessary to explore the effectiveness of performance feedback within a multi-component intervention. The current study sought to expand the literature by comparing a multi-
component intervention to an explicit timing intervention combined with performance feedback. It explored the effectiveness of a simple intervention when it is applied to students with varying instructional levels. In addition, this study attempted to examine whether an added component of performance feedback would have any impact on the effectiveness of peer tutoring intervention.

This literature review has detailed the utility of acquisition and fluency interventions, and has discussed peer tutoring for a framework within which a whole class of students can receive individualized intervention. The importance of performance feedback throughout all of these procedures has been detailed throughout this section. The following section provides a detailed account of a study that sought to explore the importance of performance feedback when the acquisition intervention of positive practice overcorrection is combined with the fluency intervention of explicit timing. The two interventions are presented within the context of a reciprocal peer-tutoring model.

**Tying It All Together: A Combined Intervention**

Thus far, this paper has demonstrated that error-correction, explicit timing, and peer tutoring have been found to be effective methods for improving rates of fluent responding to basic math facts. In addition, it has been shown that while performance feedback is an integral part of these interventions, it can also be used in a more explicit manner for the purposes of increasing student engagement (Codding et al., 2005). While these types of interventions have been found to be effective in isolation, relatively little research looks at multi-component interventions. This is particularly concerning given that Codding et al. (2011) found that students who were performing in the frustrational range, according to the new criteria put forth by Burns et al. (2006), benefited more from
interventions that included multiple treatment components, with modeling as a key feature. Despite this finding, Burns et al. (2010) note that Rhymer et al. (2000) conducted the only study in which acquisition and fluency interventions have been combined into one treatment package. In addition to being the only study that has combined these two types of interventions, it is also one of two studies that have ever utilized a positive practice overcorrection method in an attempt to improve math fluency skills (Rapp et al., 2012). Interestingly, despite the rare features of this study and the fact that it was found improve fluency scores, these procedures have not been further investigated. This section will provide a detailed account of the Rhymer et al. study as the purpose of the present study was to provide additional support for the combining of acquisition and fluency interventions with and without an additional component of performance feedback.

Rhymer et al. (2000) used an alternating treatments design to investigate the effectiveness of a multicomponent intervention that combined fluency and acquisition procedures within the framework of a reciprocal peer tutoring procedure. Participants for this study were four fourth-grade students. When analyzing the data, Rhymer et al. reported the number of problems correct per minute, rather than DCPM, which makes it impossible to determine the instructional level of these students. As previously discussed, both Deno and Mirkin (1977) and Burns et al. (2006) use DCPM to determine where a student is performing on the instructional hierarchy.

In Rhymer et al.’s (2000) study, the problem sets were based on those previously developed by Skinner et al. (1989). There were 36 problems divided into three sets, each containing 10 mutually exclusive one-digit by one-digit multiplication problems using
digits 2 through 9. No problems were included in more than one set. The use of these sets allowed Rhymer et al. (2000) to use a set of problems for each condition of the experiment. For example, set A could be used only for the tutor condition, set B could be used for the tutee condition, and set C could be used for the baseline condition. This allowed experimenters to differentiate the effectiveness of experimental conditions. If the student showed progress for set B, but not set C, one could suggest that that intervention was successful.

During the baseline condition, the students completed three separate math sheets. Each sheet contained a different set of 12 problems as described above. Students were given one minute to complete each sheet and asked to raise a hand if they finished early. As the initial intervention phase began, students were grouped into two dyads. Each dyad had a chess clock. Students were assigned a different set of problems for the tutor condition, the tutee condition, and the control condition. Index cards for the tutoring intervention were color coded. To begin the peer tutoring intervention, the chess clock was set for two minutes. One student acted as a tutor and showed the tutee an index card from the prescribed set. If the tutee answered correctly, the tutor would say “Correct” and then place the index card in a box marked “Correct.” If the tutee answered incorrectly, the tutor would say, “Incorrect,” and the tutee would write the problem with the correct answer three times on a sheet of scrap paper. The tutor would then place that index card in a box marked “Incorrect.” This process would be repeated for two minutes. Immediately following this session, both students would have one minute to complete a math probe that contained the set of problems that had just been practiced. Following this assessment, the tutor and the tutee would swap roles and repeat the process. After
the next two minutes, both students would have one minute to complete a probe that contained the same set of problems that were practiced during the second session. Finally, both students would have one minute to complete a control set of problems that neither student had practiced. Following these procedures, the experimenter would record the number of flash cards that were answered correctly and incorrectly for each student. At the start of every session, each student was given performance feedback regarding his/her performance during the previous intervention session. Specifically, each student was given a sheet that listed the number of correct and incorrect cards they had completed in the prior session. They were encouraged to complete more correct and fewer incorrect cards for the new session. In this way, they were receiving performance feedback on their fluency rate.

During the second intervention phase, all procedures were held constant but a new component of performance feedback was added to the assessment phase. Specifically, before the assessment probes were administered, each student was given a sheet with the highest number of problems correct he or she had completed during any previous assessment, without regard for the particular set of problems to which the score was attributed.

Rhymer et al. (2000) reported that three of the four students showed slight increases in the number of problems they were able to complete per minute after the implementation of the first intervention phase across tutor and tutee conditions. However, two of the students showed a greater improvement for items that fell in the tutee condition, indicating that explicit timing, active responding, and the positive practice overcorrection procedure lead to greater improvements than simply evaluating a
peer’s responses (Rhymer et al., 2000). Additionally, no students showed improvement on the control problems, indicating that there was a benefit to both the tutor and tutee conditions. Furthermore, it was found that three of the four students showed a greater increase in fluency rates when the performance feedback for the assessments was added to the intervention (Rhymer et al., 2000).

Need for further research. When considering practical implications of using academic interventions in the classroom setting, it is necessary to consider the amount of educational resources required to implement the proposed interventions. Rhymer et al. (2000) make several suggestions for further research that would explore possible ways to decrease the amount of time and teacher demands required for the implementation of successful academic interventions. For example, they note that the inclusion of the added component of performance feedback may have been confounded because a version of performance feedback was already part of the initial intervention. They suggest that the use of performance feedback and repeated assessments be investigated in isolation as this process may lead to increased fluency scores without the addition of the peer tutoring intervention. This may be especially true for students who are performing in the instructional range, as fluency interventions have been found to be more effective than acquisition interventions for students at this instructional level (Codding et al., 2007). This line of research could be particularly useful if it is possible to determine the minimal number of treatment components necessary for student success based on instructional level. Codding et al. (2011) note that students benefit more from interventions containing multiple components, but also report that these students were largely performing in the frustrational range. If students performing in the instructional range require less intensive
interventions, this means that improvements can be made with the use of interventions that require less time and fewer educational resources.

Additionally, there were only four participants in the Rhymer et al. (2000) study. They suggest, therefore, that future researchers seek to extend the external and educational validity of this procedure by presenting it in a class-wide format. If these procedures can be applied to a CWPT model, fewer educational resources are necessary as one teacher can utilize individualized interventions for several students at one time (Cates, 2005). While it seems clear that the multi-component intervention proposed by Rhymer et al. shows great potential for use in the classroom setting, further research is necessary to better understand the individual components of this intervention and how they can be applied to a class-wide intervention.

**Rationale for the Current Study**

Currently, much importance has been placed on students’ skills in basic and complex mathematics (National Council of Teachers of Mathematics, 2000; National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010). To date, we have learned much about effective ways for improving students’ basic math skills. We know that accuracy and fluency matter and we have some evidence-based ways of intervening to help students improve in these two areas (Burns et al., 2010; Codding et al., 2011). Yet, we need to hone our understanding of how students’ skill levels impact the effectiveness of these interventions so that we are best able to help our students gain the mathematical competency they need to compete in the world beyond schools.
**Student struggles in the area of mathematics.** As previously noted, students in the United States continue to struggle in the area of mathematics (NAEP, 2013; OECD, 2013b; Provasnik et al., 2012). The results of the NAEP, TIMSS, and PISA studies also show that students from low-income families continue to fall behind their higher-income peers in the United States (OECD, 2013b; Provasnik et al., 2012). Many of these students are consistently failing to develop competence when performing basic math skills and they struggle to apply these skills to higher-level reasoning and application tasks (NAEP, 2013). This is concerning because basic understanding of mathematics is essential for many basic life activities, including purchasing goods and services, balancing a budget, and meeting typical work demands (Minskoff & Allsopp, 2003). Furthermore, students who struggle in the area of mathematics are likely to struggle in other subjects such as science, economics, and computer literacy (Minskoff & Allsopp, 2003). Thus, there is clearly a need for sound, empirically-based interventions in the area of mathematics that are effective for a broad range of students, specifically those who are low income and are falling behind.

Despite this need, the main focus of past research has been in the area of reading and has only recently shifted to include the area of mathematics (Lembke et al., 2012). Because of the historic focus on reading, there are relatively few empirically-supported interventions for increasing math skills (Codding et al., 2009). While there has been a recent increase in the number of research studies attributed to the area of mathematics, there remains a need to close the gap. The present study sought to close that gap by contributing further investigation of the use of several empirically-supported interventions.
**The importance of fluency for the development of math skills.** Educators have to consider the goals of the current curriculum when considering the types of math interventions that would be effective in ensuring student success. Fluent computation is a goal for mathematics instruction (National Council of Teachers of Mathematics, 2000; National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010), and basic computation skills are the foundation upon which success in applications and concepts is built (Haring et al., 1978; Gickling, 1977; Shapiro, 2011). Thus, it remains clear that interventions targeting fluent rates of accurate responding are inherently important to mathematics instruction. The present study sought to expand the literature by further investigating evidence-based interventions that target fluency building in the area of math.

**RTI: A framework for applying academic interventions in schools.** In response to the clear need for improved instruction in the area of mathematics, many schools are attempting to address academic difficulties through the use of RTI, which requires the use of progress monitoring to inform high-quality, evidence-based instruction (Fuchs & Fuchs, 2006). If educators are to apply RTI principles to the instruction of mathematics, it is necessary to have a wide repertoire of evidence-based procedures from which they can choose appropriate strategies for students who present with various types of difficulties and characteristics (Coddington et al., 2009; Lembke et al., 2012).

One of the key components of RTI is matching instructional interventions to the specific skill deficits that are displayed by a student (Vaughn & Fuchs, 2003). Thus, it is necessary to have a consistent and evidence-based process for doing so in order to ensure
that instructional decisions are made in an objective manner. Current research indicates that the fluency criteria proposed by Burns et al. (2006) can be applied to the instructional delivery model (Gickling, 1977) in order to identify appropriate interventions from the instructional hierarchy (Haring et al., 1978) for this purpose. These procedures were used to identify students’ instructional needs and assess the utility of the interventions, based on the level of these needs.

**CBM: A standardized procedure for assessing student skill levels.** In order to use the fluency criteria, it is necessary to have valid and reliable assessments that can yield DCPM information in order to have a consistent and accurate procedure for evaluating students’ academic skill levels. Furthermore, it is necessary to understand the nature of students’ progress so that interventions can be adapted and tailored to meet students’ specific needs. Shapiro (2011) suggests the use of CBM for determining the instructional level of a student, as CBM is a procedure that evaluates student performance through the use of ongoing assessment (Gickling & Thompson, 1985). Thus, CBM was used to assess student mathematical skills for the purposes of the current study.

**Academic interventions: Need for further research.** When considering the need for evidence-based mathematics interventions, Ardoin and Daly (2007) suggest that researchers continue to rely on the recommendations put forth in the instructional hierarchy (Haring et al., 1978). According to this model, students in the acquisition stage require modeling, feedback, and praise for effort, while students in the fluency stage require frequent opportunities for repetition and practice, followed by targeted feedback regarding the accuracy and fluency of performance, as well as praise for increased fluency (Haring et al., 1978). Performance feedback is clearly part of both acquisition
and fluency interventions. The present study sought to further explore the role of performance feedback in student learning.

Ardion and Daly (2007) note that the concept of the instructional hierarchy was developed in a way that conceptualizes academic instruction as a form of behavioral shaping. Haring et al. (1978) suggest that specific academic skills be considered target behaviors that can be brought under stimulus control through instruction and intervention. The present study sought to further evaluate the effectiveness of positive practice overcorrection and explicit timing, both of which contain elements from the instructional hierarchy.

When considering the implementation of an academic intervention one must consider the efficacy of this intervention. This means that it is necessary for researchers to understand how the intervention can be applied in the actual school setting. One factor that contributes to the efficacy of an intervention is the amount of educational resources that are required for the intervention to take place. The present study sought to address this issue by implementing academic interventions within the context of a CWPT procedure.

**Error-correction interventions and mathematics.** While many acquisition interventions have been found to be effective when attempting to increase student accuracy and fluency when completing basic math facts, relatively little research exists with regard to the use of error-correction procedures in mathematics (Rapp et al., 2012). Interestingly, error-correction procedures have yielded positive gains in accuracy and/or fluency rates in the area of reading (e.g., Ferkis et al., 1997; Marvin et al., 2010; Worsdell et al., 2005), suggesting that error-correction procedures involving opportunities for
response repetition in response to errors are beneficial for student learning. As only two studies (Rapp et al., 2012; Rhymer et al., 2000) have employed the use of the specific error-correction procedure known as positive practice overcorrection within the area of mathematics, further research is necessary to build upon the body of research. Thus, the present study utilized a positive practice overcorrection procedure as part of the intervention package in order to further assess its utility within the area of mathematics.

**Explicit timing combined with other academic interventions.** Research has shown that the fluency intervention explicit timing increases fluency rates of students (Codding et al., 2007; Rhymer et al., 2002; Van Houten & Thompson, 1976). It has been noted that performance feedback within these procedures is given when students are able to see the number of problems they have completed within the specified time frame (Codding et al., 2007). Although explicit timing has been found to be effective in isolation, little research exists with regard to how this simple procedure might be combined with other effective procedures (Rhymer et al., 2000). It is possible that students may be motivated to work more quickly and accurately when completing a different type of intervention if they are given a time constraint (Rhymer et al., 2000). This may allow students to build accuracy and fluency simultaneously, which would allow them to move through the acquisition stage at a faster pace. Thus, the present study sought to further the understanding of how explicit timing procedures can be used in conjunction with an acquisition intervention.

**Performance feedback as an isolated component.** Performance feedback is typically an integral part of both acquisition and fluency interventions. While explicit timing interventions allow students to see how many items they have completed in a
specified time period (Codding et al., 2007), error-correction procedures rely heavily on immediate corrective feedback following each item (Rapp et al., 2012). Additionally, Codding et al. (2007) suggested that performance feedback can also be used as an added component to existing interventions. However, relatively few studies have examined the utility of this practice and the results have been mixed. Codding et al. found that the addition of a performance feedback component did not enhance student performance, while Rhymer et al. (2000) found that it did for three out of four subjects. This highlights the need for further research assessing the utility of adding performance feedback to academic interventions.

Furthermore, Rhymer et al. (2000) suggest that the use of performance feedback and repeated assessments be investigated in isolation as this process may lead to increased fluency scores. This may be especially true for students who are performing in the instructional range (Codding et al., 2007). This line of research could be particularly useful if it is possible to determine the minimal number of treatment components necessary for student success based on instructional level (Codding et al., 2011). These findings would allow educators to make intentional decisions about how to use their time to maximize student learning.

The present study sought to contribute to the literature by investigating performance feedback in several ways. First, it was investigated as an individual component in conjunction with repeated assessments. Students received feedback regarding prior scores on CBA probes and were encouraged to beat those scores. Additionally, it was used as an inherent part of the positive practice overcorrection procedure. It was also explicitly added as part of a multi-component intervention, as
students received feedback regarding the number of overall problems they were able to answer correctly during the intervention procedure. Finally, the two procedures were combined to see if the feedback regarding scores on the CBA probes could help improve student progress following the multi-component intervention. Details regarding the implementation of these procedures will be provided in the next chapter.

**Individual academic intervention through CWPT in the classroom.**

Implementing class-wide peer tutoring (CWPT) procedures are another way to reduce the educational resources necessary to provide individualized interventions for students. This type of procedure may be largely beneficial, as many of the interventions, including positive practice overcorrection, require individual and immediate performance feedback. While significant support for the use of CWPT exists (e.g., Hawkins et al., 2009; Kunsch et al., 2007; Menesses & Gresham, 2009; Robinson et al., 2005), there remains a need for further investigation. For example, there is a lack of literature regarding the implications of student skill level and the effectiveness of these types of interventions. Additionally, although several studies have included minority students and at-risk youth, Robinson et al. (2005) note that there is a need for additional research regarding students who receive free and reduced lunch. Additionally, only one of the studies presented in this paper (Rhymer et al., 2000) utilized peer tutoring procedures as a way of delivering interventions explicitly linked to the instructional hierarchy and this study did not utilize CWPT procedures. The current study sought to expand the literature by utilizing acquisition and fluency interventions within the context of a CWPT procedure. This study also sought to examine the effectiveness of CWPT for students of poverty.
**Combined interventions.** Very few studies have employed the use of multiple types of interventions as one treatment package. In fact, Burns et al. (2010) note that there has only been one study in which both acquisition and fluency interventions have been combined into one treatment package. When these two types of interventions were combined, fluency scores for participants showed significant improvement (Rhymer et al., 2000). As previously noted, this study utilized a reciprocal, peer-tutoring procedure. Because the implication is that a multi-component intervention may be useful for students of varying ability levels and may allow students to improve fluency while building accuracy, the current study sought to expand upon the findings of Rhymer et al. (2000).

Although Rhymer et al.’s (2000) study yielded positive results, it has several limitations. First, Rhymer et al. report the number of problems correct per minute, rather than DCPM, which makes it impossible to determine the instructional level of these students. Thus, further research is necessary to determine how student ability level may impact the effectiveness of this type of intervention. Second, Rhymer et al. added a performance feedback component in which students received feedback regarding their performance on the assessment portion of the intervention after an intervention package had already been in place for some time. It was also utilized in conjunction with the intervention package, not in isolation. Rhymer et al. note that sequence effects may have confounded results and suggested that future research assess the utility of repeated assessments with the use of performance feedback as a stand-alone procedure. This line of research could be particularly useful if it is possible to determine the minimal number of treatment components necessary for student success based on instructional level.
Additionally, Rhymer et al. conducted their study with only four students; therefore, these procedures should be implemented with a larger number of participants.

Because it seems clear that the multi-component intervention proposed by Rhymer et al. (2000) shows great potential for use in the classroom setting, the present study sought to expand upon these findings in several ways. First, DCPM data was applied to the fluency criteria by Burns et al. (2006) in order to determine the instructional level in the instructional delivery model (Gickling, 1977). This allowed us to determine how student instructional levels may mediate the effectiveness of the current interventions. Additionally, as suggested by Rhymer et al., performance feedback on assessment probes was used in isolation to determine the nature of its impact on student scores. Finally, as suggested by Rhymer et al., the intervention procedures were implemented within the context of CWPT in order to further investigate the utility of this procedure in the classroom setting.

The importance of generalization. While many studies demonstrate that basic interventions can increase rates of fluent responding for basic skills, Codding et al. (2009) note that there is a lack of research regarding generalization effects. This gap in the literature is concerning because the current mathematics curriculum focuses instruction on concepts and applications and has de-emphasized the direct instruction of computational competency (Shapiro, 2011). However, it has been hypothesized that teaching academic skills through drill tasks leads to increases in student performance in other related skills (Burns, 2004). Thus, it remains clear that it is necessary to further explore the relationship between the acquisition of basic skills and improvement on academic measures that assess related and more complex skills. Thus, the present study
sought to contribute to the literature by comparing student performance on general outcome measures were administered as pre- and post-test assessments.

**Treatment integrity.** Thus far, this paper has demonstrated that fluency scores have increased as a result of various academic interventions. However, if interventions are not carried out according to specified procedures, results are confounded. Treatment integrity or fidelity refers to the extent to which procedures are followed according to written protocols. It is always important to assess for treatment integrity because if an intervention is not followed with fidelity, results may be confounded by extraneous variables. Although Rhymer et al. (2000) evaluated treatment integrity and found that the experimenter followed procedures 100% of the time, there are no reports with regards to student performance included in the article. Thus, the question remains as to how well students followed the tutoring procedures described in the study. The present study also attempted to assess procedural integrity as it applies to student participants, in addition to that of the experimenters.

**Treatment acceptability.** Treatment acceptability is also important, as it has been found to influence the fidelity with which an intervention is implemented. It has also been found to impact student outcomes with regard to improved fluency scores for basic math facts (Allinder & Oats, 1997). Treatment acceptability refers to the extent to which an individual perceives a treatment to be fair, reasonable, appropriate, and unobtrusive (Kaszdin, 1980). It has also been noted that one of the primary methods for determining the likelihood of transferring research to practice in the school setting has occurred through the study of treatment acceptability (Riley-Tillman, Chafouleas, Eckert,
& Kelleher, 2005). Thus, an important component of the present study was to determine the perceptions of students with regard to the current procedures.

**Research and Educational Implications**

As schools around the country are adopting and utilizing RTI procedures, there remains a clear need for sound, empirically-based interventions in the area of mathematics that can work for all students. However, there are relatively few empirically-supported interventions for increasing math skills as compared to those that have been studied with regard to reading (Codding et al., 2009). Additionally, because students who qualify for free and reduced lunch continue to fall behind their peers in the area of mathematics, it is necessary to ensure that interventions are selected that can be successful for students in this subgroup.

Research has shown that academic difficulties can result from a mismatch between student skill and instructional materials (Ardoin & Daly, 2007; Gickling & Thompson, 1985; Haring et al., 1978). The instructional hierarchy (Haring et al., 1978) is a heuristic that has been used in conjunction with the instructional delivery model (Gickling, 1977) as a guide for matching students’ instructional needs to appropriate interventions for over 35 years. However, relatively few research studies have shown how this heuristic can be applied to the selection and utility of brief, explicit, evidence-based interventions in the area of mathematics (Codding et al., 2011). Additionally, only a few studies have utilized the relatively new, empirically derived criteria presented by Burns et al. (2006) to draw conclusions with regard to the educational utility of these types of interventions.
The current study sought to expand the research literature and extend our understanding of how the instructional hierarchy can be used to make instructional decisions based on elementary-age students’ performance of basic math skills. It investigated the utility of combining acquisition and fluency interventions, which has only been done once before (Rhymer et al., 2000). In addition, this study investigated the use of performance feedback in isolation for the purpose of improving fluent responding to basic math facts. Thus, this study sought to expand our understanding how brief, explicit, and evidence-based interventions can be used in the general education setting in order to improve fluency skills of elementary school students in the area of mathematics. The results of this study should provide information that will allow educators to make better decisions when attempting to match interventions to student need.

**Research Questions**

The present study sought to answer the following questions:

**Question 1:** Is students’ SES, as indicated by free lunch status, related to the effectiveness of performance feedback, peer tutoring, and combined interventions?

**Question 2:** Will increased fluency scores correlate with a discernible increase in students’ scores on generalization measures?

**Question 3:** Are the performance feedback and peer tutoring interventions in isolation effective for increasing fluency of performance with regard to single-digit multiplication?

**Question 4:** Are the performance feedback and peer tutoring interventions more effective when they are combined into one multicomponent treatment?
Question 5: Is the tutee role of the peer tutoring intervention more effective than the tutor role?

Question 6: Does the students’ instructional level affect the effectiveness of these interventions?

Question 7: Does the students’ instructional level affect the amount of time necessary for students to achieve mastery levels of learning?

Hypotheses

The current research tested the relationship of a student’s ability level on the instructional hierarchy to the effectiveness of two types of interventions. The overarching hypotheses that were investigated included the relationship between the academic skill level and a multicomponent intervention, the relationship between the academic skill level and a performance feedback intervention, and the academic skill level and the multicomponent intervention in conjunction with the performance feedback intervention.

Hypothesis 1: The performance feedback, peer tutoring, and combined interventions will be equally effective for all participants, regardless of lunch status.

Hypothesis 2: Increased fluency scores on single-digit multiplication measures will correlate with a discernible increase in student scores on measures of more complicated computational skills as well as student’s understanding of mathematical concepts and applications.

Hypothesis 3: Students’ fluency scores will increase as a result of the performance feedback and performance feedback interventions.
Hypothesis 4: All students will show greater increases in fluency scores when the performance feedback and peer tutoring interventions are combined compared to their performance as a result of either intervention in isolation.

Hypothesis 5: Students will show greater improvements on problems included in the tutee condition than problems that are included in the control and tutor conditions.

Hypothesis 6: Students who are performing in the instructional range will show greater improvements in fluency scores as a result of the performance feedback intervention than will students who are performing in the frustrational range.

Hypothesis 7: All students will show an increase in fluency scores as a result of the peer tutoring intervention, with students who are performing in the frustrational range showing greater gains than those performing in the instructional range.

Hypothesis 8: Students who begin the intervention phases at the instructional range will reach mastery levels of performance sooner than students who begin the intervention at the frustrational range.
CHAPTER 3

Methods

This chapter describes the methodology for gathering data on the effectiveness of two brief mathematics interventions, in isolation and in combination, for students at two different instructional skill levels. A description of the participants and setting will be provided, followed by an explanation of the measures and procedures used in conducting this research.

Recruitment, Participants and Setting

Recruitment. The current study sought to understand how brief, evidenced-based interventions can impact fluency scores of students on single-digit multiplication assessments when these interventions are matched to students’ skill levels. One aspect of this study questioned if there is variable impact of these targeted interventions based on student levels of SES. Thus, it was necessary to find an elementary school that with an economically diverse population. Additionally, given the research design of this study which will be explained in detail later in this chapter, it was necessary to have a minimum of 16 subjects. Thus, it was essential that the possible subject pool be large enough to meet this minimum requirement in the selected school.

There are several economically diverse schools that were considered for this study. The first principal who was approached agreed for the study to take place in her school. As a result of meeting with the principal, two classes that contained third grade students were selected based on the predicted skill level of students and class size. The Principle investigator (PI) emailed the teachers of the two classes providing a brief description of the study and offered an opportunity to answer questions and provide more
information. Both teachers, Mrs. C and Mr. L, agreed to participate without further discussion.

Mrs. C taught a class that was comprised of 19 third graders, while Mr. L taught a combined second- and third-grade class that included nine third graders and six second graders. As only third-grade students were considered to be eligible, 28 permission slips were distributed prior to all experimental procedures.

Procedures for obtaining written and informed consent from parents and/or guardians of participants were approved through the Internal Review Board of the CUNY Graduate Center. Permission slips and self-addressed, stamped envelopes were distributed to students in class. Parents were given the option to mail the consent forms directly to the PI or return them to the teachers at the school (see Appendix A). Once parental consent was obtained, the PI reviewed assent forms with all potential participants (see Appendix B). To reduce the possibility that students would feel any coercion to participate, the teacher was not present during the time assent was provided by the students. Potential participants were given a choice to sign the assent form to indicate their willingness to participate in the study. The study began after all parental consent forms were collected and all participants provided their written assent. In the end, consent and assent forms were received from 15 students from Mrs. C’s third grade class and eight students from Mr. L’s combined class.

**Setting.** Students were selected from two classrooms in a small rural public elementary school in New England during the 2014-2015 school year. Demographic data for that year indicate that the total enrollment of the selected elementary school was 155 students, with 53% of students qualifying for free or reduced-price lunch (Vermont
Agency of Education [VAE], 2015). The student racial/ethnic origin was as follows: 93% Caucasian; 3.0% multi-racial; 3.0% Asian; 2.0% Hispanic or Latino; and 1% Black or African American. There were no English Language Learners enrolled and 12% of the student population qualified for special education services during this particular school year (VAE, 2015).

Due to limited space, intervention and data collection took place in the lobby of the school. Students were asked to sit in the same seat for every session. Some students worked at tables, while others sat in chairs and used clip boards to complete work. The lobby was generally quiet, but occasionally brief distractions occurred as students transitioned to and from gym class.

**Participants.** Of the 23 third graders who agreed to participate in the study, 18 students performed at within the instructional or frustrational ranges on math worksheets that assessed single-digit multiplication problems. Criteria from Burns et al. (2006) were used to determine students’ instructional level and indicate that students falling in the instructional range are able to complete between 14-31 DCPM on a given task. Students falling below this range are performing in the frustrational range of the instructional delivery model (Gickling, 1977), while students performing above this range are performing in the mastery range (Burns et al., 2006).

Based on the criteria for participation in the study, initial baseline procedures were implemented over the course of five days to determine students’ instructional levels. The data were evaluated and one student from Mr. L’s class was excluded from participation because she was already performing at the mastery level and not in need of intervention. Of the remaining 22 students, 12 earned an average score that fell in the
instructional range within the first week of the study, while the rest earned average scores in the frustrational range. Over the course of the study, one student from Mrs. C’s class opted to withdraw as she was not comfortable performing a timed task. Another student from Mrs. C’s class completed the sessions, but her data were not used because she missed a significant amount of sessions due to a family trip and, she was observed attempting to complete problems after the allotted time frame. It was unknown to the PI how often this behavior was occurring so it was impossible to determine which scores were accurate. The student was quietly asked to follow directions in the future and allowed to continue with the group and the peer tutoring procedures, but her scores were not used. Finally, after the completion of the study, the data revealed that two more students were performing in the mastery range of instruction during the first week of the baseline phase. Consequently, their data was removed from the study, but they participated in all intervention phases and data collection procedures.

Of the 18 remaining participants, 10 were performing in the frustrational range at the start of the study, while eight were performing in the instructional range (Burns et al., 2006; Gickling, 1977). Additionally, 11 of the remaining students were from Mrs. C’s third grade class, while seven were from Mr. L’s combined class. The group included 17 students who were identified as Caucasian and one student who was identified as Native Hawaiian or other Pacific Islander. There were 11 males and seven females. All students were in the third grade and their ages ranged from 8 years, 3 months to 9 years, 8 months old at the start of the study ($M = 8.86$ years; $SD = .79$). Five students (28%) from the sample were eligible to receive free lunch. Due to the experimental design, students
began intervention phases at various points throughout the study. Each student is
described in more detail below and their characteristics are summarized in Table 3.

Table 3

*Student Information*

<table>
<thead>
<tr>
<th>Student</th>
<th>Gender</th>
<th>Age</th>
<th>SES</th>
<th>Class</th>
<th>Initial Baseline Mean Score</th>
<th>Instructional Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Female</td>
<td>9</td>
<td>Paid</td>
<td>Mr. L</td>
<td>17.71 (5.70)</td>
<td>Instructional</td>
</tr>
<tr>
<td>2</td>
<td>Male</td>
<td>8</td>
<td>Paid</td>
<td>Mr. L</td>
<td>19.75 (3.66)</td>
<td>Instructional</td>
</tr>
<tr>
<td>3</td>
<td>Male</td>
<td>9</td>
<td>Paid</td>
<td>Mrs. C</td>
<td>17.80 (4.10)</td>
<td>Instructional</td>
</tr>
<tr>
<td>4</td>
<td>Male</td>
<td>8</td>
<td>Paid</td>
<td>Mrs. C</td>
<td>15.23 (6.16)</td>
<td>Instructional</td>
</tr>
<tr>
<td>5</td>
<td>Female</td>
<td>8</td>
<td>Paid</td>
<td>Mrs. C</td>
<td>8.10 (4.17)</td>
<td>Frustrational</td>
</tr>
<tr>
<td>6</td>
<td>Male</td>
<td>8</td>
<td>Free</td>
<td>Mr. L</td>
<td>10.80 (3.36)</td>
<td>Frustrational</td>
</tr>
<tr>
<td>7</td>
<td>Male</td>
<td>9</td>
<td>Free</td>
<td>Mrs. C</td>
<td>11.67 (3.84)</td>
<td>Frustrational</td>
</tr>
<tr>
<td>8</td>
<td>Male</td>
<td>8</td>
<td>Paid</td>
<td>Mrs. C</td>
<td>25.70 (5.38)</td>
<td>Instructional</td>
</tr>
<tr>
<td>9</td>
<td>Male</td>
<td>8</td>
<td>Paid</td>
<td>Mrs. C</td>
<td>18.20 (5.77)</td>
<td>Instructional</td>
</tr>
<tr>
<td>10</td>
<td>Male</td>
<td>8</td>
<td>Paid</td>
<td>Mr. L</td>
<td>20.05 (4.46)</td>
<td>Instructional</td>
</tr>
<tr>
<td>11</td>
<td>Female</td>
<td>8</td>
<td>Paid</td>
<td>Mrs. C</td>
<td>21.10 (4.41)</td>
<td>Instructional</td>
</tr>
<tr>
<td>12</td>
<td>Male</td>
<td>9</td>
<td>Paid</td>
<td>Mrs. C</td>
<td>19.07 (6.03)</td>
<td>Instructional</td>
</tr>
<tr>
<td>13</td>
<td>Female</td>
<td>9</td>
<td>Free</td>
<td>Mr. L</td>
<td>14.26 (4.34)</td>
<td>Instructional</td>
</tr>
<tr>
<td>14</td>
<td>Male</td>
<td>9</td>
<td>Free</td>
<td>Mrs. C</td>
<td>10.77 (3.45)</td>
<td>Frustrational</td>
</tr>
<tr>
<td>15</td>
<td>Female</td>
<td>8</td>
<td>Paid</td>
<td>Mr. L</td>
<td>4.26 (4.34)</td>
<td>Frustrational</td>
</tr>
<tr>
<td>16</td>
<td>Male</td>
<td>8</td>
<td>Paid</td>
<td>Mrs. C</td>
<td>10.00 (4.44)</td>
<td>Frustrational</td>
</tr>
<tr>
<td>17</td>
<td>Female</td>
<td>9</td>
<td>Free</td>
<td>Mrs. C</td>
<td>12.63 (2.87)</td>
<td>Frustrational</td>
</tr>
<tr>
<td>18</td>
<td>Male</td>
<td>8</td>
<td>Paid</td>
<td>Mr. L</td>
<td>10.53 (3.31)</td>
<td>Frustrational</td>
</tr>
</tbody>
</table>

*Note:* Free = Eligible to receive free lunch; Paid = Student pays for lunch
Student 1 was a 9-year-old female student of Caucasian descent, who did not qualify to receive free- or reduced- lunch. Student 1 received instruction in Mr. L’s combined second- and third-grade class. At the start of the study she performed at the 8th percentile on the M-Comp and at the 43rd percentile on the M-CAP. Student 1 earned an initial baseline mean score of 17.78 (5.70) which falls in the instructional range. It is notable that student 1 had the flu and missed four consecutive sessions, two of which occurred during the baseline sessions and two of which occurred during the start of the performance feedback intervention. Student 1 began the performance feedback intervention when she returned during the second week of the study and she began the combined intervention during the fourth week of the study.

Student 2 was an 8-year-old male student of Native Hawaiian or Other Pacific Islander descent. He did not qualify to receive free- or reduced- lunch. Student 2 receives instruction in Mr. L’s combined second- and third-grade class. At the start of the study he performed at the 18th percentile on the M-Comp and at the 37th percentile on the M-CAP. Student 2 earned an initial baseline mean score of 19.75 (2.66) which falls in the instructional range. He missed only one session throughout the course of the study, which occurred during the baseline period. Student 2 began the performance feedback intervention during the fourth week of the study and he began the combined intervention during the sixth and final week of the study.

Student 3 was a 9-year-old male student of Caucasian descent who did not qualify to receive free- or reduced- lunch. Student 3 receives instruction in Mrs. C’s third-grade class. At the start of the study he performed at the 60th percentile on the M-Comp and at the 27th percentile on the M-CAP. Student 3 earned an initial baseline mean score of
17.80 (4.10) which falls in the instructional range. He was present for all baseline and intervention sessions. Student 3 began the performance feedback intervention during the second week of the study and he began the combined intervention during the fourth week of the study.

Student 4 was an 8-year-old male student of Caucasian descent. He did not qualify to receive free- or reduced- lunch. Student 4 receives instruction in Mrs. C’s third-grade class. At the start of the study, he performed at the 73rd percentile on the M-Comp and at the 23rd percentile on the M-CAP. Student 4 earned an initial baseline mean score of 15.23 (6.16), which falls in the instructional range. He was present for all baseline and intervention sessions. Student 4 began the performance feedback intervention during the second week of the study and he began the combined intervention during the fourth week of the study.

Student 5 was an 8-year-old female student of Caucasian descent. She did not qualify to receive free- or reduced- lunch. Student 5 receives instruction in Mrs. C’s third-grade class. At the start of the study she performed at the 7th percentile on the M-Comp and at the 23rd percentile on the M-CAP. Student 5 earned an initial baseline mean score of 8.10 (4.17), which falls in the frustrational range. She missed one session during the baseline phase. She began the performance feedback intervention during the third week of the study and she began the combined intervention during the fourth week of the study.

Student 6 was an 8-year-old male student who was of Caucasian descent. He receives free lunch. Student 6 receives instruction in Mr. L’s combined second- and third-grade class. At the start of the study he performed at the 8th percentile on the M-
Comp and at the 27th percentile on the M-CAP. Student 6 earned an initial baseline mean score of 10.80 (3.36), which falls in the frustrational range. He missed two sessions throughout the course of the study. The first was during the Performance Feedback intervention and the second was during the Combined Intervention. He began the performance feedback intervention during the fourth week of the study and he began the combined intervention during the sixth and final week of the study.

Student 7 was a 9-year-old male student of Caucasian descent who receives free lunch. Student 7 receives instruction in Mrs. C’s third-grade class. At the start of the study he performed at the 43rd percentile on the M-Comp and at the 18th percentile on the M-CAP. Student 7 earned an initial baseline mean score of 11.67 (3.84), which falls in the instructional range. He missed two sessions throughout the course of the study. The first was during the performance feedback intervention and the second was during the combined intervention. He began the performance feedback intervention during the fifth week of the study and he began the combined intervention during the sixth and final week of the study.

Student 8 was an 8-year-old male student of Caucasian descent. He did not qualify to receive free- or reduced- lunch. Student 8 receives instruction in Mrs. C’s third-grade class. At the start of the study he performed at the 73rd percentile on the M-Comp and at the 65th percentile on the M-CAP. Student 8 earned an initial baseline mean score of 25.70 (5.38), which falls in the instructional range. He was present for all sessions throughout the course of the study. He began the peer tutoring intervention during the second week of the study and he began the combined intervention during the fourth week of the study.
Student 9 was an 8-year old-male student of Caucasian descent. He did not qualify to receive free- or reduced- lunch. Student 9 receives instruction in Mrs. C’s third-grade class. At the start of the study he performed at the 40th percentile on the M-Comp and at the 18th percentile on the M-CAP. Student 9 earned an initial baseline mean score of 18.20 (5.77), which falls in the instructional range. He missed one session throughout the course of the study. This was during the combined intervention. Student 9 began the peer tutoring intervention during the second week of the study and he began the combined intervention during the fourth week of the study.

Student 10 was an 8-year-old male student of Caucasian descent. He did not qualify to receive free- or reduced- lunch. Student 10 receives instruction in Mr. L’s combined second- and third-grade class. At the start of the study he performed at the 24th percentile on the M-Comp and at the 38th percentile on the M-CAP. Student 10 earned an initial baseline mean score of 20.05 (4.46), which falls in the instructional range. He missed two sessions throughout the course of the study. Both absences were during the first week of data collection during the initial baseline phase. Student 10 began the peer tutoring intervention during the second week of the study and he began the combined intervention during the fourth week of the study.

Student 11 was an 8-year old female student of Caucasian descent. She did not qualify to receive free- or reduced- lunch. Student 11 receives instruction in Mrs. C’s third-grade class. At the start of the study she performed at the 35th percentile on the M-Comp and at the 43rd percentile on the M-CAP. Student 11 earned an initial baseline mean score of 21.10 (4.41). She missed two sessions throughout the course of the study. The first session she missed was during the peer tutoring intervention, while the second
was during the combined intervention. Student 11 began the peer tutoring intervention during the second week of the study and she began the combined intervention during the fourth week of the study.

Student 12 was a 9-year-old male student of Caucasian descent. He did not qualify to receive free- or reduced- lunch. Student 12 receives instruction in Mrs. C’s third-grade class. At the start of the study he performed at the 55th percentile on the M-Comp and at the 14th percentile on the M-CAP. Student 12 earned an initial baseline mean score of 19.07 (6.03), which falls in the instructional range. He was present for all sessions throughout the course of the study. Student 12 began the peer tutoring intervention during the third week of the study and he began the combined intervention during the sixth and final week of the study.

Student 13 was a 9-year-old female student of Caucasian descent. She qualifies to receive free lunch. Student 13 receives instruction in Mr. L’s combined second- and third-grade class. At the start of the study she performed at the 37th percentile on the M-Comp and at the 10th percentile on the M-CAP. Student 13 earned an initial baseline mean score of 14.26 (4.34), which falls in the instructional range. Throughout the course of the study she missed only the last session, which was during the combined intervention phase. Student 13 began the peer tutoring intervention during the third week of the study and she began the combined intervention during the sixth and final week of the study.

Student 14 was a 9-year-old male student of Caucasian descent. He qualifies to receive free lunch. Student 14 receives instruction in Mrs. C’s third-grade class. At the start of the study he performed at the 5th percentile on the M-Comp and at the 52nd percentile on the M-CAP. Student 14 earned an initial baseline mean score of 10.77
(3.45), which falls in the frustrational range. Student 14 was present for all sessions throughout the course of the study. He began the peer tutoring intervention during the second week of the study and he began the combined intervention during the fourth week of the study.

Student 15 was an 8-year-old female student of Caucasian descent. She did not qualify to receive free- or reduced- lunch. Student 15 receives instruction in Mr. L’s combined second- and third-grade class. At the start of the study she performed at the 17th percentile on the M-Comp and at the 5th percentile on the M-CAP. Student 15 earned an initial baseline mean score of 12.88 (3.10), which falls in the frustrational range. Student 15 missed two sessions throughout the course of the study. The first was during the initial baseline phase, while the second was during the combined intervention. She began the peer tutoring intervention during the second week of the study and she began the combined intervention during the fourth week of the study.

Student 16 was an 8-year-old male student of Caucasian descent. He did not qualify to receive free- or reduced- lunch. Student 16 receives instruction in Mrs. C’s third-grade class. At the start of the study he performed at the 18th percentile on the M-Comp and at the 14th percentile on the M-CAP. Student 16 earned an initial baseline mean score of 10.00 (4.44), which falls in the frustrational range. Student 16 missed three sessions throughout the course of the study. The first was during the initial baseline phase, the second was during the peer tutoring intervention, and the third was during the combined intervention. He began the peer tutoring intervention during the second week of the study and he began the combined intervention during the fourth week of the study.
Student 17 was a 9-year-old female student of Caucasian descent. She qualifies to receive free lunch. Student 17 receives instruction in Mrs. C’s third-grade class. At the start of the study she performed at the 2^{nd} percentile on the M-Comp and at the 23^{rd} percentile on the M-CAP. Student 17 earned an initial baseline mean score of 12.63 (2.87), which falls in the frustrational range. Student 17 was present for all sessions throughout the course of the study. She began the peer tutoring intervention during the third week of the study and she began the combined intervention during the fifth week of the study.

Student 18 was an 8-year-old male student of Caucasian descent. He did not qualify to receive free- or reduced- lunch. Student 18 receives instruction in Mr. L’s combined second- and third-grade class. At the start of the study he performed at the 4^{th} percentile on the M-Comp and at the 27^{th} percentile on the M-CAP. Student 18 earned an initial baseline mean score of 10.53 (3.31), which falls in the frustrational range. Student 18 was present for all sessions throughout the course of the study. He began the peer tutoring intervention during the third week of the study and she began the combined intervention during the fifth week of the study.

Measures

There are two types of curriculum based measures (CBM) that were used to assess student performance: general outcome measures (GOM) and single-skill mastery probes. GOM probes were used to assess for generalization of these basic skills interventions to more complex tasks. Two different AIMSweb probes were used for this purpose. One measure is the Mathematics Computations (M-COMP) probe, which assesses students’ ability to complete complex computational skills. The other is the Mathematics Concepts and Applications (M-CAP) probe, which assesses students’ ability
to complete concepts and applications problems. Single-skill mastery probes were created to assess ongoing student performance on single-digit multiplication facts as students progressed through the intervention phases of this study.

**AIMSweb Progress Monitoring System.** AIMSweb is a web-based assessment, data management, and reporting system based on direct, frequent, and continuous student assessment designed to support the tiered assessment and instruction that is seen within the framework of RTI (Pearson, 2012a). It provides CBM assessments and procedures for the areas of reading, writing, spelling, and mathematics. AIMSweb assessments are general outcome measures (GOM) that can be used for the purposes of universal screening, progress monitoring, and program evaluation (Pearson, 2012d).

National norms have been reported for all AIMSweb measures. These norms have been stratified by gender, ethnicity, and socioeconomic status (Pearson, 2012d). Data were pulled from the AIMSweb database, but only schools that do universal screening were selected to ensure that an accurate representation of all ability levels was present in the sample (Person, 2012d). Thus, schools that only screen students who have been identified for special education services or as being in need of intervention were excluded from the norm sample. For third grade, there were 18,953 students included in the national norm for the M-COMP assessments and 11,070 students for the M-CAP assessments. The following subsections describe these measures and their psychometric properties as they were used to test for generalization in the current study.

**AIMSweb Mathematics Computation (M-COMP) probes.** AIMSweb provides M-COMP probes which contain computational math facts that are part of the typical curriculum for grades one through eight, with 33 alternative forms for progress
monitoring for each grade level (Pearson, 2012b). There are also three universal assessments for each grade. These probes are brief, standardized tests that require students to write their answers within an eight-minute time period. They were created based on recommendations from experts in mathematics curriculum (Pearson, 2012d). The third grade probes assess skills regarding column addition, basic facts, and complex computation.

The M-COMP measure is fairly new, as it replaced another AIMSweb computation assessment known as the mathematics curriculum based measure (M-CBM) in the fall of 2011 (Pearson, 2012c). This revision was necessary to address several complaints (Pearson, 2010). First, it was reformatted so that easier problems came first in the assessment in order to prevent students from skipping more difficult items. Second, items were numbered, more space was put between problems, and a box was placed around each item to make it easier for students to navigate the probes. Finally, scoring procedures were changed because educators complained about the length of time it took to count the number of digits correct for each assessment (Pearson, 2010). Scores are now calculated using an answer sheet provided by AIMSweb (Pearson, 2012b). Scores are weighted based on item difficulty, with points ranging from one to three for each individual item.

Person (2012d) conducted a national field test in order to assess the reliability and validity of the M-COMP. This sample included 7,703 subjects in grades two through eight, with sample size ranging from 916 to 1,048 across grades. Demographic information was reported at each grade level for race, gender, income level, and geographic location, indicating that a diverse sample was used (Pearson, 2012d).
Alternate form reliability of newly designed M-COMP probes was assessed for each grade, with correlations ranging from 0.85 to 0.90, suggesting high alternate forms reliability (Pearson, 2012d). Criterion validity was assessed by comparing scores on the M-COMP to scores on the Group Mathematics Assessment and Diagnostic Evaluation (G–MADE) in the first, third, and eighth grades. The sample sizes varied with 98 subjects for grades one and three and 54 for grade eight. Correlations were high and ranged from .74 to .84, indicating that the M-COMP appears to measure the intended computational constructs (Pearson, 2012d).

AIMSweb Mathematics Concepts and Applications (M-CAP). AIMSweb M-CAP assessments are a series of tests of short duration that measure the general problem-solving skills in mathematics for students in grades two through eight. Administration time is eight to ten minutes, depending on the grade level being assessed (Pearson, 2012b). Domains assessed on the M-CAP probes are consistent with the content areas recommended by the National Council of Teachers of Mathematics (National Council of Teachers of Mathematics, 2006). In third grade, the assessment content areas include understanding of facts, fractions, and geometry (Pearson, 2012b). For each grade level, there are 33 equivalent probes that can be used for progress monitoring as well as three benchmark assessments.

During the test administration, students were given a test booklet and an answer sheet. In order for an item to be scored correctly, the student must write the correct answer for the item. Test items are worth 1, 2, or 3 points, depending on the difficulty level of the item and no partial credit is given. Scores for the M-CAP probes are based on the total number of points a student earned within an 8-minute time period.
Person (2012d) conducted a national field test in order to assess the reliability and validity of the M-CAP. This sample included 6,550 subjects in grades two through eight, with sample sizes ranging from 858 students to 1,064 across grades. Demographic information was reported at each grade level for race, gender, income level, and geographic location, indicating a diverse sample was used (Pearson, 2012d). Alternate form reliability of the M-CAP probes was assessed for each grade, and they ranged from 0.80 to 0.88, suggesting high alternate forms reliability (Pearson, 2012d). Interrater reliability was also calculated between two scorers who were provided the M-CAP administration and scoring manual as a form of training. Scoring was completed for 60 cases that were randomly selected from each grade in the national field sample. Interrater reliability was extremely high, with correlations ranging from .98 to .99 (Pearson, 2012d).

Criterion validity was assessed by comparing scores on the M-CAP to scores on the North Carolina and Illinois end-of-year state mathematics tests using data from AIMSweb users for the 2009–2010 school year within these states (Pearson, 2012d). Sample sizes varied across grade and time of year (fall, winter, spring) and ranged from 553 to 959 students. Correlations were adjusted for range restriction, using the national norm sample as the reference group. Pearson (2012d) indicated that M-CAP scores correlate with state math tests for grades three through eight at a consistent level, with correlations largely in the .60’s and low .70’s.

Overall, the AIMSweb probes are a reliable and valid measure of students’ mathematics ability. These probes were used to assess for generalization of basic skills to more complex computation and applications problems in this study.
Single-skill CBM probes. Single-skill CBM probes were used to assess student progress in the targeted skill area. Three sets of single-skill probes were created in order to compare student progress on items that are presented in the baseline condition, the tutor condition, and the tutee condition. These sets were also administered during the performance feedback intervention so that scores can be compared across phases of this study. Each set of probes included 10 mutually exclusive single-digit multiplication problems that were initially developed by Skinner et al. (1989). The problems within these sets included numbers 2 through 9 and were equated in several ways (see Table 4 below for the full lists of problems). Each math problem contains a single digit number times a single digit number and equates to a double digit answer. Additionally, no problem is repeated within or across sets. Finally, the number of times the digits 2 through 9 are used does not differ by more than one across sets.

In order to track student progress throughout the baseline and intervention sessions, 30 unique probes were created for each set for each of the baseline and intervention sessions throughout this study (Appendix C). This was done according to procedures used by Coddington et al. (2006) and recommended by Shapiro (2011). For short-term monitoring, Shapiro suggests that the educator decide on the specific academic skill to monitor. In this case, single-digit multiplication has been selected. Next, it was necessary that each probe contain a sufficient number of randomly selected problems. In this study, there were 10 problems for each set. Thus, for each probe, problems were listed in forward and reverse order (e.g., 2 x 5 =, 5 x 2 =), resulting in 20 unique facts per probe. In order to ensure there were enough problems per probe, each problem was presented twice, which means each probe contained 40 math problems.
Table 4

*Skinner et al.’s Equated Sets of Multiplication Problems*

<table>
<thead>
<tr>
<th>List A</th>
<th>List B</th>
<th>List C</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 x 6 = 12</td>
<td>2 x 9 = 18</td>
<td>2 x 5 = 10</td>
</tr>
<tr>
<td>2 x 7 = 14</td>
<td>3 x 6 = 18</td>
<td>3 x 4 = 12</td>
</tr>
<tr>
<td>3 x 5 = 15</td>
<td>3 x 7 = 21</td>
<td>3 x 9 = 27</td>
</tr>
<tr>
<td>3 x 8 = 24</td>
<td>4 x 4 = 16</td>
<td>4 x 7 = 28</td>
</tr>
<tr>
<td>4 x 6 = 24</td>
<td>4 x 5 = 20</td>
<td>5 x 5 = 25</td>
</tr>
<tr>
<td>4 x 9 = 36</td>
<td>4 x 8 = 32</td>
<td>5 x 7 = 35</td>
</tr>
<tr>
<td>5 x 6 = 30</td>
<td>5 x 8 = 40</td>
<td>6 x 6 = 36</td>
</tr>
<tr>
<td>5 x 9 = 45</td>
<td>6 x 7 = 42</td>
<td>6 x 8 = 48</td>
</tr>
<tr>
<td>7 x 7 = 49</td>
<td>7 x 9 = 63</td>
<td>6 x 9 = 54</td>
</tr>
<tr>
<td>8 x 9 = 72</td>
<td>9 x 9 = 81</td>
<td>7 x 8 = 56</td>
</tr>
</tbody>
</table>


Shapiro (2011) notes that there are several websites that can provide an appropriate template for CBM probes. Codding et al. (2006) used [www.interventioncentral.org](http://www.interventioncentral.org) to generate the template for the probes they used in their study and this website was also be used for the current study. To ensure that problems were randomly presented in each probe, Codding et al. used a random number generator when creating problems for each probe. However, because the sets used in this study are mutually exclusive, this process was changed slightly. Each problem was assigned a number one through 20 (e.g., From List A, 1 was assigned to 2 x 6 = 12, 2 was assigned
to $2 \times 7 = 14$). These numbers were entered into the random number generator at www.random.org. The problems for each probe were put in the order generated by this website. The results from these probes were used to chart student progress throughout the baseline and intervention phases of this study.

**Treatment acceptability measures.** In order to assess teachers’ and students’ opinions regarding the intervention procedures two established acceptability scales were modified for the purposes of this study. It is not uncommon for researchers to modify existing acceptability scales to meet the particular needs of their studies (e.g., Codding et al., 2007; Mautone et al., 2009).

*CBM Acceptability Scale (CBM – AS).* A modified version of CBM Acceptability Scale (CBM-AS) was created to assess treatment acceptability in this study. The CBM – AS is a 20-item Likert-type scale developed to assess teacher acceptability of CBM (Oats & Allinder, 1995; as cited by Allinder & Oats, 1997). It uses a six-point scale, ranging from 1 (strongly disagree) to 6 (strongly agree), to assess teacher opinions regarding the effectiveness of CBM, the time required to use it, and amount of skill or training necessary to understand it. Good internal consistency reliability (Cronbach's alpha = .90) of the CBM-AS has been reported (Oats & Allinder, 1995; as cited by Allinder & Oats, 1997). This scale was adapted to ask the same questions regarding the use of the three interventions in this study. Additionally, it was proposed that the teacher would be asked to provide anecdotal information regarding her opinions about these interventions (see Appendix D-1, D-2, D-3, and D-4 for the teacher acceptability measures).
**Kids Intervention Profile (KIP).** The Kids Intervention Profile (KIP) is a treatment acceptability measure designed to assess students’ opinions regarding academic interventions (Eckert, Codding, Hier, Sullivan, & Malandrino, in preparation). Eckert et al. reported excellent internal consistency reliability (Chronbach’s alpha = .93) of the KIP. For the current study this measure was adapted for use with each of the three interventions (see appendices E-1, E-2, and E-3 for student acceptability scales).

On the KIP, students provide answers to questions by filling in boxes on a rating scale where the size of the boxes corresponded to their preference ranging from not at all (a very small box) to very, very much (a very large box), or never (again, a very small box) to many, many times (again, a very large box). Therefore, consistently small boxes indicated disagreement with questions while the large boxes indicated stronger agreement. For scoring purposes, the boxes were quantified so that individual item scores ranged from 1 being not at all or never (i.e., smallest box) to 5 being very, very much or many, many times (i.e., largest box) for most items. Scoring was reversed for two questions regarding each intervention because for these questions a small box indicated higher rates of acceptability and a large box indicated higher rates.

With the measures for the study in place, procedures for implementing them will now be described.

**Procedures**

Prior to data collection and intervention phases, a research assistant was trained. Probes were created, and pre-tests were administered. The study began with a baseline phase, followed by an individual intervention phase in which students received one of two interventions, either a performance feedback or a peer tutoring intervention.
Following the individual intervention phase, a combined intervention phase began in which the interventions were combined into one treatment package. Following each intervention phase students completed acceptability scales to assess their feelings about the intervention. At the end of the study, post-tests were administered to assess for generalizability. These procedures are described in depth in the remainder of this chapter.

**Recruitment and training of research assistant.** The PI received help from a research assistant from a local college. All training procedures are described in detail in Appendix F. The research assistant was required to pass a fingerprinting and background check to be allowed in the school. The training included taking the Undergraduate Citi Human Subjects Training Course and being trained on the standardized procedures for creating and scoring assessment instruments, on how to use the treatment integrity checks and on how to score probes. For the study, she conducted treatment integrity checks for 40% of the sessions and scored probes so that inter-rater reliability could be assessed. Although it was originally proposed that Ms. Parmenter would be trained to administer intervention procedures, scheduling did not allow for that so she was not trained in that area. Additionally, Ms. Parmenter did not calculate interscorer agreement or treatment fidelity percentages, as the PI completed those tasks.

**Experimental design.** A single-case, multiple baseline design with an embedded alternating treatment design was used for this study (Richards, Taylor, Ramasamy, & Richards, 1999). Interventions were implemented approximately five times per week for six weeks. Each intervention session lasted approximately 15-20 minutes. At the onset of this study, students were randomly assigned to one of two groups (group A and group
B) based on their instructional levels. The goal was that four students in each group would be performing in the frustrational range and four students in each group would be performing in the instructional range. However, the sample included 22 students, so 10 students were in group A with five students at each instructional level and 12 students were in group B with six students at each instructional level. Due to attrition, results are reported based on 9 students in group A (four at the instructional level and five at the frustrational level) and 11 students in group B (five at the instructional level and six at the frustrational level). Initially, group A received the performance feedback intervention, while group B received the peer tutoring intervention.

After five baseline sessions, the intervention phase began for some students. One student at each instructional level in group A received the performance feedback intervention, while one dyad at each instructional level in group B received the peer-tutoring intervention. All other students continued with baseline procedures. After five more sessions, another student from each instructional level in group A began the performance feedback intervention, while the remaining dyads in group B began the peer tutoring intervention. This procedure was repeated for group A until all students completed the performance feedback intervention. Students in group A progressed from the performance feedback intervention to the combined intervention, while dyads in group B progressed from the peer tutoring intervention to the combined intervention. Students were moved at specified, staggered times throughout the study (see Tables 5 and 6 for an explicit view of the research design).

Within the peer tutoring intervention, students were paired up and they took turns serving the function of tutor and tutee. They swapped roles after every five sessions.
This study sought to assess whether these roles differ in their effectiveness of improving fluency scores as compared to a control condition. Thus, an alternating treatments design (ATD) has been imbedded within the context of the larger study.

Table 6

Proposed Multiple Baseline Research Design

<table>
<thead>
<tr>
<th>Group A</th>
<th>5 Sessions</th>
<th>5 Sessions</th>
<th>5 Sessions</th>
<th>5 Sessions</th>
<th>5 Sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructional #1 (AI-1)</td>
<td>Baseline</td>
<td>PF</td>
<td>PF</td>
<td>Combined</td>
<td>Combined</td>
</tr>
<tr>
<td>Instructional #2 (AI-2)</td>
<td>Baseline</td>
<td>PF</td>
<td>PF</td>
<td>Combined</td>
<td>Combined</td>
</tr>
<tr>
<td>Instructional #3 (AI-3)</td>
<td>Baseline</td>
<td>Baseline</td>
<td>PF</td>
<td>Combined</td>
<td>Combined</td>
</tr>
<tr>
<td>Instructional #4 (AI-4)</td>
<td>Baseline</td>
<td>Baseline</td>
<td>PF</td>
<td>Combined</td>
<td>Combined</td>
</tr>
<tr>
<td>Frustrational #1 (AF-1)</td>
<td>Baseline</td>
<td>PF</td>
<td>PF</td>
<td>Combined</td>
<td>Combined</td>
</tr>
<tr>
<td>Frustrational #2 (AF-2)</td>
<td>Baseline</td>
<td>PF</td>
<td>PF</td>
<td>Combined</td>
<td>Combined</td>
</tr>
<tr>
<td>Frustrational #3 (AF-3)</td>
<td>Baseline</td>
<td>Baseline</td>
<td>PF</td>
<td>Combined</td>
<td>Combined</td>
</tr>
<tr>
<td>Frustrational #4 (AF-4)</td>
<td>Baseline</td>
<td>Baseline</td>
<td>Baseline</td>
<td>PF</td>
<td>PF</td>
</tr>
</tbody>
</table>

Note: During the combined intervention: (AI-1 paired with AI-2); (AI-3 paired with AI-4); (AF-1 paired with AF-2); (AF-3 paired with AF-4)

Group B

| Instructional #1 (BI-1)   | Baseline   | Tutoring   | Tutoring   | Combined   | Combined   | Combined   |
| Instructional #2 (BI-2)   | Baseline   | Tutoring   | Tutoring   | Combined   | Combined   | Combined   |
| Instructional #3 (BI-3)   | Baseline   | Baseline   | Tutoring   | Tutoring   | Combined   | Combined   |
| Instructional #4 (BI-4)   | Baseline   | Baseline   | Tutoring   | Tutoring   | Combined   | Combined   |
| Frustrational #1 (BF-1)   | Baseline   | Tutoring   | Tutoring   | Combined   | Combined   | Combined   |
| Frustrational #2 (BF-2)   | Baseline   | Tutoring   | Tutoring   | Combined   | Combined   | Combined   |
| Frustrational #3 (BF-3)   | Baseline   | Baseline   | Tutoring   | Tutoring   | Combined   | Combined   |
| Frustrational #4 (BF-4)   | Baseline   | Baseline   | Tutoring   | Tutoring   | Combined   | Combined   |

Note: During the Tutoring and Combined interventions: (BI-1 paired with BI-2); (BI-3 paired with BI-4); (BF-1 paired with BF-2); (BF-3 paired with BF-4)

Note: PF = Performance Feedback; Tutoring = Reciprocal Peer Tutoring Intervention; Combined = PF + Tutoring interventions.
Table 6

*Actual Multiple Baseline Research Design*

<table>
<thead>
<tr>
<th>Group A</th>
<th>5 Sessions</th>
<th>5 Sessions</th>
<th>5 Sessions</th>
<th>5 Sessions</th>
<th>5 Sessions</th>
<th>5 Sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student 1 (AI-1)</td>
<td>Baseline</td>
<td>PF</td>
<td>PF</td>
<td>Combined</td>
<td>Combined</td>
<td>Combined</td>
</tr>
<tr>
<td>Student 2 (AI-2)</td>
<td>Baseline</td>
<td>Baseline</td>
<td>Baseline</td>
<td>PF</td>
<td>PF</td>
<td>Combined</td>
</tr>
<tr>
<td>Student 3 (AI-3)</td>
<td>Baseline</td>
<td>PF</td>
<td>PF</td>
<td>Combined</td>
<td>Combined</td>
<td>Combined</td>
</tr>
<tr>
<td>Student 4 (AI-4)</td>
<td>Baseline</td>
<td>PF</td>
<td>PF</td>
<td>Combined</td>
<td>Combined</td>
<td>Combined</td>
</tr>
<tr>
<td>Student 5 (AF-1)</td>
<td>Baseline</td>
<td>Baseline</td>
<td>PF</td>
<td>Combined</td>
<td>Combined</td>
<td>Combined</td>
</tr>
<tr>
<td>Student 6 (AF-2)</td>
<td>Baseline</td>
<td>Baseline</td>
<td>Baseline</td>
<td>PF</td>
<td>PF</td>
<td>Combined</td>
</tr>
<tr>
<td>Student 7 (AF-3)</td>
<td>Baseline</td>
<td>Baseline</td>
<td>Baseline</td>
<td>Baseline</td>
<td>PF</td>
<td>Combined</td>
</tr>
</tbody>
</table>

*Note:* During the combined intervention: (AI-1 paired with AI-2); (AI-3 paired with AF-3); (AI-4 paired with outside student.); (AF-1 paired with AF-2); (AF-4 paired with AF-5)

<table>
<thead>
<tr>
<th>Group B</th>
<th>5 Sessions</th>
<th>5 Sessions</th>
<th>5 Sessions</th>
<th>5 Sessions</th>
<th>5 Sessions</th>
<th>5 Sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student 8 (BI-1)</td>
<td>Baseline</td>
<td>Tutoring</td>
<td>Tutoring</td>
<td>Combined</td>
<td>Combined</td>
<td>Combined</td>
</tr>
<tr>
<td>Student 9 (BI-2)</td>
<td>Baseline</td>
<td>Tutoring</td>
<td>Tutoring</td>
<td>Combined</td>
<td>Combined</td>
<td>Combined</td>
</tr>
<tr>
<td>Student 10 (BI-3)</td>
<td>Baseline</td>
<td>Tutoring</td>
<td>Tutoring</td>
<td>Combined</td>
<td>Combined</td>
<td>Combined</td>
</tr>
<tr>
<td>Student 11 (BI-4)</td>
<td>Baseline</td>
<td>Baseline</td>
<td>Tutoring</td>
<td>Tutoring</td>
<td>Tutoring</td>
<td>Combined</td>
</tr>
<tr>
<td>Student 12 (BI-5)</td>
<td>Baseline</td>
<td>Baseline</td>
<td>Tutoring</td>
<td>Tutoring</td>
<td>Tutoring</td>
<td>Combined</td>
</tr>
<tr>
<td>Student 13 (BI-6)</td>
<td>Baseline</td>
<td>Baseline</td>
<td>Tutoring</td>
<td>Tutoring</td>
<td>Tutoring</td>
<td>Combined</td>
</tr>
<tr>
<td>Student 14 (BF-1)</td>
<td>Baseline</td>
<td>Tutoring</td>
<td>Tutoring</td>
<td>Combined</td>
<td>Combined</td>
<td>Combined</td>
</tr>
<tr>
<td>Student 15 (BF-2)</td>
<td>Baseline</td>
<td>Tutoring</td>
<td>Tutoring</td>
<td>Combined</td>
<td>Combined</td>
<td>Combined</td>
</tr>
<tr>
<td>Student 16 (BF-3)</td>
<td>Baseline</td>
<td>Tutoring</td>
<td>Tutoring</td>
<td>Combined</td>
<td>Combined</td>
<td>Combined</td>
</tr>
<tr>
<td>Student 17 (BF-4)</td>
<td>Baseline</td>
<td>Baseline</td>
<td>Tutoring</td>
<td>Tutoring</td>
<td>Combined</td>
<td>Combined</td>
</tr>
<tr>
<td>Student 18 (BF-5)</td>
<td>Baseline</td>
<td>Baseline</td>
<td>Tutoring</td>
<td>Tutoring</td>
<td>Combined</td>
<td>Combined</td>
</tr>
</tbody>
</table>

During the Tutoring and Combined interventions: (BI-1 paired with BI-2); (BI-4 paired with outside student.); (BI-4 paired with BI-5); (BF-1 paired with BF-2); (BF-3 paired with BF-4); (BF-5 paired with BF-6)

*Note:* PF = Performance Feedback; Tutoring = Reciprocal Peer Tutoring Intervention; Combined = PF + Tutoring interventions; Outside Student = results are not reported.
**Baseline.** Baseline performance was assessed for each participant through the use of one single-skill CBM probe from each set during each baseline session. As is recommended by Richards et al. (1999), baseline data was collected for five consecutive sessions. The student was read directions from the CBM generator at www.interventioncentral.org. These directions required students to complete as many math facts as possible during a two-minute period. The experimenter later calculated the number of Digits Correct per Minute (DCPM) and the number of errors for each probe and an initial baseline mean score was calculated. Initial baseline mean scores were calculated using scores from all three sets (A, B, and C) across the first five sessions of the baseline phase. Students were considered to be performing at the instructional level if they earned an initial baseline mean score that fell within the range of 14-31 DCPM and they were considered to be in the mastery range if they earned an initial baseline score that fell below 14 DCPM (Burns et al., 2006). Student levels were determined for this study based on these initial baseline means so that students could be evenly grouped based upon their ability level when intervention sessions began.

**Performance feedback intervention.** Procedures were identical to those described for the baseline condition. However, prior to beginning the single-skill CBM assessments, students received written notice of their previous highest score (Figure 1). Scores were written on the back of a post-it, which was attached to the back of each packet. When students directions were read by the PI, students were reminded that they were should attempt to beat their previous high score. These procedures were copied from those used in by Rhymer et al. (2000).
Performance Feedback Intervention

Peer tutoring intervention. This study employed peer-tutoring procedures that were similar to those used by Rhymer et al. (2000) (Figure 2). Students were grouped into dyads. Each condition (tutor, tutee, and control) was assigned a different set of problems in order to assess student progress on items that were not part of the peer-tutoring intervention. Index cards for the tutoring intervention were color-coded. As a class-wide peer tutoring procedure was used, procedures differed slightly from those used by Rhymer et al. In the present study, instead of providing a chess clock to all dyads, the researcher had a stop watch that was set for two minutes. In this way, all student dyads began and ended the peer tutoring procedure together. Additionally, during the peer tutoring intervention, students held their role as either tutee or tutor for five consecutive sessions before swapping. More specifically, if a student began the peer tutoring intervention serving as the tutor, he/she would remain as the tutor for five sessions, after which the student would become the tutee and his/her partner would act as tutor for five
consecutive sessions. In this way, students never served as tutor and tutee during the same session, instead, they took turns on a weekly basis until the end of the study.

To begin the peer tutoring intervention, a student acted as the tutor and showed the tutee an index card from the prescribed set. If the tutee answered correctly, the tutor said, “Correct,” and then placed the index card in a pile. If the tutee answered incorrectly, the tutor said, “Incorrect,” provided the answer to the tutee who then wrote the problem with the correct answer on a sheet of scrap paper. The index card with the incorrectly answered problem was placed in a separate pile. This process was repeated for two minutes. Following these procedures, the tutors were asked to record the number of flash cards that were answered correctly for their partners. These records were collected at the end of each session by the researchers. At the start of every session, tutees were given their record sheets which allowed them to see how many flashcards they answered correctly in previous sessions. In this way, they received performance feedback regarding performance during the previous intervention sessions. Students were encouraged to complete more correct cards than their previously highest score in the upcoming session.

Immediately following this session, both students were allotted two minutes to complete a math probe that contained the set of problems that had just been practiced. Following this assessment, they were given two minutes to complete a probe that contained the set of problems that would be practiced when students swapped roles. Finally, both students had two minutes to complete a control set of problems that neither student had practiced. Two control sets were completed by the three students who served only one role (tutor or tutee) in the final week of the study.
Peer Tutoring Intervention

Figure 2. Peer tutoring intervention procedures (adapted from Rhymer et al., 2000).

Combined Intervention

Figure 3. Combined intervention procedures adapted from Rhymer et al. (2000).

**Combined intervention.** The peer tutoring procedures were implemented as described above, but now the assessment feedback provided in the performance feedback intervention was added. At the beginning of each session, students would look at a post-
it on the back of their assessment packet that contained their highest score to date on any of the assessments. All students were encouraged to beat their highest score.

**Testing for generalization.** In order to assess for generalization, students completed an M-COMP and an M-CAP probe following the initial baseline session. Following the last intervention phase, students completed another M-COMP and another M-CAP probe. These scores were compared to mean scores during the first five sessions of the baseline phase and mean scores during the last five sessions of the final intervention phase.

**Procedural integrity.** Procedural integrity was ensured in two ways and based on the procedures used by Codding et al. (2007). According to Codding et al., researchers had to follow and complete a checklist of each step of the procedure during each intervention phase to ensure that all steps are followed. Second, during 40% of sessions an independent observer completed a procedural checklist for the interventions (see Appendix G, G-1, G-2, and G-3, for complete experimenter checklists). A checklist that is identical to the protocols used by the experimenters was created for this study to assess procedural integrity for each intervention. The checklist contained the following: (a) the materials required, (b) CBM probe procedures, and (c) the intervention procedures. Each checklist also included the scripts that the experimenters were to follow. The independent observer recorded a checkmark for the presence of the required materials. A checkmark was also recorded when each step of the procedures and scripts were correctly implemented by the experimenter. The number of steps checked by the independent observer was divided by the total number of steps listed for the procedure and then multiplied by 100.
Although the proposed study also sought to assess procedural integrity as it applies to participants, only one research assistant was available for 40% of the study sessions. Thus, there was not enough time and resources to conduct this type of observation or to implement the experimenter’s integrity checklist. Checklists for student procedures were created in which the experimenter was to check off a box for each step the students completed correctly. The number of steps checked by the independent observer would be divided by the total number of steps listed for the procedure and then multiplied by 100. If students failed to meet 100% procedural integrity during the session, specific instructions regarding the mistakes would have been provided to the whole group prior to the next session. Although they were not used, the checklists are included in Appendices G-4 and G-5.

**Treatment acceptability.** As previously mentioned, the CBM – AS and KIP scales were modified for the purposes of assessing teacher’s and student’s opinions of the current interventions. Following each treatment phase, students completed the corresponding scale to provide information regarding how much they enjoyed the intervention and how helpful they found it to be. Mean scores and standard deviations were calculated for each separate item and for each of the three scales. Results of the surveys were analyzed and discussed in order to understand the aspects of the interventions that were enjoyable or seen as useful by students.

Although it was proposed that the teachers would complete the CBM – AS regarding their opinions about the interventions, teachers opted not to be present during the intervention procedures and were therefore unable to offer an opinion regarding the interventions used in this study. Thus, only student acceptability was assessed.
Chapter 4

Results

This chapter begins with a description of the data collection procedures. A central component of the study was student skill level. Therefore, procedures that were used to identify student skill level are also included in the description of data collection. This is followed by a brief description of the findings related to treatment integrity.

What follows is a section that details the statistical analyses used for this study including the various statistical methods used to analyze the data generated from this study. There were three tools used: calculation of slopes and means scores of performance, which were paired with visual analysis; Tau-U analysis, to measure the strength of effect size; and independent t-tests to compare groups and the impact of the various interventions. The findings from these analyses will be presented as well.

The next section of this chapter synthesizes the data in response to the research questions. First, there is a discussion about the impact of students’ socioeconomic status (SES) on the utility of these types of interventions, followed by a section detailing the findings related to students’ acceptability of the intervention procedures. Third, the generalizability of the skills acquired within the context of this intervention will be discussed.

In the last section, the findings that shed light on the main research questions will be presented. That is, comparisons between interventions across students will be relayed in four distinct areas. The first of which is a comparison between the performance feedback intervention and the peer tutoring intervention, while the second area compared the effectiveness of a combined intervention that was comprised of both individual
interventions to either of the interventions on their own. Third, a comparison between the results of the tutee and tutor conditions as part of the peer tutoring as well as combined interventions will be presented. Finally, individual interventions and overall improvement throughout the course of the study will be discussed within the framework of instructional levels. To answer each of these questions, visual analysis, Tau-U analysis, and independent t-tests were used to make comparisons.

The chapter will end with a summary of the relevant findings. Implications and suggestions for further research will be addressed in Chapter 5.

**Data Collection**

In order to obtain fluency measures of performance over time, third-grade students completed daily multiplication worksheets over the course of 30 sessions, which yielded DCPM scores. DCPM scores represented how quickly and accurately students perform arithmetic problems, multiplication in this case. Following the procedures used by Codding et al. (2006, 2007), students were allotted 2 minutes to complete assessment probes after which each correct digit within each answer was counted. This score was then divided by 2 in order to calculate DCPM scores. There were several students, however, who completed assessment probes in less than the allotted time frame. For these students, completion times were documented and scores were prorated accordingly. All tests were scored by the principal investigator and the research assistant, and a digit-by-digit comparison was made. Discrepancies in scoring were noted and double-checked by the principal investigator and corrections were made as needed. Inter-rater scoring was calculated for 50% of the sessions and ranged from 89%-100% with a mean of 94% agreement, which is typical for such probes.
For this study, three mutually exclusive sets of problems, initially created by Skinner et al. (1989), were used to create probes. For each session, every participant completed one probe from each set (A, B, or C) and a score for each probe was calculated, resulting in three student scores per session. On one occasion, student 8 failed to turn to the correct page when the timer sounded and he worked on the same probe for 4 minutes instead of working on two probes for 2 minutes each. Hence, he only completed one of the three probes appropriately, resulting in only one valid score for that session.

**Determination of instructional level.** Fifteen scores from the first five sessions of the study were used to sort the participants into two groups. These scores were analyzed using the fluency criteria developed by Burns et al. (2006). If a student earned an average score between 14 - 31 DCPM for the first 15 probes, he/she was considered to be in the instructional group. If average scores were below 14 DCPM, they were considered to be in the frustrational group. As previously noted, students performing at the instructional level have acquired a skill, but they are unable to perform it with fluency and automaticity, while students performing at the frustrational are in the beginning stages of learning a skill and cannot independently complete a task with accuracy (Burns et al., 2006; Codding et al., 2007; Codding et al., 2010; Haring et al., 1977; Rhymer et al., 2000; Rivera & Bryant, 1992; Vanderheyden & Burns, 2005).

Students from each instructional group were randomly assigned to create two evenly distributed intervention groups. Due to attrition, however, the final group sizes were slightly uneven; but they did contain a minimum of three students, which is generally considered to be best practice for single-case research (Horner et al., 2005). The first group, labeled group A, received the performance feedback intervention
following baseline procedures, while the second group, labeled group B, received the peer tutoring intervention. Both groups participated in a third phase of the study with a combined intervention comprised of elements from both the performance feedback and peer tutoring interventions. Group A consisted of four participants performing at the instructional level and three performing at the frustrational level. In group B, 6 students performed at the instructional level, while 5 performed at the frustrational level.

During the peer tutoring and combined intervention phases, students alternated their role, as either tutor or tutee, every five sessions. Because students did not always participate in the different interventions for an even number of sessions, there was not always an even distribution of results with students in the different roles. A specific set of problems was assigned to each condition so that comparisons could be made across roles, meaning that if a student began tutoring with set A, they always tutored with set A. If a student was a tutee with set B, they were always the tutee with set B. Whichever set was not used for the tutee or tutor condition was used as a control condition. This allowed us to determine if one role of the peer tutoring dyad was more effective than another, and/or if either or both were more effective than no tutoring intervention at all.

**Treatment integrity.** As described in the previous chapter, procedural integrity was ensured in two ways, based on the procedures used by Codding et al. (2007). First, the principal researcher followed and completed a checklist of each step of the procedure during each intervention phase. This checklist was used for each intervention for every session. Results from this checklist indicated that all steps were completed with 96% accuracy.
Second, during 40% of sessions an independent observer completed a procedural checklist for the interventions (see Appendix A, A-1, A-2, A-3, A-4, and A-5 for complete checklists). A checklist that was identical to the protocols used by the experimenters was created for this study to assess procedural integrity for each intervention. The number of steps checked by the independent observer was divided by the total number of steps listed for the procedure and then multiplied by 100. Similar to the principal investigators treatment integrity, results from the independent observer indicated that the intervention was completed with 95% accuracy.

**Treatment acceptability.** This study sought to obtain teacher and student opinions about the current interventions; however, teachers opted not to be present throughout intervention procedures. Thus, only student acceptability measures could be used.

As previously noted, the KIP (Eckert, Codding, Hier, Sullivan, & Malandrino, in preparation) was used to assess intervention acceptability for students. Scores ranged from 1 to 5, with 1 meaning students did not positively endorse the item and 5 indicating a positive response.

The mean scores for the performance feedback and peer tutoring interventions were identical ($M = 4.08; SD = 1.18$) and both indicated that students generally enjoyed these procedures. The mean score for the combined intervention was slightly higher, but essentially the same ($M = 4.18; SD = 1.11$).

Additionally, average scores for each item of each rating scale were calculated. In general, students responded positively to all items, and average scores ranged from 2.6 (1.44) to 5.9 (.00), with most items earning an average score above 4. Across
interventions students typically reported that they felt their skills had improved as mean scores for this question were 4.11 (1.27), 5.00 (.00), and 4.70 (.80) for the performance feedback, peer tutoring, and combined intervention respectively. It is noteworthy that only two scores ($M = 2.75; SD = 1.36$, $M = 2.68; SD = 1.16$ for the peer tutoring and combined interventions, respectively) fell below 3, and both of these were with regard to the error-correction procedure. When students made a mistake during the peer tutoring procedures, they did not enjoy writing the math fact on a piece of paper. This is important because this part of the peer tutoring intervention was developed from error-correction procedures in which an instructor prompts a learner to elicit a correct response following the occurrence of an error by punishing that error (Worsdell et al., 2005). Thus, a slightly negative response with regard to this procedure was to be expected and desired.

In sum, student responses on the acceptability measure indicated that students found these interventions to be useful and they enjoyed participating in the procedures. **Visual Analysis of Individual Scores**

The students’ scores on the single-skill CBM probes were recorded following each intervention session and a graph of each student’s performance across interventions was created. Although students completed three probes during each session, data were only reported for the control set and for the set assigned to the implemented condition (tutor or tutee) for each session during the peer tutoring and combined intervention phases. This made it easier to make comparisons between the conditions. The graphs for each student in Figure 4 were analyzed visually by examining trends, levels, and variability to determine if the intervention had any effect on their math fluency levels.
The design of this study was complicated, as it combined a multiple baseline across subjects design with an alternating treatments design. This allowed for both intra-subject and inter-subject replication. Intra-subject replication refers to repeating the experimental effect with the same participant, while inter-subject replication refers to repeating the experimental effect with different participants (Gast, D. L. & Ledford, 2014). Both types of replication are necessary to ensure external and internal validity. Combining the two designs, however, made visual analysis of scores challenging, intervention effects are interpreted differently for each design.

When conducting single-case research, interventions can be deemed to be effective when prediction, verification, and replication occur (Richards et al., 1999). In a multiple baseline across subjects design, the baseline phase is initiated simultaneously across subjects. Then, an intervention is introduced to participants in a step-wise fashion. Effectiveness of an intervention (as well as experimental control) within this design is determined by the presence of a constant data path throughout the baseline phase before changing in a predictable direction upon implementation of the intervention. Verification occurs when each data point falls in the predicted direction, while replication of prediction and verification occurs when this pattern is observed across subjects (Richards et al., 1999).

For alternating treatments design, prediction, verification, and replication are assessed differently. Instead of using baseline performance to predict future performance, each data point in a condition serves as a predictor of future behavior within each condition. Each successive data point verifies the prediction of the preceding point, and replication occurs when trend continues in the predicted direction and
differentiation between interventions is established. Therefore, differentiation is used to confirm and describe the presence of experimental control. For this study, the baseline served to ensure that a stable trend was present and to ensure that problem sets A, B, and C were equally difficult for students. In order to make the findings meet the requirements for experimental control, one of the problem sets for each student was considered a control set and it was not targeted for intervention upon implementation of the tutee or tutor conditions.

Because both designs were used, it was necessary to interpret the data in various ways. As the performance feedback intervention was not introduced within the context of the alternating treatments design, data from this phase had to be compared to baseline levels of performance. However, when comparing the level of effectiveness of the tutee and tutor conditions of the peer tutoring and combined interventions, a different approach was used. First, it was necessary to determine if differentiation between conditions was present within each phase of the study. If scores for sets assigned to the tutee or tutor condition were higher than one another and/or from the control condition, differentiation between sets was present and verification of effectiveness was observed, independent of student performance in previous phases. Nonetheless, the goal of this study was for DCPM scores to improve over time, and comparisons were made between and within all phases. Finally, as replication across subjects is necessary for external validity, inter-subject replication was also assessed across all phases of this study.

It was also important to visually analyze the variability in each student’s performance over time. If an individual showed little variability in performance (i.e., fluency scores remain similar) or a stable trend (i.e., fluency scores are always
increasing), the more certain we could be that data points were indicative of the student’s true ability level (Richards et al., 1999). If stable scores and trend lines were observed, then fewer data points were needed to ascertain the interventions’ effectiveness.

In order to gain further understanding of individual student progress over time, means scores and standard deviations across phases were calculated by determining the average DCPM score of each set of problems for every phase of the study. Tables 7 and 9 display the mean DCPM scores for each student and set across all of the phases of the study. Standard deviations were also calculated. Finally, in order to assist with visual analysis and to determine the trend of student performance, individual slopes were calculated for each set of problems for each phase of the study (see Tables 8 and 10 for these scores).

Finally, visual analysis of Figure 5, which depicts the number of errors each student made over time, will be used to illustrate the accuracy with which students completed each multiplication probe. Errors were recorded as DICPM scores, which indicated how many mistakes each student made per minute.
Table 7

**Mean Scores across Intervention Settings for Group A**

<table>
<thead>
<tr>
<th>Group A: Performance Feedback Intervention</th>
<th>Condition</th>
<th>Baseline</th>
<th>Performance Feedback</th>
<th>Combined Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student 1</td>
<td>A – tutor</td>
<td>21.00 (8.00)</td>
<td>28.00 (0.00)</td>
<td>40.88 (14.09)</td>
</tr>
<tr>
<td>Instructional #1 (AI-1)</td>
<td>B – tutee</td>
<td>15.33 (1.15)</td>
<td>24.94 (4.46)</td>
<td>54.85 (11.09)</td>
</tr>
<tr>
<td></td>
<td>C – control</td>
<td>17.00 (6.24)</td>
<td>21.13 (5.22)</td>
<td>29.90 (13.40)</td>
</tr>
<tr>
<td>Student 2</td>
<td>A – tutee</td>
<td>26.04 (4.98)</td>
<td>38.10 (8.01)</td>
<td>55.60 (8.93)</td>
</tr>
<tr>
<td>Instructional #2 (AI-2)</td>
<td>B – control</td>
<td>22.25 (4.31)</td>
<td>25.25 (3.90)</td>
<td>29.90 (2.38)</td>
</tr>
<tr>
<td></td>
<td>C – control</td>
<td>20.46 (4.68)</td>
<td>23.00 (6.04)</td>
<td>28.10 (0.55)</td>
</tr>
<tr>
<td>Student 3</td>
<td>A – tutee</td>
<td>21.60 (4.45)</td>
<td>40.20 (4.49)</td>
<td>57.70 (3.33)</td>
</tr>
<tr>
<td>Instructional #3 (AI-3)</td>
<td>B – control</td>
<td>19.40 (0.89)</td>
<td>30.85 (4.11)</td>
<td>47.80 (7.94)</td>
</tr>
<tr>
<td></td>
<td>C – tutor</td>
<td>12.40 (3.03)</td>
<td>21.50 (5.87)</td>
<td>47.65 (9.19)</td>
</tr>
<tr>
<td>Student 4</td>
<td>A – tutor</td>
<td>18.50 (6.24)</td>
<td>26.70 (5.74)</td>
<td>34.55 (9.00)</td>
</tr>
<tr>
<td>Instructional #4 (AI-4)</td>
<td>B – control</td>
<td>13.40 (5.27)</td>
<td>18.30 (7.41)</td>
<td>27.33 (5.63)</td>
</tr>
<tr>
<td></td>
<td>C – tutee</td>
<td>13.80 (6.76)</td>
<td>12.50 (4.69)</td>
<td>29.30 (8.06)</td>
</tr>
<tr>
<td>Student 5</td>
<td>A – control</td>
<td>11.67 (3.75)</td>
<td>10.20 (2.95)</td>
<td>12.47 (4.13)</td>
</tr>
<tr>
<td>Frustrational #1 (AF-1)</td>
<td>B – tutee</td>
<td>6.56 (3.24)</td>
<td>8.80 (2.17)</td>
<td>9.70 (4.42)</td>
</tr>
<tr>
<td></td>
<td>C – tutor</td>
<td>6.61 (2.37)</td>
<td>8.60 (2.61)</td>
<td>9.90 (3.25)</td>
</tr>
<tr>
<td>Student 6</td>
<td>A – control</td>
<td>13.87 (4.29)</td>
<td>16.06 (5.77)</td>
<td>12.25 (7.80)</td>
</tr>
<tr>
<td>Frustrational #2 (AF-2)</td>
<td>B – control</td>
<td>8.80 (2.83)</td>
<td>8.81 (3.33)</td>
<td>9.13 (4.13)</td>
</tr>
<tr>
<td></td>
<td>C – tutee</td>
<td>10.47 (2.83)</td>
<td>9.58 (2.71)</td>
<td>14.63 (6.80)</td>
</tr>
<tr>
<td>Student 7</td>
<td>A – tutor</td>
<td>12.03 (3.16)</td>
<td>8.70 (3.30)</td>
<td>9.30 (4.09)</td>
</tr>
<tr>
<td>Frustrational #3 (AF-3)</td>
<td>B – control</td>
<td>8.13 (3.57)</td>
<td>7.40 (3.47)</td>
<td>7.00 (5.80)</td>
</tr>
<tr>
<td></td>
<td>C – control</td>
<td>9.97 (3.95)</td>
<td>8.60 (4.27)</td>
<td>8.20 (6.95)</td>
</tr>
</tbody>
</table>
Table 8

*Slopes across Intervention Settings for Group A*

<table>
<thead>
<tr>
<th>Condition</th>
<th>Baseline</th>
<th>Performance Feedback</th>
<th>Combined Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructional #1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(AI-1) Student 1</td>
<td>A – tutor</td>
<td>4.00</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>B – tutee</td>
<td>1.00</td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td>C – control</td>
<td>6.00</td>
<td>.30</td>
</tr>
<tr>
<td><strong>Student 2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructional #2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(AI-2) Student 2</td>
<td>A – tutee</td>
<td>.62</td>
<td>2.24</td>
</tr>
<tr>
<td></td>
<td>B – control</td>
<td>.24</td>
<td>-.16</td>
</tr>
<tr>
<td></td>
<td>C – control</td>
<td>.27</td>
<td>.05</td>
</tr>
<tr>
<td><strong>Student 3</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructional #3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(AI-3) Student 3</td>
<td>A – tutee</td>
<td>1.40</td>
<td>1.34</td>
</tr>
<tr>
<td></td>
<td>B – control</td>
<td>-.50</td>
<td>1.24</td>
</tr>
<tr>
<td></td>
<td>C – tutor</td>
<td>.90</td>
<td>1.51</td>
</tr>
<tr>
<td><strong>Student 4</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructional #4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(AI-4) Student 4</td>
<td>A – tutor</td>
<td>3.00</td>
<td>1.23</td>
</tr>
<tr>
<td></td>
<td>B – control</td>
<td>1.70</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>C – tutee</td>
<td>3.80</td>
<td>.35</td>
</tr>
<tr>
<td><strong>Student 5</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frustrational #1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(AF-1) Student 5</td>
<td>A – control</td>
<td>.17</td>
<td>-0.90</td>
</tr>
<tr>
<td></td>
<td>B – tutee</td>
<td>.10</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>C – tutor</td>
<td>.10</td>
<td>-1.20</td>
</tr>
<tr>
<td><strong>Student 6</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frustrational #2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(AF-2) Student 6</td>
<td>A – control</td>
<td>.28</td>
<td>-1.72</td>
</tr>
<tr>
<td></td>
<td>B – control</td>
<td>.07</td>
<td>-.63</td>
</tr>
<tr>
<td></td>
<td>C – tutee</td>
<td>-.16</td>
<td>-.12</td>
</tr>
<tr>
<td><strong>Student 7</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frustrational #3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(AF-3) Student 7</td>
<td>A – tuteor</td>
<td>-.15</td>
<td>-1.20</td>
</tr>
<tr>
<td></td>
<td>B – control</td>
<td>.04</td>
<td>-.70</td>
</tr>
<tr>
<td></td>
<td>C – control</td>
<td>.12</td>
<td>-.40</td>
</tr>
</tbody>
</table>

**Group A: Performance feedback intervention at the instructional level.** As previously noted, students 1 through 4 received the performance feedback intervention following the baseline condition. This intervention consisted of a performance feedback phase first, then a combined intervention, which included a performance feedback component. These students were considered to be performing at the instructional level if
they earned an initial baseline mean score that fell within the range of 14-31 DCPM (Burns et al., 2006). Initial baseline mean scores differ from the baseline means that were reported in Tables 7 and 9 for each student in two ways. First, initial baseline means were calculated based on the first five sessions of the baseline phase, whether or not the student continued with baseline procedures for additional sessions. These scores are reported in Table 3.

The baseline means, on the other hand, were calculated for all sessions that fell within the baseline phase. Second, initial baseline means were calculated by including scores from all three sets, while baseline means were calculated separately for each set (A, B, and C). Student levels were determined for this study based on these initial baseline means so that students could be evenly grouped based upon their ability level when intervention sessions began. Both initial baseline means and baseline means will be reported for each student throughout the individual analysis.
Figure 4. *DCPM scores for each student across baseline, single intervention, and combined intervention phases.*
Figure 5 (cont.).

**Student 3 - (AI - 3)**

- **Baseline**
- **Performance Feedback Intervention**
- **Combined Intervention**

**Student 4 (AI - 4)**

- **Baseline**
- **Performance Feedback Intervention**
- **Combined Intervention**

---

**Observation Session**
Figure 4 (cont.).
Student 5 (AF - 1)

Student 6 (AF - 2)

Student 7 (AF - 3)

Observation Session
Figure 4 (cont.).
Student 8 (BI-1)

Student 9 (BI-2)

Student 10 (BI-3)

Observation Session
Figure 4 (cont.).

Student 11 (BI - 4)

Student 12 (BI - 5)

Student 13 (BI - 6)

Observation Session
Observation Session

Figure 4 (cont.).
Student 14 (BF - 1)

Student 15 (BF - 2)

Student 16 (BF - 3)
Figure 4 (cont.).

**Student 17 (BF - 4)**

- **Baseline**
- **Peer Tutoring Intervention**
- **Combined Intervention**

**Digits Correct Per Minute**

- **A - tutor**
- **B - control**
- **C - tutee**

**Student 18 (BF - 5)**

- **Baseline**
- **Peer Tutoring Intervention**
- **Combined Intervention**

**Digits Correct Per Minute**

- **A - tutee**
- **B - control**
- **C - tutor**

**Observation Session**
Figure 5. DICPM scores for each student across baseline, single intervention, and combined intervention phases.
Figure 5 (cont.).

Observation Session
Figure 5 (cont.).

Observation Session
Figure 5 (cont.).

Student 8 (BI-1)

Errors

Baseline | Peer Tutoring Intervention | Combined Intervention

Digits Incorrect Per Minute

Student 9 (BI-2)

Errors

Baseline | Peer Tutoring Intervention | Combined Intervention

Digits Incorrect Per Minute

Student 10 (BI-3)

Errors

Baseline | Peer Tutoring Intervention | Combined Intervention

Digits Incorrect Per Minute

Observation Session
Figure 5 (cont.).

Student 11 (BI-4)

Errors

Baseline

Peer Tutoring Intervention

Combined Intervention

Observation Session

Student 12 (BI-5)

Errors

Baseline

Peer Tutoring Intervention

Combined Intervention

Student 13 (BI-6)

Errors

Baseline

Peer Tutoring Intervention

Combined Intervention

Observation Session
Figure 5 (cont.).

**Student 14 (BF-1)**

**Errors**
- Baseline
- Peer Tutoring Intervention
- Combined Intervention

- **A - tutee**
- **B - tutor**
- **C - control**

**Observation Session**

**Student 15 (BF-2)**

**Errors**
- Baseline
- Peer Tutoring Intervention
- Combined Intervention

- **A - tutee**
- **B - tutor**
- **C - control**

**Observation Session**

**Student 16 (BF-3)**

**Errors**
- Baseline
- Peer Tutoring Intervention
- Combined Intervention

- **A - tutor**
- **B - tutee**
- **C - control**

**Observation Session**
Observation Session

**Student 1.** It is noteworthy that student 1 missed four consecutive days of school due to illness. As a result, she missed 2 of 5 baseline sessions and the first 2 of 10 intervention sessions. She earned an initial baseline mean score of 17.78 (5.70), but this score may not be representative of her abilities due to her absences. Means for performance during the baseline phase for sets A ($M = 21.00; SD = 8.00$), B ($M = 15.33$);
$SD = 1.15$) and C ($M = 17.00; SD = 6.24$) all fell in the instructional range. Steady improvement was observed during this phase, as indicated by positive slopes for sets A ($b_i = 4.00$), B ($b_i = 1.00$), and C ($b_i = 6.00$). These numbers indicated that student 1 should have continued to improve whether or not an intervention was implemented. Although DCPM scores improved within the baseline phase, no improvement was seen with regard to errors. Student 1 made frequent errors, mainly due to skipped problems, throughout the baseline phase with DICPM scores ranging from 3 to 20 across sets and sessions.

The performance feedback means for student 1 for sets A ($M = 28.00; SD = 0$), B ($M = 24.94; SD = 4.46$), and C ($M = 21.13; SD = 5.22$) all continued to fall in the instructional range. Although baseline predictions indicated that DCPM scores should continue to increase as the performance feedback phase began, improvement came to a halt for set A ($b_i = 0.00$), and continued to slowly increase for sets B ($b_i = 1.09$) and C ($b_i = 0.30$). Additionally, student 1 continued to show no improvement with regard to errors and, in fact, she seemed to make more errors throughout the performance feedback phase. Thus, verification of the performance feedback intervention was not demonstrated for student 1.

During the combined intervention phase, student 1 was partnered with a student who was performing at the mastery range at the start of the study. Due to an oversight, her partner continued to participate in the study, but as his scores were already in the mastery range data were not reported for this student. Student 1 received the combined intervention for 15 sessions. She began and ended the combined phase serving the role of tutor, and acted as the tutee for the five middle sessions. For student 1, set A was
assigned to the tutor condition, set B was assigned to the tutee condition, and set C was assigned to the control condition.

Trends for the combined intervention indicated that there was significant improvement overall throughout this intervention phase ($b_i = 2.37, 5.03, \text{ and } 2.48$ for sets A, B, and C, respectively). Additionally, improvement was such that student 1’s mean scores for sets A ($M = 37.68, SD = 12.75$) and B ($M = 44.75, SD = 13.33$) improved from the instructional to the mastery range, while her score for set C ($M = 29.90, SD = 13.40$) continued to fall within the instructional range. Additionally, student 1 began to consistently perform at mastery levels across all sets, including the control set C. She began to achieve mastery levels of performance for the tutee set B during session 21 and she consistently earned scores in the mastery range for sets A and C during session 26.

When considering DICPM, an obvious pattern was evident. Upon introduction of the tutor condition, the trend of data shifted and the error rate decreased for set A, with the exception of one outlier score. This pattern was not observed for the control set C. Although DICPM scores were erratic for set A, a steady decrease was observed upon the second implementation of the tutor condition towards the end of the study. Likewise, upon introduction of the tutee phase, scores significantly decreased for set B (tutee set) and held at 0 DICPM for the four final sessions of this phase. The implications of these findings will be discussed in the next chapter.

Although marked improvement was observed for all sets, particularly those targeted by the tutee and tutor conditions, visual analysis of the graph for student 1 in Figure 4 indicated that no improvement was observed across sets when tutor condition was first implemented. Although scores for set A were consistently higher than scores
for set C within this condition, a similar pattern was observed throughout the baseline and performance feedback sessions. Accordingly, differentiation could not be attributed to intervention effects.

However, when the tutee condition began, student 1 showed an immediate and significant improvement in fluency scores for set B, which was assigned to the tutee condition. Simultaneously, student 1 began to show improvement, though not as drastic, for the control C throughout the tutee condition. Differentiation was clearly observed, as scores for set B more than doubled those of the control set C. A pattern such as this verified that the tutee condition of the peer tutoring intervention was effective.

Following the tutee condition, the tutor condition was reestablished and student 1 showed immediate and significant improvement for set A, which was assigned to the tutor condition, with the exception of one outlier score that fell well below all other scores for that set. At the same time, the student improved at a comparable rate for set C, which was the control condition. As scores were similar for sets A and C, differentiation was not observed. Consequently, the improvement observed for set A could not be attributed to intervention effects. Thus, the combined intervention was partially verified as effective for student 1.

**Student 2.** Student 2 earned an initial baseline mean score of 19.75 (3.66) which fell in the instructional range. He was absent for 1 out of 15 baseline sessions. The baseline means for sets A \( (M = 26.04; SD=4.98) \), B \( (M = 22.25; SD=4.31) \) and C \( (M = 20.46; SD=4.68) \) all fell in the instructional range. Baseline trends indicated that little to no growth would occur in the absence of intervention \( (b_i = 0.62, 0.24, \text{ and } 0.27 \text{ for sets A, B, and C}) \).
B, and C respectively). Visual analysis of student 2’s graph in Figure 5 indicated errors were erratic across sets with no discernible pattern. DICPM scores ranged from 2 to 14.

Upon implementation of the performance feedback intervention an immediate shift in trend was apparent for set A ($b_i = 2.24$), but no significant change in trend was observed for sets B or C ($b_i = -0.16$ and 0.05 respectively). The performance feedback mean for A ($M = 38.10; SD=8.01$) was considerably higher than the baseline mean, as the score improved into the mastery range. Student 2 began consistently performing in the mastery range for set A during session 18. Means from this phase for sets B ($M = 25.25; SD=3.90$) and C ($M = 23.00; SD=6.04$) showed only slight improvement over baseline means and both remained in the instructional range.

Prediction based on baseline trends indicated that student 2 would not show much improvement unless an intervention was implemented. Visual analysis indicated that this trend changed upon implementation of the performance feedback intervention for set A, but not for sets B or C. Additionally, visual analysis indicated no change in the pattern associated with DICPM, as errors continued to fluctuate with no decreasing trend. Thus, verification was not demonstrated.

Student 2 did not begin the combined intervention phase until the last five sessions of the study, when he acted as a tutee. He never had an opportunity to serve as a tutor. For student 2, set A was assigned to the tutee condition, while sets B and C were assigned to the control condition. Student 2 was paired with student 7. No visible change in scores was noticeable for either control set upon implementation of the combined intervention. Slopes remained relatively flat ($b_i = 0.38$ and -0.25 for sets B and C, respectively) and combined intervention means were only slightly higher than those
present in the performance feedback phase \( (M = 29.90; SD=2.38 \text{ and } M = 28.10; SD=0.55 \) for sets B and C, respectively). Student 2 did not achieve scores in the mastery range for either of the control sets B and C.

However, visual analysis of student 2’s graph in Figure 4 revealed an immediate and noticeable shift in scores for set A, which was assigned to the tutee condition. Although the slope for set A was negative in the combined intervention phase \( (b_i = -2.40) \), the mean score for this set \( (M = 55.60; SD = 8.93) \) showed considerable improvement. An outlier score of 41 was obtained in the final intervention session which negatively impacted these results. Differentiation was clearly present, as scores for set A doubled those of the control conditions. Although scores for set A were considerably higher than sets B and C in the previous intervention phase, a shift in trend was observed for set A upon introduction of this condition, with scores increasing at an even greater rate than previously seen. Similarly, when considering error rates, an abrupt and sizable shift occurred upon implementation of the tutee condition for set A (see graph for student 4 in Figure 5). DICPM scores immediately dropped to zero and remained below one throughout this intervention phase. Thus, the tutee condition of the peer tutoring intervention was verified as effective for student 2.

**Student 3.** Student 3 earned an initial baseline mean score of 17.80 (4.10), which fell in the instructional range. His baseline means for sets A \( (M = 21.6; SD = 4.45) \) and B \( (M = 19.4; SD = 0.89) \) both fell in the instructional range, while his mean score for C \( (M = 12.40; SD = 3.03) \) fell in the frustrational range. Set C was clearly the most difficult for this student. Baseline trends indicated that scores would slowly improve over time for sets A \( (b_i = 1.40) \) and B \( (b_i = 0.90) \), but remain relatively stable for set C \( (b_i \)
 Upon implementation of the performance feedback intervention an immediate and obvious increase in DCPM scores was observed for sets A and B; however, set C showed no immediate improvement. While scores for set C started lower than those of sets A and B, all three sets showed a similar rate of change throughout the intervention phase ($b_i = 1.34, 1.24, \text{ and } 1.51$ for sets A, B, and C respectively). This positive trend was notable for set C because baseline predictions indicated growth was unlikely for this set in the absence of an intervention. Intervention means improved for all three sets, with the mean for set A ($M = 40.20; SD = 4.49$) improving from the instructional range to the mastery range, as student 3 began consistently performing in the mastery range during session 6 for this set. Further, the mean for set C ($M = 21.5; SD = 5.87$) improved from the frustrating to the instructional range and student 3 began consistently performing in the mastery range during session 12. The mean score for set B ($M = 30.85; SD = 4.11$) remained in the instructional range, but also showed significant improvement. Based on the immediate increase in scores for sets A and B and the increased rate of improvement for set C upon introduction of the performance feedback intervention, it appeared that the performance feedback intervention was effective. As a result, intervention effectiveness was verified for this student.

Student 3 was paired with student 4 for the 15 sessions of the combined intervention. This student began and finished the intervention as the tutor and served as the tutee during the five sessions in the middle of this phase. For student 3, set A was assigned to the tutee condition, set B was assigned to the control condition, and set C was
assigned to the tutor condition. The mean for the combined intervention for set A \( (M = 50.57; SD = 8.95) \) indicated continued improvement and remained in the mastery range. Additionally, mean scores for sets B \( (M = 47.80; SD = 7.94) \) and C \( (M = 48.18; SD = 7.65) \) both improved from the instructional range to the mastery range. Student 3 began to consistently perform in the mastery range for set C, which was assigned to the tutor condition, during session 17.

Interestingly, an encouraging trend was also observed when assessing the error data. Although student 3 began the study with relatively high levels of accuracy, he tended to make a few mistakes on each probe throughout the baseline and performance feedback interventions. However, upon implementation of the combined intervention DICPM scores were consistently 0, with the exception of a few errors that were made in the control set B. This pattern is similar to that which was observed for previous students in which increases in accuracy were observed as a result of the implementation of the combined intervention.

Trends from the performance feedback intervention predicted that DCPM scores would continue to increase at a similar rate if the intervention remained in place without change. The calculated slopes for the combined intervention phase \( (b_i = 1.45, 1.45, \text{ and } 1.43 \text{ for sets A, B, and C respectively}) \) were similar to those observed during the performance feedback intervention, which indicated that the combined intervention was not more effective than the performance feedback intervention in isolation. Although visual analysis showed that scores for set A were higher than those of set C during the tutee condition, differentiation could not be attributed to intervention effects as the same pattern was evident throughout the performance feedback intervention. Additionally,
growth for sets B and C followed a similar trajectory, despite the fact that no additional intervention was in place for set C. Thus, the combined intervention could not be verified as effective for student 3.

**Student 4.** Student 4 earned an initial baseline mean score of 15.23 (6.16) which fell in the instructional range. His baseline means for sets A ($M = 18.50; SD = 6.24$), B ($M = 13.4; SD = 5.27$), and C ($M = 13.8; SD = 6.76$) all fell in the instructional range. Baseline trends indicated that student performance should continue to improve in the absence of additional intervention ($b_i = 3.00, 1.70, and 3.80$ for sets A, B, and C respectively). A visual analysis of the graph for student 4 in Figure 5 indicated that he began the study with high rates of accuracy, making occasional mistakes across sets and sessions throughout the baseline and intervention phases, as DICPM scores were consistently zero.

Upon implementation of the performance feedback intervention, scores for all three sets showed an initial decline. Scores quickly recovered and surpassed baseline scores, but then dropped again for sets B and C. The performance feedback means were slightly higher for sets A ($M = 26.70; SD = 5.74$) and B ($M = 18.30; SD = 7.41$) but the mean was lower for set C ($M = 12.50; SD = 4.69$). The intervention means for sets A and B remained in the instructional range, while the mean for set C declined into the frustrational range.

Although the trends observed throughout the baseline phase predicted that student scores would continue to improve, the rates of improvement slowed upon the introduction of the performance feedback intervention ($b_i = 1.23, 1.04, and 0.35$ for sets A, B, and C respectively). Thus, the performance feedback intervention did not appear
to have a significant effect on the performance of student 4 and verification was not
evident.

Student 4 was paired with student 3 for the combined intervention phase, which
he began and ended as the tutee. He served as the tutor for the five middle sessions. Set
A was assigned to the tutor condition, set B was assigned to the control condition, and set
C was assigned to the tutee condition. Mean scores for the combined intervention were
all indicative of improvement with mean scores of 34.55 (9.00), 27.33 (5.63), and 29.30
(8.06) for sets A, B, and C, respectively. Growth was such that DCPM scores increased
to mastery levels for set A during session 23, for set B during session 26, and for set C
during session 28. Although all sets showed significant improvement, sets A and C
showed slightly larger gains than the control set B.

Upon implementation of the tutee condition, scores for set C began to show
improvement; however, there was also improvement for the control set B. Therefore,
differentiation was not present and we could not attribute growth to the tutee condition of
the combined intervention. However, when the tutor condition was reinstated at the end
of the combined intervention phase, differentiation between sets was established as
DCPM scores for set C continued to improve above and beyond those of the control set B
during the final three sessions.

When considering rates of improvement an interesting pattern emerged. The
performance feedback trends predicted that growth would continue to occur across sets,
with faster rates of improvement predicted for set B (control set) than for set C (tutee
condition). This trend changed during the combined intervention, with less growth
observed for the control set C (b_i = 0.61) and more growth observed for set B (b_i = 1.27).
When taken together, data indicated that the tutee condition of the combined intervention was effective for student 4.

Additionally, when the tutor condition was implemented, DCPM scores for set A showed rapid improvement \( (b_i = 5.28) \) and differentiation was established as scores climbed well above those of the control set B. The growth observed for the tutor condition of the combined intervention clearly surpassed rates predicted by the calculated baseline and performance slopes. Thus, the evidence suggests that the tutor condition of the combined intervention was effective. Consequently, both conditions of the combined intervention were verified as effective for student 4.

**Summary for students in group A performing at the instructional level.** The combined intervention cannot be deemed effective in its entirety, but various elements of it impacted the students differently. When considering the effect of the performance feedback in isolation, only 1 of 4 students demonstrated significant improvement upon implementation of this intervention. As this finding was not replicated, it could not be verified as an effective intervention for students performing in the instructional range at the start of the study.

Mixed results were observed with regard to the roles students played in the combined intervention. Three of 6 students showed marked improvement as the result of the tutee condition, but only 1 of 6 students showed increased scores as the result of the tutor condition. Additionally, while accuracy rates improved for all students, a significant drop in DICPM scores was attributed to the tutee condition for 2 of 6 students.

In a multiple-baseline across-subjects design, one must observe change across a minimum of three subjects in order for an intervention to be verified as effective (Horner
et al., 2005). In this case, the performance feedback intervention was only verified as effective for student 3, so it could not be considered as an effective intervention for this group. Students 1, 2, and 4 all showed improvement as the result of the tutee condition of the combined intervention, so this condition could be verified as effective for this group of students. However, the tutor condition was only verified as effective for student 4, indicating that the combined intervention was only partially verified as effective for this group of students.

There were other noteworthy patterns within this group. First, all students showed marked improvement over the course of the study. While student 3 was the only one that earned mean scores in the mastery range for all three sets within the combined intervention phase, all four students began to consistently perform at the mastery level under the tutee condition before the end of the study. Further, students 1, 3, and 4 began to consistently perform within the mastery range for all three sets throughout the course of the intervention phases.

Additionally, the patterns suggest that the combined intervention addressed accuracy issues, which correlated with improved fluency scores for these students. Further implications of this finding will be discussed in the following chapter.

**Group A (continued): Performance feedback intervention at the frustrational Level.** Students 5 through 7 received an initial baseline average score that fell in the frustrational range (Burns et al., 2006). These students also received the performance feedback intervention following the baseline phase.

**Student 5.** Student 5 earned an initial baseline mean score of 8.10 (4.17) which fell in the frustrational range. Her baseline means for sets A ($M = 11.67; SD = 3.75$), B
(\(M = 6.56; SD = 3.24\)), and C (\(M = 6.61; SD=2.37\)) all fell in the frustrational range.

Baseline trends predicted that little or no improvement would occur in the absence of intervention (\(b_i = 0.16, 0.08, \text{ and } 0.08\) for sets A, B, and C respectively). Visual analysis of error patterns indicated that student 5 struggled with accuracy. Throughout the baseline phase DICPM scores ranged from 1-17 across sets and sessions, with most scores falling above 5. No discernible trend was present for errors.

No improvement was observed upon the implementation of the performance feedback intervention, as the mean score for set A (\(M = 10.20; SD = 2.95\)) was lower than the one observed during the baseline phase and only minimal improvement was observed in the mean scores for sets B (\(M = 8.80; SD = 2.17\)) and C (\(M = 8.60; SD = 2.61\)). All scores continued to fall in the frustrational range.

When considering trends, student 5 showed little improvement for set B (\(b_i =0.40\)) throughout the course of the performance feedback intervention, while scores for sets A and C declined over time (\(b_i = -0.90 \text{ and } -1.20\) for sets A and C respectively). Visual analysis of error patterns indicated that DICPM scores immediately increased upon implementation of the performance feedback intervention, and then declined throughout the course of the intervention. Moreover, the improvements in accuracy did not correlate with improved fluency scores. Thus, there was not sufficient evidence to suggest that the performance feedback was an effective intervention for this student.

During the combined intervention phase student 5 was partnered with a student who was performing at the mastery range at the start of the study. Due to an oversight, she continued to participate in the study, but as her scores were already in the mastery range data were not reported for this student. Student 5 began and ended the combined
intervention in the tutee role and served as the tutor for the middle five sessions. Set A was assigned to the control condition, set B was assigned to the tutee condition, and set C was assigned to the tutor condition for this student.

Visual analysis of the graph for student 5 in Figure 4 yielded no discernible change in scores or trends upon implementation of the combined intervention. Mean scores for this phase were slightly higher than those from the baseline and performance feedback phases, with mean scores of 12.47 (4.13), 9.70 (4.42), and 9.90 (3.25) for sets A, B, and C, respectively. Slopes across sets continued to show a decline for sets A ($b_i = -0.22$) and C ($b_i = -0.09$) and minimal growth for set B ($b_i = 0.41$). Furthermore, differentiation was not observed as scores for the control condition were higher than those of the tutee and tutor condition throughout much of the combined intervention phase. Additionally, student 5 displayed little to no growth throughout the intervention phases; she never earned scores in the instructional or mastery range of performance.

Visual analysis of trends for errors was difficult as student 5’s performance was erratic. DICPM scores for set B showed a steep increase during the first implementation of the tutee condition and then showed a steep decrease during the second implementation of the tutee condition. Lack of consistency makes it impossible to attribute intervention effects on these scores. Student 5 also displayed erratic performance for set C, although visual analysis indicated a marginally decreasing trend. Further, student 5 earned the lowest DICPM scores for the control set A over the course of the combined intervention phase. In addition to erratic DICPM scores, any improvements in accuracy that may have occurred did not positively impact fluency scores for student 5.
In brief, student 5 showed little or no improvement upon the introduction of the performance feedback intervention or the combined intervention. This was true for both DCPM and DICPM scores. Hence, there is no evidence that either intervention was effective for this student.

**Student 6.** Student 6 earned an initial baseline mean score of 10.80 (3.36) which fell in the frustrating range. For student 6, baseline means for sets A \((M = 13.87; SD = 4.29)\), B \((M = 8.80; SD = 2.83)\), and C \((M = 10.47; SD = 2.83)\) all fell in the frustrating range. Analysis of trends indicated that there would be little or no improvement in the absence of intervention \((b_1 = 0.28, 0.07, \text{ and } -0.16 \text{ for sets A, B, and C respectively})\). Visual analysis of error patterns indicated erratic performance with scores ranging from zero to 15 across sets and sessions throughout the baseline phase.

Visual analysis yielded no obvious improvement upon the implementation of the performance feedback intervention, with the exception of set A. An immediate improvement was observed for set A, but scores quickly returned to baseline levels. The performance feedback means for sets A \((M = 15.75; SD = 5.53)\) and B \((M = 9.33; SD = 3.19)\) showed slight improvement and were not indicative of noteworthy change, while the intervention mean for set C \((M = 9.39; SD = 2.43)\) decreased slightly.

When considering trend, visual analysis revealed no discernible change in data upon the implementation of the performance feedback intervention. Slopes confirmed that improvement did not take place \((b_1 = -1.66, -0.62, \text{ and } -0.07 \text{ for sets A, B, and C respectively})\) and, instead, indicated that scores decreased over time.

Visual analysis of errors during the performance feedback phase for student 6 indicated that DICPM scores continued to be erratic during the first six sessions and
became more stable, consistently falling below 5 DICPM for the last four sessions. Although improvement was noted with regard to errors, it took place towards the end of the intervention phase, making it difficult to attribute changes to the implementation of the intervention. Additionally, the increase in accuracy did not coincide with an increase in fluency scores. Consequently, the performance feedback cannot be deemed effective for student 6.

Student 6 was paired with a student who dropped out of the study but agreed to help by serving the role of tutor for the final week of the study. Because the combined intervention did not begin for student 6 until the final five sessions of the study, student 6 only acted as the tutee and did not serve the role of tutor. Additionally, this student was only present for four sessions of the tutee intervention, making it difficult to draw any conclusions given the limited amount of data. Visual analysis of the graph for student 6 in Figure 4 revealed erratic performance within the combined intervention phase. For this student, set C was assigned to the tutee condition, while sets A and B were assigned to the control condition.

When considering the means for the combined intervention, improvement was noted for set C ($M = 14.63; SD = 6.80$) as the mean score for this set improved from the frustrational range to the instructional range. Growth was evident even though an outlier score of six negatively impacted this mean score, as three of four scores were above those that had previously been obtained. These scores were also above the DCPM scores for control sets A and B; however, scores were variable for all three sets throughout this intervention phase and differentiation was not established.
The combined intervention mean score for set A \( (M = 12.25; SD = 7.80) \) was lower than the mean scores for the baseline and performance feedback intervention phases and while set B yielded a combined intervention mean \( (M = 9.13; SD = 4.13) \) that was higher than the baseline mean, but lower than the performance feedback mean. The means for sets A and B remained in the frustrational range. Error rates throughout the combined condition remained low across all sets and sessions. However, this pattern was observed towards the end of the performance feedback phase and cannot be attributed to the combined intervention. Thus, the combined intervention was not verified as effective for this student.

**Student 7.** Student 7 earned an initial baseline mean score of 11.67 (3.84), which fell in the frustrational range. For student 7, baseline means for sets A \( (M = 12.03; SD = 3.16) \), B \( (M = 8.13; SD = 3.57) \), and C \( (M = 9.97; SD = 3.95) \) all fell in the frustrational range. The observed trends from the baseline phase indicated that improvement would not occur in the absence of some type of intervention \( (b_i = -0.15, 0.04, \text{ and } 0.12 \text{ for sets A, B, and C respectively}) \). Visual analysis of error patterns in the graph for student 7 in Figure 5 indicated erratic performance throughout the baseline phase with scores ranging from 0 to 17.

No change in trend or performance was observed upon implementation of the performance feedback intervention. In fact, the intervention means for sets A \( (M = 8.70; SD = 3.13) \), B \( (M = 7.40; SD = 3.33) \), and C \( (M = 8.60; SD = 4.13) \) all decreased slightly, while remaining comparable with mean scores from the baseline, with the largest decrease in scores observed for set A. All mean scores continued to fall in the frustrational range. Trends for the performance feedback phase indicated that student 7
did not progress as a result of this intervention ($b_i = -1.20, -0.70, \text{ and } -0.40$ for sets A, B, and C respectively), thus a prediction could be made that scores would decrease without added intervention. When considering error patterns, DICPM scores remained erratic, with a slight decrease observed for set A. Therefore, the performance feedback intervention was not effective for this student.

Student 7 was paired with student 2 for the combined intervention phase. As the combined intervention was not implemented until the final week of the study, student 7 served the role of tutor, but did not have an opportunity to act as the tutee. For this student, set A was assigned to the tutor condition, while sets B and C were assigned to the control condition. Upon initiation of the combined intervention no discernible change in scores was evident for any of the three sets, including set A. Scores were also erratic, and differentiation between sets was not observed. The mean score for set A ($M = 9.30; SD = 4.09$) showed slight improvement over the performance feedback mean, but it did not reach the score achieved during the baseline phase. The mean scores for sets B ($M = 7.00; SD = 5.80$) and C ($M = 8.20; SD = 6.95$) were lower than those observed in previous phases of the study. Although a positive trend was observed for sets A ($b_i = 0.75$) and C ($b_i = 0.90$), this was largely due to a drop in scores at the start of this intervention phase. The slope for set B remained essentially flat ($b_i = -0.18$). As no improvement was observed throughout the study, student 7 continued to perform within the frustrational range, earning no scores in the instructional or mastery ranges.

Similarly, no significant change was observed with regard to errors. Although DICPM scores were lower for set A, which was assigned to the tutor session, the same pattern was observed throughout the performance feedback intervention, and so the
discrepancy between set A and the other two sets cannot be attributed to the tutoring condition of the combined intervention. When taken together, the data suggest that the combined intervention was not effective for this student.

**Summary for students in group A performing at the frustrational level.** Based on data from students 5, 6, and 7, the performance feedback intervention was ineffective. It is notable that none of these students showed considerable improvement throughout any phases of the study. Only 1 of the 3 students showed minimal improvement under the tutee condition of the combined intervention, but consistent scores were not observed, making it impossible to draw sound conclusions.

Although DCPM scores did not improve, DICPM scores indicated improved rates of accuracy for the tutee set for student 5 and for all three sets for student 6. However, improvements in accuracy did not coincide with improved fluency rates. Thus, neither the performance feedback intervention, nor the combined intervention were verified as effective for this group.

**Summary of the data for group A: Students receiving the performance feedback intervention in isolation.** As previously noted, students 3 of 4 students who began the study in the instructional range showed improvement as the result of the tutee condition of the combined intervention, so this condition could be verified as effective for this group of students. However, the tutor condition was only verified as effective for student 4, indicating that the combined intervention was only partially verified as effective for this group of students. When considering these results in conjunction with students performing in the frustrational range, additional verification of effectiveness was not present. None of the 3 students from the frustrational group showed improved
fluency scores as the result of the performance feedback intervention. Student 6 showed minimal improvement under the tutee condition of the combined intervention, but not enough improvement to verify its effectiveness.

There were other noteworthy patterns within this group. First, all students in the instructional range showed marked improvement over the course of the study, while no students in the frustrational range showed the same degree of improvement. Second, in regard to the baseline phase, 3 of 4 students in the instructional range displayed significant growth for at least one set, while none of the students in the frustrational range displayed this pattern. Third, while accuracy improved for some students in both groups, as evidenced by decreased DICPM scores for students 5 of 7 students, increased accuracy did not correlate with increased fluency scores for students performing in the frustrational range, while they did for those performing in the instructional range. Further implications of these findings will be discussed in the following chapter.

Table 9

<table>
<thead>
<tr>
<th>Group B: Peer Tutoring Intervention</th>
<th>Condition</th>
<th>Baseline</th>
<th>Performance Feedback</th>
<th>Combined Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>A - tutor</td>
<td>3.65</td>
<td>2.24</td>
<td>1.59</td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>Instructional #</td>
<td>Condition</td>
<td>Baseline</td>
<td>Performance Feedback</td>
</tr>
<tr>
<td>--------------</td>
<td>----------------</td>
<td>------------</td>
<td>----------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Student 8</td>
<td>Instructional #1 (BI-1)</td>
<td>B - control</td>
<td>-1.50</td>
<td>1.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C - tutee</td>
<td>1.60</td>
<td>0.72</td>
</tr>
<tr>
<td>Student 9</td>
<td>Instructional #2 (BI-2)</td>
<td>A - tutee</td>
<td>2.70</td>
<td>3.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B - control</td>
<td>-1.45</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C - tutor</td>
<td>-1.90</td>
<td>-0.15</td>
</tr>
<tr>
<td>Student 10</td>
<td>Instructional #3 (BI-3)</td>
<td>A - control</td>
<td>-1.10</td>
<td>-0.79</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B - tutor</td>
<td>2.75</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C - tutee</td>
<td>-1.10</td>
<td>1.95</td>
</tr>
<tr>
<td>Student 11</td>
<td>Instructional #4 (BI-4)</td>
<td>A - control</td>
<td>2.40</td>
<td>1.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B - tutee</td>
<td>-1.01</td>
<td>2.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C - tutor</td>
<td>0.08</td>
<td>2.42</td>
</tr>
<tr>
<td>Student 12</td>
<td>Instructional #5 (BI-5)</td>
<td>A - control</td>
<td>2.58</td>
<td>1.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B - tutor</td>
<td>2.26</td>
<td>2.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C - tutee</td>
<td>2.68</td>
<td>0.99</td>
</tr>
<tr>
<td>Student 13</td>
<td>Instructional #6 (BI-6)</td>
<td>A - tutor</td>
<td>2.25</td>
<td>-0.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B - tutee</td>
<td>1.80</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C - control</td>
<td>-0.20</td>
<td>0.34</td>
</tr>
<tr>
<td>Student 14</td>
<td>Frustrational #1 (BF-1)</td>
<td>A - tutor</td>
<td>1.10</td>
<td>-1.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B - tutee</td>
<td>0.25</td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C - control</td>
<td>-0.10</td>
<td>-0.32</td>
</tr>
<tr>
<td>Student 15</td>
<td>Frustrational #2 (BF-2)</td>
<td>A - tutee</td>
<td>-1.50</td>
<td>-1.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B - tutor</td>
<td>-0.40</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C - control</td>
<td>-0.25</td>
<td>0.34</td>
</tr>
<tr>
<td>Student 16</td>
<td>Frustrational #3 (BF-3)</td>
<td>A - tutor</td>
<td>2.65</td>
<td>-1.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B - tutee</td>
<td>4.20</td>
<td>-0.39</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C - control</td>
<td>3.55</td>
<td>-1.13</td>
</tr>
<tr>
<td>Student 17</td>
<td>Frustrational #4 (BF-4)</td>
<td>A - tutor</td>
<td>-0.45</td>
<td>-0.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B - control</td>
<td>0.04</td>
<td>-0.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C - tutee</td>
<td>-0.93</td>
<td>-0.11</td>
</tr>
<tr>
<td>Student 18</td>
<td>Frustrational #5 (BF-5)</td>
<td>A - tutee</td>
<td>1.06</td>
<td>-0.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B - control</td>
<td>0.93</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C - tutor</td>
<td>0.38</td>
<td>0.33</td>
</tr>
</tbody>
</table>

Table 10

Mean Scores across Intervention Settings for Group B

<table>
<thead>
<tr>
<th>Group B: Peer Tutoring Intervention</th>
<th>Condition</th>
<th>Baseline</th>
<th>Performance Feedback</th>
<th>Combined Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>Instructional #</td>
<td>A - tutor</td>
<td>B - control</td>
<td>C - tutee</td>
</tr>
<tr>
<td>---------</td>
<td>----------------</td>
<td>-----------</td>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>8</td>
<td>(BI-1)</td>
<td>25.10 (6.00)</td>
<td>28.00 (5.34)</td>
<td>24.00 (5.10)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>35.55 (6.46)</td>
<td>25.65 (5.49)</td>
<td>23.60 (2.51)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>57.10 (10.39)</td>
<td>38.87 (12.12)</td>
<td>34.98 (9.47)</td>
</tr>
<tr>
<td>9</td>
<td>(BI-2)</td>
<td>22.40 (6.32)</td>
<td>17.90 (3.99)</td>
<td>14.30 (4.41)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>38.40 (14.51)</td>
<td>17.30 (3.74)</td>
<td>16.60 (4.35)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>42.55 (6.08)</td>
<td>20.82 (7.20)</td>
<td>23.17 (4.32)</td>
</tr>
<tr>
<td>10</td>
<td>(BI-3)</td>
<td>20.83 (3.55)</td>
<td>17.90 (3.99)</td>
<td>17.65 (3.00)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20.30 (5.56)</td>
<td>17.30 (3.74)</td>
<td>16.60 (4.35)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>27.97 (9.20)</td>
<td>20.82 (7.20)</td>
<td>23.17 (4.32)</td>
</tr>
<tr>
<td>11</td>
<td>(BI-4)</td>
<td>30.25 (8.72)</td>
<td>23.65 (8.68)</td>
<td>21.70 (10.21)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>52.89 (11.61)</td>
<td>44.87 (16.50)</td>
<td>50.95 (5.92)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>71.44 (8.69)</td>
<td>52.89 (11.61)</td>
<td>54.75 (6.62)</td>
</tr>
<tr>
<td>12</td>
<td>(BI-5)</td>
<td>15.50 (6.50)</td>
<td>13.30 (3.87)</td>
<td>14.00 (2.35)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20.20 (7.85)</td>
<td>16.00 (4.83)</td>
<td>15.06 (5.79)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25.40 (6.69)</td>
<td>25.40 (6.69)</td>
<td>17.83 (3.59)</td>
</tr>
<tr>
<td>13</td>
<td>(BI-6)</td>
<td>13.20 (2.95)</td>
<td>8.70 (4.02)</td>
<td>10.40 (12.50)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12.00 (7.25)</td>
<td>7.00 (2.12)</td>
<td>5.00 (2.42)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19.30 (9.70)</td>
<td>19.20 (9.83)</td>
<td>17.83 (3.59)</td>
</tr>
<tr>
<td>14</td>
<td>(BF-1)</td>
<td>12.50 (3.08)</td>
<td>12.50 (3.65)</td>
<td>13.63 (3.35)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14.90 (5.18)</td>
<td>11.00 (4.85)</td>
<td>7.55 (5.29)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15.45 (8.70)</td>
<td>9.50 (8.35)</td>
<td>6.00 (6.28)</td>
</tr>
<tr>
<td>15</td>
<td>(BF-2)</td>
<td>12.50 (3.08)</td>
<td>12.50 (3.65)</td>
<td>13.63 (3.35)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14.90 (5.18)</td>
<td>11.00 (4.85)</td>
<td>7.55 (5.29)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15.45 (8.70)</td>
<td>9.50 (8.35)</td>
<td>6.00 (6.28)</td>
</tr>
<tr>
<td>16</td>
<td>(BF-3)</td>
<td>9.63 (3.50)</td>
<td>9.50 (5.45)</td>
<td>10.88 (5.36)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>18.25 (4.65)</td>
<td>18.75 (10.18)</td>
<td>10.50 (5.48)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12.39 (4.23)</td>
<td>21.60 (5.13)</td>
<td>12.21 (7.54)</td>
</tr>
<tr>
<td>17</td>
<td>(BF-4)</td>
<td>13.35 (2.76)</td>
<td>9.90 (1.52)</td>
<td>11.10 (3.36)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13.40 (2.51)</td>
<td>13.10 (3.66)</td>
<td>13.50 (2.45)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>18.00 (11.14)</td>
<td>20.70 (6.97)</td>
<td>28.90 (9.07)</td>
</tr>
<tr>
<td>18</td>
<td>(BF-5)</td>
<td>14.55 (4.91)</td>
<td>11.05 (5.52)</td>
<td>11.00 (3.78)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14.30 (2.49)</td>
<td>10.20 (4.37)</td>
<td>10.80 (2.56)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14.40 (6.35)</td>
<td>12.90 (4.64)</td>
<td>17.50 (5.28)</td>
</tr>
</tbody>
</table>

**Group B: Peer tutoring intervention for students at the instructional level.**

As previously noted, there were three phases to this study. Group B first participated in the baseline condition to ascertain their level, then they received a peer tutoring...
intervention, and, lastly, the combined intervention was implemented in which students received performance feedback in addition to a peer tutoring intervention. In the same way that students from group A were identified, students 8-13 were described as performing at the instructional level because they produced an initial baseline average score that fell in the instructional range. This score was calculated using scores from all three sets within the first five sessions of the baseline phase.

**Student 8.** Student 8 earned an initial baseline mean score of 25.70 (4.38) which fell in the instructional range (Burns et al., 2006). Baseline means for sets A \((M = 25.10; SD = 6.00)\), B \((M = 28.00; SD = 5.34)\), and C \((M = 24.00; SD = 5.10)\) were all in the instructional range as well. Baseline trends for this student indicated that scores should improve over time for sets A \((b_i = 3.65)\) and C \((b_i = 1.60)\) whether or not an intervention was implemented, while scores for set B decreased over time \((b_i = -1.50)\), indicating that growth was unlikely to occur without intervention.

Student 8 was paired with student 9 for the peer tutoring and combined interventions. He engaged in the peer tutoring intervention for a total of 10 sessions, beginning as the tutee and ending as the tutor. For student 8, set A was assigned to the tutor condition, set B was assigned to the control condition, and set C was assigned to the tutee condition. The peer tutoring mean for set A \((M = 35.55; SD = 6.46)\) improved from the instructional range to the mastery range, while the mean scores for sets B \((M = 25.65; SD = 5.49)\) and C \((M = 23.60; SD = 2.51)\) remained in the instructional range.

Analysis of trends indicated that student 8 displayed moderate improvement for sets B and C \((b_i = 1.32, \text{ and } 0.90 \text{ for sets B and C respectively})\) during the peer tutoring intervention. However, he displayed a decrease in scores for the first session of the peer
tutoring phase when he served as tutee, so as scores returned to baseline levels a positive trend was observed. Additionally, differentiation between sets B and C was not observed as DCPM scores were similar for both sets. Only scores for set A began to exceed baseline levels of performance upon implementation of the tutor condition and differentiation was observed as scores for set A were significantly higher than those for control set B for the final three sessions of the tutor condition. Additionally, when considering the slope for set A \( (b_i = 3.95) \), growth occurred at a slightly faster rate than was predicted by the baseline trends. Finally, student 8 began consistently performing at mastery levels for set A during session 13. Based on the data patterns only the tutor condition of the peer tutoring intervention could be verified as effective for this student.

Student 8 received the combined intervention for 15 sessions. He began and ended the intervention phase as the tutee, and served as the tutor for the five middle sessions. No immediate change was evident for DCPM scores upon the implementation of the tutee condition of the combined intervention for sets B (control set) or C (tutee set).

Although scores showed an increased rate of improvement for set C \( (b_i = 1.39) \) upon the reintroduction of the tutee condition, scores also improved on a similar trajectory for the control set B \( (b_i = 2.31) \). Mean scores for sets B \( (M = 38.87; SD = 12.12) \) and C \( (M = 34.98; SD = 9.47) \), which corresponded with the control and tutee conditions, improved from the instructional range to the mastery range during the combined intervention. Further, student 8 began consistently performing in the mastery range for set B during session 21 and for set C during session 28. Because similar rates of improvement were observed for sets B and C and, in fact, it seems that growth
occurred faster for the control set B, differentiation between sets was not observed and it was impossible to attribute improved performance to the tutee condition of the combined intervention. Thus, the tutee condition of the combined intervention could not be verified as effective. Instead, one must consider the possible positive effects of repeated practice for student 8, as significant improvement was observed for all sets throughout this study. Implications of this finding will be discussed in the next chapter.

Upon implementation of the tutor condition, DCPM scores for set A were erratic, with 2 of the 5 scores showing significant improvement. Yet differentiation was clearly present as all five scores remained above the control set B throughout this phase, and the mean score for set A ($M = 57.10; SD = 10.39$), which was assigned the tutee condition, showed considerable improvement and remained the mastery range. Although growth slowed ($b = 1.59$ for set A), this was to be expected due to possible ceiling effects. Thus, the tutee condition was verified as effective for this student.

Visual analysis of the error patterns presented in the graph for student 8 in Figure 5 error patterns further supports these conclusions. DICPM scores remained relatively stable throughout the baseline phase. Overall, the error rate decreased over the course of the study, with scores consistently reaching 0 for all 3 sets by the completion of the study. It is interesting to note that DICPM scores steadily declined for set A during the tutee condition in the peer tutoring intervention, reaching scores of 0 by the final session. Although error scores also decreased for set C, they did not decrease upon implementation of the tutee condition. Further, the decrease in scores for sets B (control set) and C (tutee set) followed a similar trajectory, making it impossible to attribute positive effects to the tutee condition. These findings parallel those observed for DCPM
scores, providing further evidence that the tutor condition of the peer tutoring and combined interventions was effective, while significant improvement was also observed, albeit at a slower rate, for the tutee and control conditions

**Student 9.** Student 9 earned an initial baseline mean score of 18.20 (5.77) which fell in the instructional range (Burns et al., 2006). Baseline means for sets A ($M = 22.40; SD = 6.32$), B ($M = 17.90; SD = 3.99$), and C ($M = 14.30; SD = 4.41$) all fell in the instructional range. With the exception of set A ($b_i = 1.59$), for which a positive slope was observed, baseline trends predicted that scores would decline in the absence of intervention ($b_i = -1.45$ and -1.90 for sets B and C respectively). A visual analysis of the graph for student 9 in Figure 5 indicated a pattern in which error rates increased throughout the course of the baseline phase with DICPM scores ranging from 0 to 9.5.

Student 9 was paired with student 8 for the peer tutoring and combined phases. He was assigned set A for the tutee condition, set B for the control condition, and set C for the tutor condition. He began the peer tutoring intervention, which lasted 10 sessions, as the tutor (with set C) and finished as the tutee (with set A). No change in trend was observed for DICPM or DCPM scores upon the implementation of the tutor condition of the peer tutoring intervention and differentiation between sets B and C was not observed. Thus, the tutor condition of the peer tutoring intervention was not verified as effective.

However, as can be seen in the graphs for student 9 in Figures 4 and 5, visual analysis indicated that immediate and significant improvement was seen for both DCPM and DICPM scores when student 9 was the tutee. In fact, differentiation was clearly established as DCPM scores for set A more than doubled those observed for sets B and C (the control and tutor conditions) and DICPM score fell to 0. The improvement was such
that student 9 began to consistently perform in the mastery range at session 13, resulting in a peer tutoring mean score for set A ($M = 38.40; SD = 14.51$) that was in the mastery range, while the means for sets B ($M = 17.30; SD = 3.74$) and C ($M = 16.60; SD = 4.35$) remained in the instructional range. Thus, verification of effectiveness was present for the tutee condition, and the peer tutoring intervention was partially verified as effective.

Student 9 began and ended the combined intervention as the tutee and served as the tutor during the five sessions in the middle of the intervention phase. Upon implementation of the tutor condition, student 9 consistently demonstrated improvement for set C, which was assigned to the tutor condition. Scores were consistently above those of control set B throughout all four sessions of the tutor condition. However, the slope for the control set B ($b_i = 1.38$) was greater than the slope for set C ($b_i = 0.48$) within this intervention phase. Consequently, scores for both sets were similar at the end of the combined intervention phase and differentiation between sets was not observed. It is also noteworthy that student 9 obtained only one score in the mastery range for control set B in session 29, but he did not consistently reach mastery levels of performance for either set B or C.

The mean scores for sets B and C remained in the instructional range, with a slightly greater improvement observed for the set C ($M = 23.17; SD = 4.32$), which was assigned the tutor condition than for set B ($M = 20.82; SD = 7.20$), which was assigned the control condition, while the mean score for set A remained in the mastery range ($M = 42.55; SD = 6.08$). Additionally, DICPM scores remained much higher for sets B and C, but were at 0 for set A.
The slope for set A continued to indicate significant growth \( (b_i = 3.13) \). Furthermore, scores for set A (tuttee condition) remained visibly higher than scores for sets B and C. Consequently, differentiation was clearly present and the tuttee condition of the combined intervention was verified as effective for this student.

**Student 10.** Student 10 earned an initial baseline mean score of 20.05 (4.46) which fell in the instructional range (Burns et al., 2006). Baseline means for sets A \( (M = 20.83; \ SD = 3.55) \), B \( (M = 20.67; \ SD = 7.64) \), and C \( (M = 18.67; \ SD = 2.08) \) all fell in the instructional range. Baseline trends indicated that DCPM scores would continue to decrease for sets A \( (b_i = -1.21) \) and C \( (b_i = -1.14) \) in the absence of intervention. Although the slope for set B \( (b_i = 2.50) \) indicated that growth would occur without the implementation of an intervention, this trend was largely due to an outlier score in the final session of the baseline. A visual analysis of errors indicated that student 10 was fairly accurate when completing assessment probes, with DICPM scores ranging from 1 to 6.

Student 10 was paired with another student who completed the study, but whose scores were not reported due to excessive absences and because she failed to comply with directions when she was present and completing the probes. Student 10 began the peer tutoring intervention, which lasted 10 sessions as the tutor, and finished as the tuttee. For student 10, set A was assigned to the control condition, set B was assigned to the tutor condition, and set C was assigned to the tuttee condition. An analysis of error patterns indicated that DICPM scores dropped for all sets upon implementation of the peer tutoring intervention. Student 10 began making almost no errors across sets and sessions, with the exception of two sessions in which he made several errors on set A.
When considering fluency scores during the peer tutoring intervention, only set A (the control condition) showed an immediate improvement, and after two elevated data points, scores dropped back down to baseline levels. Although there was no initial improvement of scores for set B, a significant increase was observed during the fourth and fifth sessions of the tutoring condition as DCPM scores for this set were well above those observed for set A. Similarly, differentiation was observed between sets A and C as a considerable improvement in scores could be seen for set C throughout the implementation of the tutee condition, with 4 of the 5 data points falling above set A (the control condition). Further, the slope for set A continued to predict poorer outcomes in the absence of additional intervention ($b_i = -0.79$), while the slopes for sets B ($b_i = 4.75$) and C ($b_i = 3.40$) both predicted significant growth. Interestingly, the peer tutoring mean scores for sets A ($M = 20.30; SD = 5.56$) and B ($M = 20.40; SD = 8.45$) were similar to those found in the baseline phase, while the mean score for set C ($M = 24.40; SD = 7.57$), when the subject served as a tutee, showed improvement. All three mean scores remained in the instructional range.

When taken together the data for student 10 suggest that the tutee condition of the peer tutoring intervention was effective, as significant growth was observed and DCPM scores consistently fell above those achieved within the control set (set B) throughout the implementation of this condition. While the data suggest that the tutor condition may be effective, improved scores were not observed steadily enough, as three consistent scores are necessary to verify effectiveness (Horner et al., 2005) and only two consecutive scores showed improvement within this phase. As a result, it is impossible to infer that
the tutor condition was effective. Thus, the peer tutoring intervention was partially verified as effective when considering data from this set in isolation.

The combined intervention was implemented for 15 sessions. Student 10 began and ended this intervention phase as the tutor and served as the tutee for the five sessions in the middle. Visual analysis of student 10’s graph in Figure 4 revealed a pattern similar to that which was observed throughout the peer tutoring intervention. Differentiation was clearly present as scores for the sets B and C (tutor and tutee conditions respectively) remained higher than those in the control condition during most sessions. When considering mean scores for the combined intervention phase we saw considerable growth across all three sets, with greater improvement observed for sets B ($M = 38.65; SD = 8.91$) and C ($M = 40.15; SD = 10.16$), than for set A ($M = 27.97; SD = 9.20$). It is notable that the mean scores for sets B and C improved from the instructional range to the mastery range, while the mean for set A remained in the instructional range. Student 10 began to perform in the mastery range for set C during session 19, and for sets A and B during session 23. A visual analysis of the graph for student 10 in Figure 5 indicated that DICPM scores remained low throughout the combined intervention phase, with most scores being zero indicating no errors.

The trends for sets A ($b_l = 1.28$) and B ($b_l = 1.17$) showed similar rates of growth, while the trend for set C ($b_l = 4.00$), the tutee condition, showed a greater rate of improvement. Interestingly, DCPM scores showed improvement for set A, the control condition, during the combined intervention, which had not previously occurred. This change in trends indicated that the addition of performance feedback may have positively impacted student 10’s performance.
When considering the data for the combined intervention it could be concluded that both conditions were effective, as evidenced by increased scores and rates of improvement when the different roles of tutee and tutor were implemented. Further, scores for the control set A began to improve during the combined intervention, which provided evidence that the performance feedback component of the combined intervention was effective as well. Thus, the combined intervention was verified as effective for student 10.

**Student 11.** Student 11 earned an initial baseline mean score of 21.10 (4.41) which fell in the instructional range (Burns et al., 2006). Student 11’s baseline means for sets A ($M = 15.50; SD = 6.50$), B ($M = 13.30; SD = 3.87$), and C ($M = 14.00; SD = 2.35$) all fell in the frustrational range. An analysis of trends indicated that scores improved steadily for sets A ($b_i = 2.25$) and B ($b_i = 1.80$) suggesting that scores should continue to improve over time in the absence of intervention. However, scores for set C declined over time ($b_i = -.20$), indicating that scores would decrease over time unless an intervention was implemented. Visual analysis of errors for student 11 in Figure 5 indicated that DCIPM scores were erratic throughout the baseline phase as scores ranged from zero to 24 with no discernible trend.

Student 11 was paired with student 14 for the 10 sessions of the peer tutoring intervention. Student 11 began the intervention as the tutor and finished as the tutee. Set A was assigned to the tutor condition, set B was assigned to the tutee condition, and set C was assigned to the control condition. No noticeable change in DCPM or DICPM scores was observed for the control set C within this condition. The mean score for control set C
was similar to the baseline mean score, and slope indicated that relatively little growth occurred throughout this intervention phase \((b_i = .20)\).

Data from set A displayed a positive slope \((b_i = 2.10)\), which was similar to the slope observed within the baseline phase, and the mean score improved slightly as compared to the mean baseline score \((M = 20.20; SD = 7.85)\). DCPM scores did not immediately improve upon implementation of the peer tutoring intervention. After the initial intervention session, however, scores for set A, when the student served as tutor, increased considerably during the following three sessions. These three DCPM scores from the tutor condition were well above those observed for set C, the control condition. Additionally, scores for these sessions were above baseline levels to the degree predicted by baseline trends. However, DCPM scores decreased for the fifth and final tutor session. The erratic nature of the scores for set A made it difficult to determine whether or not differentiation was present. Therefore, we cannot attribute improved scores to intervention effects. Although it is noteworthy that DICPM scores for set A fell to 0 for the last three sessions of the tutor condition, improved accuracy did not coincide with consistently elevated DCPM scores. Thus, the tutor condition of the peer tutoring intervention could not be verified as effective.

No observable change for DCPM scores was present for set B upon implementation of the tutee condition; however, DICPM scores fell to 0. The intervention mean for set B \((M = 16.00; SD = 4.83)\) was indicative of minimal improvement and remained in the instructional range. Additionally, although set B displayed a positive slope \((b_i = 1.94)\), this was largely due to a decrease in DCPM scores at the start of the tutee phase, and scores did not exceed those observed within the
baseline phase. Thus, the tutee condition of the peer tutoring phase was not verified as effective for this student.

Student 11 participated in the combined intervention for 15 sessions. She began and ended the intervention phase serving as the tutor and acted as the tutee for the five middle sessions. The intervention mean for sets A ($M = 25.40; SD = 6.69$), B ($M = 25.40; SD = 2.88$), set C ($M = 17.83; SD = 3.59$) were all higher than the baseline and peer tutoring means, although scores for set C showed only minimal improvement. The intervention means remained in the instructional range for all three sets. When considering slopes, little to no improvement was observed across sets, with minimal growth observed for the tutor set A ($b_i = .33$) and a decrease in scores observed for sets B ($b_i = -.30$) and C ($b_i = -.14$). Visual analysis of the graph for student 11 in Figure 5 indicated that most DICPM scores remained at 0 throughout the combined intervention phase for the tutee and tutor conditions. While DICPM scores remained erratic for the control set C, a generally decreasing trend was observed.

Although mean scores improved, no immediate increase in scores was observed for any of the three data sets upon the implementation of the performance feedback condition as part of the combined intervention. Further, scores never reached mastery levels for any of the sets. However, during the third session of the tutor condition an increase was observed for set A. Consequently, differentiation was observed as student 11 continued to show elevated scores for set A whenever the tutor condition was implemented. Interestingly, scores for set B increased upon the implementation of the tutee condition to the extent that scores fell above those observed for the control set C. Scores continued to be erratic throughout the combined intervention phase, with scores
for both the tutee and tutor conditions falling higher than those for the control condition throughout most of the combined intervention phase. Thus, differentiation was established for both conditions of the combined intervention, with a more pronounced difference observed between sets A (tutor condition) and C (control condition). When taken together, the data suggested that both the tutor and tutee conditions of the combined intervention were verified as effective for student 11.

**Student 12.** Student 12 earned an initial baseline mean score of 19.07 (6.03) which fell in the instructional range (Burns et al., 2006). The baseline mean for sets A ($M = 30.25; SD = 8.72$), B ($M = 23.65; SD = 8.68$), and C ($M = 21.70; SD = 10.21$) all fell in the instructional range. Set A was clearly easier than sets B and C for this student. An analysis of trends indicated that scores improved steadily for all three sets throughout the baseline phase ($b_i = 2.58, 2.26, \text{ and } 2.68$ for sets A, B, and C respectively). These trends predicted that improvement would continue to occur whether or not an intervention was implemented. In fact, improvement was such throughout the baseline phase that student 12 reached mastery levels of performance for set A during session 5, for set B during session 7, and for set C during session 9.

Student 12 was paired with student 13 for the peer tutoring intervention. He began and ended the intervention as the tutor and acted as the tutee during the middle five sessions. For student 12, set A was assigned to the control condition, set B was assigned to the tutor condition, and set C was assigned to the tutee condition. Consistent improvement for student 12 occurred across all three sets regardless of the intervention condition that was implemented. The rate of improvement was similar to the growth observed in the baseline phase for set B ($b_i = 2.29$) but it was slower for set A ($b_i = 1.30$),
the control condition, and faster for set C \((b_i = 3.60)\), when he served as tutee. A visual analysis of errors showed that DICPM scores remained mostly at 0, which was also observed for the final sessions of the baseline phase.

All three peer tutoring means were greater than baseline means, as the mean scores for sets A \((M = 52.89; SD = 11.61)\), B \((M = 44.87; SD = 16.50)\), and C \((M = 50.95; SD = 5.92)\) all improved from the instructional range to the mastery range.

Although student 12 showed significant improvement over time, improvement did not occur at a faster rate upon implementation of the peer tutoring intervention for any of the sets, with the exception of set C. Further, visual analysis of student 12’s graph in Figure 4 indicated that the trajectory of improvement observed for the control condition was nearly identical to that observed for sets B and C. Finally, scores for the intervention sets B and C did not exceed those observed for the control set A. Thus, differentiation was not established and the peer tutoring intervention was not verified as effective.

Student 12 received the combined intervention for five sessions and acted as the tutee throughout this intervention phase. No significant improvement was observed for sets A or C. Additionally, scores were erratic within this phase, and growth stagnated for sets A \((b_i = -.08)\) and C \((b_i = -.48)\), which is noteworthy given that set C was associated with the tutee condition. It is possible that scores were impacted by ceiling effects, as student 12 had already begun performing well within the mastery range. The mean scores for sets A \((M = 59.15; SD = 13.18)\) and C \((M = 54.75; SD = 6.62)\) both fell within the mastery range. DICPM scores remained equal to or close to zero throughout this intervention phase. Therefore, there is no evidence to suggest that the combined
intervention was effective, as differentiation was not present and DCPM scores showed little to no improvement upon implementation of the intervention.

In sum, student 12 made significant gains throughout the course of the study. Baseline trends, however, predicted that scores would improve over time even in the absence of intervention. Based on these data, neither intervention was verified as effective, but his gains may have been due to repeated practice. Implications of these findings will be discussed in the next chapter.

**Student 13.** Student 13 earned an initial baseline mean score of 14.26 (4.34) which fell in the instructional range (Burns et al., 2006). The baseline mean for sets A ($M = 27.00; SD = 8.69$), B ($M = 19.30; SD = 4.83$), and C ($M = 17.65; SD = 3.00$) all fell in the instructional range. Set A was clearly easier than sets B and C for student 13. Trends predicted that scores would increase for set A ($b_i = 2.40$), decrease for set C ($b_i = -1.01$), and remain relatively stable for set B ($b_i = .08$). It is of note that improvement for set A was such that student 13 began performing in the mastery range during session 8. Visual analysis of the error graph for student 13 in Figure 5 indicated that DICPM scores were erratic as they ranged from 1 to 23.5. Error rates decreased, however, across all three sets throughout the baseline condition.

Student 13 was paired with student 10 for the 15 sessions of the peer tutoring intervention. She began and ended the intervention as the tutee and was the tutor for the five middle sessions. Set A was assigned to the control condition, set B the tutee condition, and set C was assigned to the tutor condition. Significant growth was observed for all three sets across the peer tutoring intervention phase. However, visual analysis of the graph for student 13 in Figure 4 revealed that scores for set A remained
higher than scores for sets B and C, and immediate improvement did not appear to occur upon introduction of either the tutee or tutor condition. Although slopes for sets B ($b_i = 2.53$) and C ($b_i = 3.30$) indicated that improvement for these sets occurred at a faster rate than set A ($b_i = 1.71$), the slope for set A was negatively impacted by an outlier score at the end of the intervention phase.

The peer tutoring mean for the control set A ($M = 45.33; SD = 12.58$) increased from the instructional range to the mastery range, while the means for sets B ($M = 28.45; SD = 15.05$) and C ($M = 22.40; SD = 9.18$) remained in the instructional range. However, student 13 began performing in the mastery range for set B during session 22. Visual analysis of DICPM the graphs for student 13 in Figure 5 indicated that error scores continued to remain erratic but, as predicted from baseline trends, a general decrease in errors could be observed throughout the intervention phase. As improvement was more significant for the control condition than it was for the peer tutoring conditions, differentiation was not observed and experimental control was not demonstrated. Thus, the peer tutoring intervention could not be verified as effective.

Given that student 13 only received the combined intervention for five sessions, she only acted as the tutor. As a result, scores for set C (tutor condition) and A (control condition) were reported, but scores for set B were not reported for this intervention phase. Upon implementation of the combined intervention, slopes for sets A ($b_i = 6.73$) and C ($b_i = 7.58$) indicated a faster rate of improvement than was observed within the peer tutoring intervention. Further, mean scores for sets A ($M = 71.44; SD = 8.69$) and C ($M = 50.94; SD = 10.04$) both fell in the mastery range and were considerably higher than those that were previously observed. Student 13 began to achieve mastery levels for set
C in session 26. Although scores were lower for set C (the tutor condition), the overall pattern of scores might suggest that the additional component of performance feedback as part of the combined intervention was helpful for student 13. However, visual analysis indicated a similar trajectory for set B at the end of the peer tutoring intervention, suggesting that perhaps growth rates for student 13 were random and could not be attributed to intervention effects. As a result, the marked improvement displayed by student 13 could not be attributed to intervention effects for the peer tutoring or combined interventions. Implications of these findings will be discussed in the next chapter.

Summary for students in group B performing at the instructional level. As with group A, the students performing in the instructional range within group B showed variable results with regard to the effectiveness of the interventions. Within the context of the peer tutoring intervention the tutee condition was verified to be effective for 2 of the 6 students, while the tutor condition was verified as effective only for student 8. Horner et al. (2005) note that in order for an intervention to be verified as effective, there has to be instances of change for at least three subjects. In this case, neither condition was verified as effective three times, indicating that the peer tutoring intervention as not verified as an effective intervention for this group.

Conversely, the combined intervention was verified to be effective for 2 of the 6 students, as the rate of improvement increased for all three sets as compared to the peer tutoring intervention. Furthermore, the tutee condition of the combined intervention was verified for student 9, and the tutor condition was verified to be effective for student 8. Thus, replication was observed three times for each condition, which means that the combined intervention was verified as an effective intervention for this group. Though, it
is noteworthy that both conditions were not effective for all students. Implications of these findings will be presented in the next chapter.

There were other noteworthy patterns within this group. First, even when growth could not be attributed to intervention effects, with the exception of student 11, all students showed marked improvement over the course of the study, as 5 of 6 students began to consistently perform in the mastery range. Student 8 earned average scores in the mastery range for all three sets in the combined intervention phase, while student 12 earned average scores in the mastery range during the peer tutoring intervention. Student 10 achieved average scores in the mastery range for the sets assigned to the tutee and tutor conditions within the combined intervention. Student 12 earned scores in the mastery range for both the control set and the tutor set in the combined intervention. Finally, student 9 earned a mean score in the mastery range for his tutee set. Further, 4 of the 6 students began to consistently perform within the mastery range for all three sets throughout the course of the intervention phases. Student 11 was the only student in this group who did not achieve any mastery level scores. These findings will be discussed further in chapter 5.

Another pattern observed is that all 6 students within this group showed significant growth for at least one set within the baseline phase. This finding is interesting, as most students’ performance continued to improve through the course of the study. The third pattern was that all of the students’ accuracy improved for at least one set over the course of the study. Four of the 6 students showed improvement across all sets, although some improved at a faster rate for sets assigned to the tutee and/or tutor condition. Student 9 showed significant improvement with regard to accuracy for the
tutee phase, while student 11 showed significant improvement for the sets assigned to the peer tutoring conditions, but not for the control set. Interestingly, student 11 did not show improved fluency scores. Further implications of these findings will be discussed in the following chapter.

**Group B (continued): Peer tutoring for students at the frustrational level.**

Students 14 through 18 received an initial baseline average score that fell in the frustrational range (Burns et al., 2006). All students in this group first received the peer tutoring then the combined intervention (which included the added element of performance feedback) following baseline procedures.

**Student 14.** Student 14 earned an initial baseline mean score of 10.77 (3.45) which fell in the frustrational range (Burns et al., 2006). Student 14’s baseline means for sets A \( (M = 13.20; SD = 2.95) \), B \( (M = 8.70; SD = 4.02) \), and C \( (M = 10.40; SD = 1.95) \) all fell in the frustrational range. An analysis of trends demonstrated that scores improved steadily for set A \( (b_i = 1.10) \), suggesting that scores should continue to improve over time in the absence of intervention. However, slopes for sets B \( (b_i = .25) \) and C \( (b_i = -.10) \) indicated that significant improvement would not occur in the absence of intervention. Visual analysis of errors for student 14 in Figure 5 indicated that DICPM scores were erratic throughout the baseline phase as scores ranged from 2 to 15 with no discernible trend.

Student 14 was paired with student 11 for the peer tutoring intervention. This intervention was implemented for 10 sessions. Student 14 began the intervention as the tutee and finished as the tutor. Set A was assigned to the tutee condition, set B was assigned to the tutor condition, and set C was assigned to the control condition. DCPM
scores for all sets decreased upon implementation of the peer tutoring intervention.

Intervention means for sets A (M = 12.40; SD = 2.56), B (M = 7.00; SD = 2.12), and C (M = 5.00; SD = 2.42) were all lower than those observed during baseline levels and mean scores remained in the frustrational range. Additionally, slopes for the intervention phase were negative for all sets (bi = -.40, -1.10, and -.32 for sets A, B, and C respectively) indicating that DCPM scores decreased over time. Visual analysis of the graph for student 14 in Figure 4 indicated that scores for sets A and B were slightly higher than those observed for control set C. However, DCPM scores for set A were also higher than they were for Set C throughout the baseline phase, making it impossible to attribute this small difference to intervention effects. Additionally, the slight difference observed between sets B and C is not consistent or large enough to draw any definitive conclusions regarding intervention effects. Thus differentiation between sets was not observed. Finally, visual analysis of the graph for student 14 in Figure 5 indicated that DICPM scores continued to be erratic, but generally decreased over time. However, the decrease in errors appeared to be related to a decrease in attempted problems, as DCPM scores also decreased. When taken together, the peer tutoring intervention could not be verified as effective for student 14.

Student 14 participated in the combined intervention for 15 sessions. He began and ended the intervention phase serving as the tutee and acted as the tutor for the five middle sessions. The intervention means for sets A (M = 19.30; SD = 9.70), and B (M = 19.20; SD = 9.83), that were assigned to the tutee and tutor conditions respectively, showed considerable growth, as both improved from the frustrational range to instructional range. The mean for Set C (M = 5.13; SD = 3.16), the control condition, was
similar to that which was observed during the peer tutoring intervention, but it was considerably lower than the baseline mean. The mean score for set C remained in the frustrational range. DICPM scores for set C were lower than baseline levels, but similar to those observed in the peer tutoring intervention. Additionally, the slope for the control set C was relatively flat \( b_i = .08 \). Student 14 did not consistently reach mastery levels for any set, but he did earn one score in the mastery range for set A and one for set B.

The mean score improved for set A, and although the slope indicated only minor growth across this intervention phase \( b_i = .51 \), it was greatly impacted due to erratic scores. More specifically improvement was observed for several sessions of the tutee phases, but scores peaked at session 26 and showed a sharp decline over the course of the last four sessions, with the final score falling back at baseline levels. Nonetheless, differentiation was clearly established between set A and C as all but one data point for set A fell well above those of the control set C. Additionally, visual analysis of the graph for student 14 in Figure 5 DICPM showed that scores continued to decrease upon implementation of the combined intervention for set A, as scores quickly fell to 0. Although a consistent upward trend is preferred, the data suggest that the tutee condition of the combined intervention was effective for student 14.

Conversely, the slope for set B \( b_i = 4.40 \) suggests significant growth occurred within throughout the tutee condition of the combined intervention. Visual analysis of the graph for student 14 in Figure 4 confirmed that scores for set B were higher than those that were observed in previous phases, and a large increase in scores was observed in the final session of the tutor condition. Furthermore, differentiation was clearly established as scores for set B consistently fell above those of the control set C. Finally,
DICPM scores for set B showed less variability and were generally lower than those that were observed in prior phases of the study. When taken together, the tutor condition of the combined intervention could be verified as effective. Thus, the combined intervention could be verified as effective for student 14.

**Student 15.** Student 15 earned an initial baseline mean score of 4.26 (4.34) which fell in the frustrational range (Burns et al., 2006). Student 15’s baseline means for sets A ($M = 12.50; SD = 3.08$), B ($M = 12.50; SD = 3.65$), and C ($M = 13.63; SD = 3.35$) all fell in the frustrational range as well. An analysis of trends indicated that scores would likely decrease for all three sets in the absence of intervention ($b_i = -1.50, -1.40, \text{ and } -0.25$ for sets A, B, and C respectively). Visual analysis of errors for student 15 in Figure 5 indicated that DICPM scores were erratic throughout the baseline phase as scores ranged from 0.5 to 8 with no discernible trend.

Student 15 was paired with student 16 for the peer tutoring intervention. This intervention was implemented for 10 sessions. Student 15 began the intervention as the tutee and finished the intervention as the tutor. Set A was assigned to the tutor condition, set B was assigned to the tutee condition, and set C was assigned to the control condition. DCPM scores were erratic throughout the intervention phase, which made assessing trends problematic. For example, the slope for set A was negative ($b_i = -1.35$), yet the greatest gains were observed for set A when considering the peer tutoring mean for this set ($M = 14.90; SD = 5.18$). Further, the intervention mean for set B ($M = 11.00; SD = 4.85$) was lower than the baseline mean, but the slope for this set ($b_i = 1.30$) was indicative of faster rate of improvement than was observed in the baseline phase. Finally, the mean for set C ($M = 7.55; SD = 5.29$) was also lower than the baseline mean, and the
slightly positive slope ($b_i = .34$), was largely due to one outlier score that fell above baseline levels.

Visual analysis of the graph for student 15 in Figure 5 indicated that DICPM generally decreased after the first intervention session with scores ranging from 0-2. Visual analysis of the graph for student 15 in Figure 4 indicated that differentiation between sets A and C was present as the DCPM scores for set A were clearly higher than set C when the tutee condition was implemented. Further, although growth was not consistently above baseline levels for set A, improvement was observed for 2 of the 5 data points. Because differentiation was established, the tutee condition of the peer tutoring condition was verified as effective. However, because gains were minimal as compared to baseline scores, the practical utility of this intervention is called into question. This finding will be discussed further in the following chapter.

Although several scores for set B were higher than set C when the tutor condition was implemented, variability in performance and overlap between the two sets makes differentiation impossible to establish. Additionally, scores did not consistently exceed baseline levels, thus, the tutor condition of the peer tutoring intervention could not be verified as effective.

Student 15 participated in the combined intervention for 14 sessions. She began and ended the intervention phase serving as the tutee and acted as the tutor for the four middle sessions, as she was absent during session 23. The intervention mean for set A ($M = 15.45; SD = 8.70$) improved from the frustrational range to the instructional range, but the means for sets B ($M = 9.50; SD = 8.35$) and C ($M = 6.00; SD = 6.28$) remained in the frustrational range as they were both lower than baseline and peer tutoring means.
Further, an analysis of trends indicated that growth was not observed for any of the three sets ($b_i = -0.29, 0.00, \text{ and } -0.10$ for sets A, B, and C respectively). However, student 15 continued to perform erratically, which makes assessment of trends impossible. When considering errors, DICPM scores remained low throughout the combined intervention with scores mostly ranging from 0-3, with two outlier scores of 7.5, indicating an overall level of improved accuracy.

Visual analysis of the graph for student 15 in Figure 4 indicated that differentiation was observed between the tutee and control condition as scores for set A (tutee condition) were higher than the control set C for most intervention sessions. However, contrary to what was observed throughout the peer tutoring intervention phase, data points began to exceed previous levels, as four scores for set A fell above baseline levels. When considering these data points in conjunction with the improved mean score, it seems clear that the tutee condition of the combined intervention could be verified as effective for student 15.

Conversely, there was not enough data to establish differentiation between sets B and C, as only two scores for set B exceeded those of set C. Further, gains were minimal as compared to previous phases of the study. Consequently, the tutor condition of the combined intervention could not be verified as effective.

**Student 16.** Student 16 earned an initial baseline mean score of 12.63 (2.87) which fell in the frustrational range (Burns et al., 2006). Student 16’s baseline means for sets A ($M = 9.63; SD = 3.50$), B ($M = 9.50; SD = 5.45$), and C ($M = 10.88; SD = 5.36$) all fell in the frustrational range as well. Analysis of trends indicated that improvement would occur for all three sets in the absence of intervention ($b_i = 2.65, 4.20, \text{ and } 3.55$ for
sets A, B, and C respectively). Visual analysis of errors of the graph for student 16 in Figure 5 indicated that DICPM scores increased throughout the course of the baseline phase.

Student 16 was paired with student 15 for the peer tutoring intervention. Student 16 began the 10-session intervention as the tutor and finished as the tutee. Set A was assigned to the tutee condition, set B was assigned to the tutor condition, and set C was assigned to the control condition. Visual analysis of the graph for student 16 in Figure 5 indicated that DICPM scores became erratic with no discernible trend and scores ranging from 0 to 13.

Within the peer tutoring phase, mean scores for sets A ($M = 18.25; SD = 4.65$) and B ($M = 18.75; SD = 10.18$) showed considerable improvement, as they now fell in the instructional range, while the mean score for set C ($M = 10.50; SD = 5.48$) remained essentially the same as baseline mean scores. Due to erratic scores, trends were difficult to assess. The slope for set A was flat ($b_i = 0.03$), while slopes for sets B ($b_i = -2.30$) and C ($b_i = -1.08$) indicated a decrease in scores throughout this intervention phase.

Within the peer tutoring intervention phase, differentiation was clearly present between the control set C and the targeted sets A and B. DCPM scores for sets A and B were consistently higher than those for set C, with the exception of the first session of the tutor condition in which scores for sets A and C were the same. Thus, intervention effects were confirmed and both the tutee and tutor conditions were verified as effective for the peer tutoring condition. It is noteworthy that growth from the baseline phase to the intervention phase was minimal at best, which again calls the functional utility of this
intervention into question. Implications of this finding will be explored in the following chapter.

Student 16 participated in the combined intervention for 15 sessions. He began and ended the intervention phase serving as the tutor and acted as the tutee for the five middle sessions. The combined intervention mean for set A ($M = 12.39; SD = 4.23$) fell below the peer tutoring mean, while the means for sets B ($M = 21.60; SD = 5.13$) and C ($M = 12.21; SD = 7.54$) both showed slight improvement. Only the mean score for set B remained in the instructional range. The mean score for set A returned to the frustrational range and the mean score for set C stayed in the frustrational range. Student 16 did not reach mastery levels of performance for any of the three sets. Further, an analysis of trends indicated that slight growth was observed sets A ($b_i = .46$) and B ($b_i = .50$), while scores decreased over time for set C ($b_i = -.30$). Visual analysis of the graph for student 16 in Figure 5 indicated that DICPM scores ranged from 0 to 16 with no discernible pattern.

As we have seen for students starting in the frustrational range, student 16 continued to perform erratically. As a result, differentiation between sets could not be clearly established. Scores for the targeted sets A and B consistently overlapped with those observed for the control set C throughout the first 10 sessions of this intervention phase. Although a pattern emerged in which scores for set A fell above those for set C throughout the last five sessions of the intervention phase, scores for set A fell below scores for set C that were observed throughout earlier sessions of the intervention phase. Thus, the combined intervention was not verified as effective for this student.
**Student 17.** Student 17 earned an initial baseline mean score of 12.63 (2.87) which fell in the frustrational range (Burns et al., 2006). For student 17, the baseline means for sets A (M = 13.35; SD = 2.76), B (M = 9.90; SD = 1.52), and C (M = 11.10; SD = 3.36) all fell in the frustrational range as well. Slopes for all three sets predicted that improvement would not occur without intervention, as the slopes for sets A (b_i = -.45) and C (b_i = -.93) predicted that scores would decrease over time, while the slope for set B (b_i = .04) predicted that scores would remain the same. Visual analysis of the graph for student 17 in Figure 5 indicated that she struggled with accuracy across sets as DICPM scores ranged from 6 to 17.5 with no discernible pattern.

Student 17 was paired with student 18 for the peer tutoring intervention. This intervention was implemented for 10 sessions. Student 17 began the intervention as the tutor and finished as the tutee. Set A was assigned to the tutor condition, set B was assigned to the control condition, and set C was assigned to the tutee condition.

Within the peer tutoring phase, mean scores for sets A (M = 13.40; SD = 2.51), B (M = 13.10; SD = 3.66), and C (M = 13.50; SD = 2.45) were all similar to baseline means, with slight improvement observed for set B, the control condition. All three remained in the frustrational range. Additionally, slopes for all three sets were negative, indicating that scores did not improve throughout the course of the intervention (b_i = -.70, -.68, -.30 for sets A, B, and C respectively). Finally, differentiation was not established as scores for all sets were similar throughout the intervention phase.

Conversely, improvement was observed in terms of accuracy when considering error patterns. Upon implementation of the tutor condition, DICPM scores for set A dropped significantly and became more consistent with scores ranging from 0 to 4.
Conversely, DICPM scores for set B, the control condition, remained high and erratic, with scores ranging from 8 to 23. Upon implementation of the tutee condition, DICPM scores began to decrease for both sets B and C, with scores in the final session of 4 and 2, respectively. However, the improvements in accuracy did not coincide with increased fluency scores. Thus, the peer tutoring intervention was not verified as effective for student 17.

Student 17 participated in the combined intervention for 10 sessions. She began the session as the tutor and ended as the tutee. The combined intervention means for sets A \((M = 18.00; SD = 11.14)\), B \((M = 20.70; SD = 6.97)\), and C \((M = 28.90; SD = 9.07)\) all showed considerable improvement from the frustrational to the instructional range. An analysis of trends indicated that student 17 showed significant improvement throughout the course of the combined intervention for sets A and B \((b_i = 6.10, \text{ and } 2.70, \text{ respectively})\) and a slower rate of improvement for set C \((b_i = .33)\). Visual analysis of the graph for student 17 in Figure 4 indicated that this slowed rate of improvement was due to the extreme variability in DCPM scores that was observed for set C throughout the intervention phase. DICPM scores for the combined intervention phase were comparable to those observed within the peer tutoring phase, with highest rates of accuracy continuing for set A. Overall, DICPM scores decreased from the baseline phase.

Visual analysis revealed some interesting patterns throughout the combined intervention. First, greater gains in DCPM scores were observed for the targeted sets A and B as compared to the control set C, indicating that combining the peer tutoring and performance feedback interventions may have been the most effective methods for this student. Additionally, trends indicated that scores for sets A and C were beginning to
climb higher than those for the control set C. However, differentiation could not be confirmed, as there were not enough data points to replicate this finding. Yet, improvements were observed for all three sets upon the implementation of the performance feedback condition of the combined intervention. Thus, the combined intervention was verified as partially effective, although conclusions could not be drawn about whether the tutor or tutee condition were more effective than the control condition. Implications of this finding will be discussed in the next chapter.

**Student 18.** Student 18 earned an initial baseline mean score of 10.53 (3.31) which fell in the frustrational range (Burns et al., 2006). For student 18, the baseline means for set A \( (M = 14.55; SD = 4.91) \) fell in the instructional range, while the mean scores for sets B \( (M = 11.05; SD = 5.52) \), and C \( (M = 11.00; SD = 3.78) \) fell in the frustrational range. Slopes for all sets predicted that improvement would occur without intervention, with slopes for sets A \( (b_i = 1.06) \) and B \( (b_i = .93) \) predicting slightly faster rates of improvement than the slope for set C \( (b_i = .38) \). Visual analysis of the graph for student 18 in Figure 5 indicated that he struggled with accuracy across sets as DICPM scores ranged from 1 to 24 with no discernible pattern.

Student 18 was paired with student 17 for the peer tutoring intervention. This intervention was implemented for 10 sessions. Student 18 began the intervention as the tutee and finished as the tutor. Set A was assigned to the tutee condition, set B was assigned to the control condition, and set C was assigned to the tutor condition.

Within the peer tutoring phase, mean scores for sets A \( (M = 14.30; SD = 2.49) \), B \( (M = 10.20; SD = 4.37) \), and C \( (M = 10.80; SD = 2.56) \) were all similar to baseline means. Additionally, only set A yielded a slope that would indicate growth \( (b_i = 1.40) \). Set B
showed slight improvement throughout the course of the intervention ($b_i = .40$), while scores for set C remained relatively stable ($b_i = .10$). Although the slope for set A indicated growth, scores did not improve above baseline levels. It is of note that, as with scores for the baseline phase, scores for set A were considerably higher than those of control set C. However, as the magnitude of this difference was consistent, clear differentiation could not be established. Additionally, DICPM scores remained erratic with scores ranging from 0 to 21. Thus, the peer tutoring intervention was not verified as effective for this student.

Student 18 participated in the combined intervention for 10 sessions. He began as the tutee and ended as the tutor. The combined intervention means for sets A ($M = 14.40; SD = 6.35$) and B ($M = 12.90; SD = 4.64$) remained relatively similar to previous mean scores, while the mean score for set C ($M = 17.50; SD = 5.28$), when he was the tutor, showed considerable improvement from the frustrational to the instructional range. However, student 18 never reached mastery levels of performance. An analysis of trends indicated that student 18 showed a pattern similar to that which was observed during the peer tutoring phase. The slope for set A ($b_i = 1.70$) was indicative of considerable improvement throughout the course of the intervention phase, while the slope for set B ($b_i = .30$), the control, showed minimal improvement, and the slope for set C ($b_i = .10$), the tutor condition, showed almost no improvement. These scores indicated that the performance feedback component of the combined intervention did not have positive effects. DICPM scores for the combined intervention phase remained erratic, but an overall decreasing trend was observed for set B. Scores generally ranged from 0 to 14.
A visual analysis of the graph for student 18 in Figure 4 yielded some interesting patterns. DCPM scores for set A showed a decrease, followed by a rapid increase during the implementation of the tutee condition. However, this improvement was predicted by trends in previous phases and not consistent enough to be indicative of treatment effectiveness. It is notable that scores for set C, the tutor condition, were consistently above scores for set B. When considering this pattern in conjunction with the improved mean score for set C, one might consider the tutor condition of the combined intervention to be effective. However, improvement was not considerable and scores were erratic for all sets throughout the course of the study. More data would be necessary to draw sound conclusions regarding differentiation and intervention effectiveness. Thus, the combined intervention was not verified as effective for student 18.

**Summary for students in group B performing at the frustrational level.** When analyzing the performance of the 6 students who initially performed at the frustrational range and received the peer tutoring intervention in isolation, the tutee condition of the peer tutoring intervention was verified as effective for student 15 and both the tutee and tutor conditions were verified as effective for student 16. As neither condition was verified as effective three times (Horner et al., 2005), the peer tutoring intervention was not verified as effective for this group. It is also noteworthy that these conditions were verified as effective due to differentiation within the intervention phase and scores did not increase significantly as compared to baseline levels. Implications of these findings will be discussed in the next chapter.

Similarly, the combined intervention was verified as effective for individual students but not for the group as a whole. It was effective for student 14 for both the
tutor and tutee conditions, and the tutee condition of the combined intervention was verified as effective for student 15. Furthermore, although differentiation did not take place within the combined intervention for student 17, scores the data path changed for all three sets upon the implementation of the combined intervention. As intervention effects could not be attributed to the tutee or tutor conditions, this change in trajectory indicated that the added component of performance feedback was effective for student 17 and the combined intervention was partially verified as effective in this way. When considering the data, the combined intervention was not verified as effective for this group, as replication did not occur three times for any of the conditions.

Within this group 3 of the 5 students displayed slopes during the baseline phase for at least one set that indicated growth would occur in the absence of intervention, while two did not. Additionally, an improvement in accuracy was observed for students 15 and 16 across all three sets, while student 17 showed an improvement within the tutor condition. However, increased accuracy did not always translate into increased fluency scores. Implications of these patterns will be discussed in the next chapter.

**Summary of the data for group B.** The results of the peer tutoring intervention were mixed for students performing at both the instructional and frustrational levels. As with group A, the students performing in the instructional range within group B showed variable results with regard to the effectiveness of the interventions. When considering the peer tutoring intervention for group B, 1 of 11 students responded positively to both the tutee and tutor condition, and as previously noted gains were small. Additionally, the tutee condition was found to be effective for students 3 of 11 students and the tutor
condition was effective only for student 8. When taken together, only the tutee condition of the peer tutoring intervention was verified as effective for group B.

When considering all students for group B, the combined intervention was verified as effective for 3 of 11 students. Additionally, the tutee condition of the combined intervention was verified as effective for 2 of 11 students, while the tutor condition was verified as effective only for student 8. Additionally, the performance feedback component was identified as effective for student 17. Thus, the combined intervention was verified as an effective treatment for students in group B. The implications of this finding will be discussed in the next chapter.

There were other noteworthy patterns within the entire group. First, even when growth could not be attributed to intervention effects, with the exception of student 11, all students in the instructional range showed marked improvement over the course of the study, as 9 of 10 students earned scores in the mastery range before the conclusion of the study. Conversely, only student 14 reached mastery levels of performance, and he did so inconsistently. Further, visual analysis indicated that improvement for students at the frustrational level did not achieve the same magnitude of growth, meaning that DCPM scores for the instructional groups typically increased at a faster rate and gains were greater than was observed for the frustrational groups.

Another pattern observed is that 12 of 18 students (9 in the instructional group and 3 in the frustrational group) displayed growth for at least one set within the baseline phase. Further discussion of this finding is necessary to understand the impact this may have had on student performance throughout intervention phases. A final pattern suggested that the accuracy improved for 11 if 18 students (7 from the instructional group...
and 4 in the frustrational group) for at least one set over the course of the study. Further implications of these findings will be discussed in the following chapter.

**Summary of Individual Analysis of Groups A and B.** Several patterns were evident when looking at the data across both groups. These patterns are displayed in Table 14. As previously noted, the performance feedback intervention in isolation was only verified as effective for 1 of 7 students. Thus, replication was not present and this intervention was not effective for this group of students. Conversely, both the tutee and tutor conditions of the peer tutoring intervention in isolation were verified to be effective for 1 of 11 students, the tutee condition was verified to be effective for students 3 of 11 students, and the tutor condition was verified to be effective only for student 8. Consequently, only the tutee condition could be verified as effective for group B, as the tutor condition was only found to be effective twice.

The combined intervention was verified as effective for students 4 of 18 students for both conditions, while 4 more students showed improvement as the result of the tutee condition of the combined intervention. Further, the tutor condition was verified as effective for student 8. Finally, the performance feedback component of the combined intervention was found to be effective for student 17, although differentiation was not present within this phase and conclusions could not be drawn regarding the effectiveness of the peer tutoring conditions. In sum, the combined intervention was verified as effective for students with stronger support for the tutee condition of this intervention. As a result, performance feedback might be an important component of helping students improve. These findings will be explored further in an attempt to understand why these interventions were effective for some students and not for others in the fifth chapter.
Several other important patterns were observed throughout the study. First, 9 of 10 students in the instructional range showed marked improvement over the course of the study, while no students in the frustrational range showed significant improvement. No improvement was observed for students in group A who were performing in the frustrational range, while minimal improvement was observed for students in group B who were performing at the frustrational range.

When considering baseline trends, an interesting pattern emerged in that 9 of the 10 students performing at the instructional level displayed considerable growth for at least one set during the baseline phase, while only 3 of 8 students in the frustrational range displayed this pattern. That is, 5 students performing at the frustrational range and 1 student at the instructional range displayed slopes that indicated no growth would occur in the absence of intervention.

A final pattern suggested that the accuracy improved for 13 of 18 students (9 from the instructional group and 4 in the frustrational group) for at least one set over the course of the study. Further implications of these findings will be discussed in the following chapter.

As noted throughout, 9 of 10 students in the instructional range achieved mastery levels of performance at various times throughout this study. Conversely, only 1 of 8 students in the frustrational range was able to inconsistently perform in the mastery range. Based on the data reported here, it seemed that DCPM scores at the start of the study greatly influenced if students were able to achieve this goal. This finding will be discussed in more detail in a later section of this chapter.
In sum, the results of the individual analysis led to several findings. First, repeated practice appeared to be beneficial for students performing in the instructional range. Second, when considering the nature of the interventions, peer tutoring was effective for more students than the performance feedback intervention, while the combined intervention appeared to be effective for more students than either of the individual interventions. Further, more students responded positively to the tutee condition than the tutor condition, however a select few responded better to the tutor condition. These findings will be further discussed in later sections of this chapter.
### Summary of Findings from the Individual Analysis

<table>
<thead>
<tr>
<th>Student</th>
<th>Intervention Verified as Effective</th>
<th>Accuracy Improved</th>
<th>Improved DCPM Scores</th>
<th>BL Slope $&gt;1.0$</th>
<th>Mastery Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PF/Tutor</td>
<td>PT/Tutor</td>
<td>Combined</td>
<td>PF/Not Intervention</td>
<td>PT/Combined</td>
</tr>
<tr>
<td>1 (AI-1)</td>
<td></td>
<td></td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>2 (AI-2)</td>
<td></td>
<td></td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>3 (AI-3)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>4 (AI-4)</td>
<td></td>
<td></td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>5 (AF-1)</td>
<td></td>
<td></td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>6 (AF-2)</td>
<td></td>
<td></td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>7 (AF-3)</td>
<td></td>
<td></td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>8 (BI-1)</td>
<td></td>
<td></td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>9 (BI-2)</td>
<td></td>
<td></td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>10 (BI-3)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>11 (BI-4)</td>
<td></td>
<td></td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>12 (BI-5)</td>
<td></td>
<td></td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>13 (BI-6)</td>
<td></td>
<td></td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>14 (BF-1)</td>
<td></td>
<td></td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>15 (BF-2)</td>
<td></td>
<td></td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>16 (BF-3)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>17 (BF-4)</td>
<td></td>
<td></td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>18 (BF-5)</td>
<td></td>
<td></td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
</tbody>
</table>

**Note.** PF = Performance Feedback; PT = Peer Tutoring; Intervention deemed effective if DCPM scores increased as a result of intervention effects; Accuracy Improved = DICPM scores decreased; PF/Not Due to Intervention = Decrease in DICPM scores for any set that did not occur within the PT or combined interventions; Improved DCPM Scores = DCPM scores increased, but not due to intervention conditions. BL Slope $>1.0$ = Baseline Slope $>1.0$ for at least one set during baseline procedures; * = Scores improved due to PF and differentiation between conditions was not observed.
Statistical Analysis

This section provides a detailed explanation of the statistical analyses that were used for this study. Data were analyzed in addition to the visual analysis so that comparisons between groups and interventions could be clarified. The Tau-U statistic was used to measure the effect size of the interventions for each individual and for various intervention groups and independent t-tests were used to assess whether differences between various groups could be considered statistically significant.

The Tau-U statistic and its application to this study will be described, followed by general explanation of the findings from the independent t-tests. There were additional statistics used to answer individual research questions. For example, the findings associated with Pearson product-moment correlation coefficients that were calculated to assess whether improved fluency scores correlated with improved scores on generalization measures will be presented. These findings will be embedded in the relevant research questions, as they are only pertinent to specific areas of inquiry.

**Tau-U.** Tau-U is a method for analyzing effect sizes so that improvement, or lack thereof, within the context of visual analysis can be quantified. This is important because reporting effect size can provide added credibility, reliability, and defensibility of conclusions based upon visual analysis within single-case research (Vannest & Ninci, 2015).

Tau-U is a non-parametric, non-overlap effect size that quantifies the difference in trend and level between two intervention phases (Parker, Vannest, Davis, & Sauber, 2011). Tau-U compares all data points in each phase and calculates the percent of data showing improvement from one intervention phase to the next (Parke et al.). Because
Tau-U includes all data points, effect sizes are less influenced by outliers and are considered reliable even with a small number of data points (Parker et al.). Brossart, Vannest, Davis, and Patience (2014) noted that the Tau-U has a number of special features: it has the ability to control for monotonic trend, it can be used when autocorrelation is present, it uses all the data involved, and it is conservative but not overly so. As a result, Brossert et al. stated that it is often the best non-parametric method for analyzing single-case data. Further, because of its design, Tau-U can work with any type of experimental design and any type of data. Trends do not have to be linear for the Tau-U to be applied (Brossert et al.). Finally, the Tau-U has superior statistical power and it is considered the most discriminating compared to all other nonoverlap indexes (Parker et al., 2011).

When interpreting the Tau-U, a .20 improvement is considered a small change, .20 to .60 reflects a moderate change, .60 to .80 reflects a large change, and above .80 indicates a large to very large change (Vannest & Ninci, 2015). However, it was noted that these ranges should be carefully considered in the context of expected growth for any given research study. The established ambitious growth rate for third grade students is .5 DCPM increase per week for performance on mathematics CBM probes (Fuchs, Fuchs, Hamlett, & Waltz; 1999), indicating that expected rate of improvement for these types of measures is rather small. Thus, it was determined that the proposed ranges would be acceptable for the purposes of this study.

In order to calculate phase contrasts for Tau-U in this study several steps were taken. First, baseline slopes were calculated. As research suggests, correction for baseline trends were utilized whenever a positive baseline slope was greater than 0.40
As a result, 13 students required this correction for one or more sets. Data were entered into the web-based calculator (www.singlecasedesignresearch.org) and phase contrasts for Tau-U were derived for individual students, for group A, for group B, for group A and B together, and for students performing in each instructional range within groups A and B. Contrasts were calculated for the performance feedback intervention by calculating weighted averages using all three problem sets. Phase contrasts for the peer tutoring and combined interventions were determined for each student by calculating weighted averages using the two problem sets that had been assigned to the tutee and tutor conditions. Additionally, phase contrasts were calculated individually for the tutee, tutor, and control sets so that these conditions could be compared. Comparisons were made between the baseline phase and the single intervention phase (either the performance feedback or peer tutoring intervention) and between the single intervention phase and the combined phase. Together, these data allow us to quantify the findings discussed within the framework of visual analysis.

The online Tau-U calculator (www.singlecasedesignresearch.org) produces $p$-values and confidence intervals (CIs) so that the reliability of each Tau-U value can be assessed. Vannest et al. (2015) noted that it is important to include these values in the context of research in conjunction with graphed data and visual analysis in order to gain a full understanding of the participants’ trajectory. As such, this information is clearly reported throughout this chapter. The significance of Tau-U values allows us to ascertain the probability that the same effect would be likely to occur again, with low $p$-values indicating high levels of probability. Significant effect sizes, when combined with clear
visual data, are the most reliable way of understanding whether or not an intervention was effective.

However, nonsignificant Tau-U values indicative of moderate and large effect sizes were calculated for several participants in this study. These values indicated that growth occurred, but the data were not consistent or robust enough to ensure that this rate of improvement would be observed again. When these values were obtained, the 90% or 95% CI indicated that the lower limit fell below 0 and it could not be predicted that growth would occur again.

Vannest et al. (2015) reported that CI’s and p-values are not necessary for the purposes of making low- or medium-stakes decisions and note that combining effect size estimates with graphed data provides a numerical estimate of change that is depicted in the visual representation of the data. For the purposes of this study, nonsignificant Tau-U values should be interpreted as a way of quantifying individual performance, but it cannot be assumed that results based on nonsignificant values could be generalized to other students.

Tau-U values and the related effect sizes and 90% or 95% confidence intervals will be presented and discussed later in this chapter within the context of the research questions proposed at the beginning of this chapter.

**T-test.** When testing for significant difference between two means with a relatively small sample size ($N < 30$), it is recommended that a $t$-test be used (Gall, Gall, & Borg, 2003). Independent $t$-tests were then used to compare interventions across students using z-scores that were derived for each student by the Tau-U calculator. The results of these tests can be observed in Table 12. All but three of these scores indicated
that there was no significant difference between interventions or groups. The one effect for which there was a significant impact was a student’s instructional level, with students in the instructional group showing greater gains than students in the frustrational group when participating in the performance feedback and combined interventions. This statistical finding reflects the differences that have been observed throughout the visual data analysis thus far with regard to instructional level and intervention effects. These findings will be discussed in detail in upcoming sections of this chapter.

Table 12

*Means, Standard Deviations, and Independent t-test Results for Z-Scores across Groups*

<table>
<thead>
<tr>
<th>Comparison Groups</th>
<th>Group 1</th>
<th>Group 2</th>
<th>t-value</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF</td>
<td>.158 (2.20)</td>
<td>.124 (1.06)</td>
<td>.37</td>
<td>9</td>
</tr>
<tr>
<td>PT</td>
<td>.45 (1.47)</td>
<td>1.10 (1.38)</td>
<td>1.07</td>
<td>20</td>
</tr>
<tr>
<td>PT + PF</td>
<td>.89 (1.82)</td>
<td>1.29 (1.24)</td>
<td>.78</td>
<td>30</td>
</tr>
<tr>
<td>PF AI</td>
<td>2.97 (1.57)</td>
<td>-.26 (1.38)</td>
<td>2.89*</td>
<td>5</td>
</tr>
<tr>
<td>Combined AI</td>
<td>1.94 (1.25)</td>
<td>1.13 (0.69)</td>
<td>1.07</td>
<td>5</td>
</tr>
<tr>
<td>PT BI</td>
<td>1.13 (0.94)</td>
<td>-.37 (1.67)</td>
<td>1.79</td>
<td>6</td>
</tr>
<tr>
<td>Combined BI</td>
<td>1.89 (1.22)</td>
<td>.14 (0.88)</td>
<td>2.75*</td>
<td>9</td>
</tr>
<tr>
<td>Combined (AI+BI)</td>
<td>1.91 (1.16)</td>
<td>.51 (0.92)</td>
<td>2.84**</td>
<td>16</td>
</tr>
<tr>
<td>PT Tutor</td>
<td>.27 (1.31)</td>
<td>.56 (0.92)</td>
<td>.59</td>
<td>18</td>
</tr>
<tr>
<td>PT Tutor</td>
<td>.16 (1.61)</td>
<td>.27 (1.31)</td>
<td>.18</td>
<td>19</td>
</tr>
<tr>
<td>PT Control</td>
<td>.16 (1.61)</td>
<td>.56 (0.92)</td>
<td>.71</td>
<td>16</td>
</tr>
<tr>
<td>Combined Tutor</td>
<td>1.29 (1.41)</td>
<td>.54 (1.61)</td>
<td>1.35</td>
<td>28</td>
</tr>
<tr>
<td>Combined Tutor</td>
<td>.91 (1.28)</td>
<td>1.29 (1.41)</td>
<td>.81</td>
<td>28</td>
</tr>
<tr>
<td>Combined Control</td>
<td>.91 (1.28)</td>
<td>.54 (1.61)</td>
<td>.89</td>
<td>23</td>
</tr>
<tr>
<td>PT + C Tutor</td>
<td>.85 (1.43)</td>
<td>.55 (1.34)</td>
<td>.89</td>
<td>50</td>
</tr>
<tr>
<td>PT + C Tutor</td>
<td>.65 (1.42)</td>
<td>.85 (1.43)</td>
<td>.42</td>
<td>53</td>
</tr>
<tr>
<td>PT + C Control</td>
<td>.65 (1.42)</td>
<td>.65 (1.42)</td>
<td>.54</td>
<td>53</td>
</tr>
</tbody>
</table>

Note. *p < .05, two-tailed. **p < .05, two-tailed; PF = Performance Feedback Intervention; PT = Performance Feedback Intervention; C = Combined Intervention; AI/BI = Groups A or B at the Instructional Level; AF/BF = Groups A or B at the Frustrational Level.
Findings

This section summarizes the results in the context of the research questions that were posed at the start of the study. First, socioeconomic status (SES) as a contributing factor to student success will be discussed. Next, the findings associated with Pearson product-moment correlation coefficients that were calculated to assess the relationship between improved fluency scores throughout this study and improved scores on the AIMSweb M-CAP and M-COMP pre- and post-test generalization measures will be presented.

Finally, the findings to the research questions that address student performance on the multiplication probes of this study will be presented. That is, comparisons between interventions across students will be relayed for four different areas using visual analysis, Tau-U analysis, and independent t-tests. The first question sought to determine the effectiveness of the performance feedback and peer tutoring interventions, while the second question sought to determine if a combined intervention that was comprised of both interventions was more effective than either of the interventions on their own. Third, there will be a discussion comparing the tutee and tutor conditions as part of the peer tutoring and combined interventions. Finally, the impact of student instructional level will be presented. When answering the question about students’ instructional level, the effectiveness of the interventions and intervention components will be discussed as well as their impact on students’ ability to achieve mastery levels of performance. These findings will be interpreted and discussed in Chapter 5.

SES as a mitigating factor. The first research question of this study sought to determine if interventions were equally beneficial for students who received free lunch
and those who did not. It was hypothesized that SES would not be a contributing factor to the relative effectiveness of the intervention.

As noted in the literature review, there is a difference in mathematics performance between those students in poverty and those who are not (NAEP; 2012; OECD, 2013b; Provasnik et al., 2012). We define students in poverty as those who qualify for the federally-funded, free- and reduced-lunch program; this program provides free and reduced lunches for students based on income level and family size (Department of Agriculture, 2013).

In order to answer this question, a school was chosen in which 53% of students qualify for free or reduced-price lunch (Vermont Agency of Education [VAE], 2015). Despite school-wide SES representation, when the sample was established, only 5 of 18 students, or 28% of the sample, fell into this category.

When considering the students who qualified for free and reduced lunch, four of the five students began the study in the frustrational range. When considering individual performance, 3 of the 5 students (2 instructional, 1 frustrational) showed significant improvement throughout the study and 2 of the 5 (1 instructional, 1 frustrational) reached mastery levels of performance. However, given such a small number of students in poverty that participated in this study, conclusions regarding the impact of this factor could not be assessed within the context of this study.

**Generalization.** The second research question of this study sought to determine if increased fluency scores for students could be generalized to improvement on more complex mathematical tasks.
As previously noted, generalization was assessed by using alternate versions of both the AIMSweb M-COMP and M-CAP probes as pre- and post- tests. The third grade M-COMP probes assess skills regarding column addition, basic facts, and complex computation (Pearson, 2012d), while the third grade M-CAP probes assessed understanding of facts, fractions, and geometry (Pearson, 2012b). The pre- and post-tests were scored and differences between the administrations were calculated. Next, initial baseline mean scores were compared to average scores from the final week of the study for each student. The difference between these scores was then calculated.

Finally, a Pearson product-moment correlation coefficient was computed to assess the relationship between the change scores for both the M-COMP and M-CAP measures to change in fluency scores over the course of the intervention. There was a significant positive relationship between increased fluency rates throughout the study and increased scores on the M-COMP; $r (16) = .69, p = .002$. Conversely, there was not a significant relationship observed between increased fluency and M-CAP scores; $r (16) = .31, p = .22$.

However, only students in the instructional range reached mastery levels of performance. Because the instructional hierarchy suggests that generalization does not begin to occur until mastery levels are reached, Pearson correlations were also calculated comparing the change scores for students who began the study in the instructional range. Interestingly, there was a nonsignificant relationship between increased fluency rates and increased scores on the M-COMP; $r (9) = .62, p = .05$, but a positive and significant relationship was observed between increased fluency scores and increased scores on the M-CAP; $r (9) = .79, p = .006$. 
The hypothesis for the third research question proposed that increased fluency scores would correlate with increased scores on both generalization measures. Thus, the hypothesis was partially supported. The findings for the group are not surprising given that M-COMP assesses skills that are more aligned with calculation tasks, whereas the M-CAP assesses math reasoning abilities. However it is surprising that the students performing in the instructional range yielded a positive and significant relationship between increased fluency scores and the M-CAP, but not for the M-COMP. These findings will be explored further in the next chapter.

**The performance feedback intervention and the peer tutoring intervention.**

The third research question sought to determine if a performance feedback intervention, in which students simply practiced math worksheets and attempted to beat their previous high scores and a peer tutoring intervention, in which students briefly tutored one another before completing the same math work sheets were effective for increasing students’ fluency scores. It was hypothesized that these interventions would lead to increased fluency scores for students.

*Performance feedback and peer tutoring: Visual analysis.* Visual analysis indicated that the performance feedback intervention was not verified as effective as intervention effects could only be attributed to improved scores for 1 out of 7 students. When considering the peer tutoring intervention, the tutor condition was only verified as effective for 2 students and the tutee condition was verified as effective for 4 of 11. Consequently, the peer tutoring intervention was only partially verified as effective according to established guidelines for interpreting single-case research, which require verification to be replicated a minimum of three times (Horner et al., 2005). Although,
there was minimal support for the use of the peer tutoring intervention to increase student’s fluency scores, it could be inferred that neither intervention was very effective for the purpose of improving students’ fluency scores.

Table 13

*Tau-U Effect Sizes for Intervention*

<table>
<thead>
<tr>
<th>Group</th>
<th>Condition</th>
<th>95% Confidence Interval</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>LL</td>
<td>Tau-U</td>
</tr>
<tr>
<td>A</td>
<td>Performance Feedback</td>
<td>.14</td>
<td>.28</td>
</tr>
<tr>
<td>B</td>
<td>Peer Tutoring (Tutor)</td>
<td>-.07</td>
<td>.15</td>
</tr>
<tr>
<td></td>
<td>Peer Tutoring (Tutee)</td>
<td>-.07</td>
<td>.15</td>
</tr>
<tr>
<td>B</td>
<td>Peer Tutoring (Tutee+Tutor)</td>
<td>-.01</td>
<td>.15</td>
</tr>
<tr>
<td>B</td>
<td>Peer Tutoring (Control)</td>
<td>-.09</td>
<td>.09</td>
</tr>
<tr>
<td>A</td>
<td>Combined (Tutor)</td>
<td>.15</td>
<td>.39</td>
</tr>
<tr>
<td>A</td>
<td>Combined (Tutee)</td>
<td>.23</td>
<td>.49</td>
</tr>
<tr>
<td>A</td>
<td>Combined (Tutee+Tutor)</td>
<td>.25</td>
<td>.45</td>
</tr>
<tr>
<td>A</td>
<td>Combined (Control)</td>
<td>.12</td>
<td>.32</td>
</tr>
<tr>
<td>B</td>
<td>Combined (Tutor)</td>
<td>.15</td>
<td>.38</td>
</tr>
<tr>
<td>B</td>
<td>Combined (Tutee)</td>
<td>.17</td>
<td>.40</td>
</tr>
<tr>
<td>B</td>
<td>Combined (Tutee+Tutor)</td>
<td>.23</td>
<td>.39</td>
</tr>
<tr>
<td>B</td>
<td>Combined (Control)</td>
<td>.07</td>
<td>.23</td>
</tr>
<tr>
<td>A + B</td>
<td>Combined (Tutor)</td>
<td>.20</td>
<td>.39</td>
</tr>
<tr>
<td>A + B</td>
<td>Combined (Tutee)</td>
<td>.26</td>
<td>.44</td>
</tr>
<tr>
<td>A + B</td>
<td>Combined (Tutee + Tutor)</td>
<td>.28</td>
<td>.41</td>
</tr>
<tr>
<td>A + B</td>
<td>Combined (Control)</td>
<td>.13</td>
<td>.25</td>
</tr>
</tbody>
</table>

*Note.* LL = Lower Limit, UL = Upper Limit.
### Table 14

**Tau-U Effect Sizes for the Performance Feedback Intervention**

<table>
<thead>
<tr>
<th>Student/Group</th>
<th>95% Confidence Interval</th>
<th>$p$ – value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Upper Limit</td>
<td>Tau-U</td>
</tr>
<tr>
<td>S1 (AI-1)</td>
<td>.05</td>
<td>.51</td>
</tr>
<tr>
<td>S2 (AI-2)</td>
<td>.23</td>
<td>.50</td>
</tr>
<tr>
<td>S3 (AI-3)</td>
<td>.54</td>
<td>.91</td>
</tr>
<tr>
<td>S4 (AI-4)</td>
<td>-.13</td>
<td>.24</td>
</tr>
<tr>
<td>S5 (AF-1)</td>
<td>-.15</td>
<td>.23</td>
</tr>
<tr>
<td>S6 (AF-2)</td>
<td>-.35</td>
<td>-.06</td>
</tr>
<tr>
<td>S7 (AF-3)</td>
<td>-.60</td>
<td>-.27</td>
</tr>
</tbody>
</table>

### Table 15

**Tau-U Effect Sizes for the Peer Tutoring Intervention**

<table>
<thead>
<tr>
<th>Student/Group</th>
<th>95% Confidence Interval</th>
<th>$p$ – value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Upper Limit</td>
<td>Tau-U</td>
</tr>
<tr>
<td>S8 (BI-1)</td>
<td>-.41</td>
<td>.12</td>
</tr>
<tr>
<td>PT</td>
<td></td>
<td>.65</td>
</tr>
<tr>
<td>Control</td>
<td>-.64</td>
<td>-.10</td>
</tr>
<tr>
<td>S9 (BI-2)</td>
<td>-.19</td>
<td>.34</td>
</tr>
<tr>
<td>PT</td>
<td></td>
<td>.87</td>
</tr>
<tr>
<td>Control</td>
<td>-.62</td>
<td>-.08</td>
</tr>
<tr>
<td>S10 (BI-3)</td>
<td>-.35</td>
<td>.27</td>
</tr>
<tr>
<td>PT</td>
<td></td>
<td>.89</td>
</tr>
<tr>
<td>Control</td>
<td>-.82</td>
<td>-.17</td>
</tr>
<tr>
<td>S11 (BI-4)</td>
<td>-.42</td>
<td>.13</td>
</tr>
<tr>
<td>PT</td>
<td></td>
<td>.68</td>
</tr>
<tr>
<td>Control</td>
<td>-.30</td>
<td>.24</td>
</tr>
<tr>
<td>S12 (BI-5)</td>
<td>.06</td>
<td>.47</td>
</tr>
<tr>
<td>PT</td>
<td></td>
<td>.89</td>
</tr>
<tr>
<td>Control</td>
<td>.22</td>
<td>.61</td>
</tr>
<tr>
<td>S13 (BI-6)</td>
<td>-.04</td>
<td>.37</td>
</tr>
<tr>
<td>PT</td>
<td></td>
<td>.37</td>
</tr>
<tr>
<td>Control</td>
<td>.14</td>
<td>.53</td>
</tr>
</tbody>
</table>

*Note: LL = Lower Limit, UL = Upper Limit; PT = Peer tutoring intervention.*
Performance feedback and peer tutoring: Tau-U analysis. In contrast, Tau-U analysis for the performance feedback and peer tutoring interventions indicated that students receiving performance feedback intervention (group A) earned moderately greater scores than did those receiving the peer tutoring intervention (group B). When comparing the baseline condition to the performance feedback intervention (Table 13), a significant Tau-U \((p = .001)\) indicated 28% improvement from baseline to intervention. According to established guidelines, this level of improvement is considered to be moderate growth (Vannest & Ninci, 2015). On the contrary, when comparing the baseline intervention with the peer tutoring intervention, a non-significant Tau-U \((p = .14)\) was observed, indicating only a 15% improvement (Table 13). Additionally, a non-significant Tau-U value \((p = .98)\) indicating only a 9% increase when comparing the student performance within the control condition from baseline to intervention phase was obtained. Thus, a pattern emerged in which a significant and moderate effect size was observed for the performance feedback intervention and nonsignificant effect sizes indicative of minimal growth were observed for the peer tutoring and control conditions. Nonetheless, improvements were minimal and not indicative of significant growth. Therefore, one might infer that performance feedback had a moderate impact on students’ fluency rates; while the peer tutoring did not impact students’ fluency rates. Yet, Tau-U analysis can quantify effect size, it cannot confirm causality. It is always necessary to consider Tau-U analysis within the context of individual visual analysis.

Tau-U values were calculated to obtain an effect size for each individual student in groups A and B using the Tau-U online calculator (www.singlecaseresearch.org). For group A, the performance feedback group, these values were obtained by combining the
effect sizes for problem sets A, B, and C, while for group B, Tau-U values were obtained by combining the effect sizes for the tutee and tutor conditions. Individual Tau-U values and their accompanying confidence intervals are presented in Tables 14 and 15. Tau-U analysis indicated significant effects for 3 of 7 students in group A and 1 of 11 students in group B; however, visual analysis indicated that intervention effects were only present for 1 of these 4 students. It is also noteworthy that moderate effect sizes were indicated by nonsignificant Tau-U values for 3 additional students within the peer tutoring condition, whose improvement was attributed to intervention effects.

Based solely on Tau-U analysis, one might conclude that the performance feedback intervention was effective, while the peer tutoring intervention was not. However, visual analysis concluded that most of the improvements observed within the performance feedback condition were not result of this intervention. Further, Tau-U analysis was not indicative of significant improvement for the peer tutoring intervention, meaning that intervention effects that were observed within the context of visual analysis were not statistically significant. Due to the conflicting evidence and weak conclusions, it is likely that neither intervention – the performance feedback in isolation or the peer tutoring - could be verified as effective.

Combined intervention. The fourth research question of this study sought to determine if a combined intervention, which included both the peer tutoring and performance feedback interventions, was better than either intervention in isolation. It was hypothesized that all students will show greater increases in fluency scores when the two interventions are combined compared to their performance as a result of either intervention in isolation.
Table 16

**Tau-U Effect Sizes within the Combined Intervention: Group A**

<table>
<thead>
<tr>
<th>Student/Group</th>
<th>Condition</th>
<th>90% or 95% CI</th>
<th>p-value</th>
<th>Student/Group</th>
<th>Condition</th>
<th>90% or 95% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>Combined</td>
<td>.17</td>
<td>.60</td>
<td>.007</td>
<td>S5</td>
<td>Combined</td>
<td>-.26</td>
</tr>
<tr>
<td>(AI-1) Control</td>
<td>-</td>
<td>-.11</td>
<td>.31</td>
<td>.74</td>
<td>(AF-1) Control</td>
<td>-.16</td>
<td>.56</td>
</tr>
<tr>
<td>S2</td>
<td>Tuttee</td>
<td>-.23</td>
<td>.22</td>
<td>.67</td>
<td>S6</td>
<td>Tuttee</td>
<td>-.01</td>
</tr>
<tr>
<td>(AI-2) Control</td>
<td>-</td>
<td>.02</td>
<td>.56</td>
<td>1.10</td>
<td>(AF-2) Control</td>
<td>-.51</td>
<td>.09</td>
</tr>
<tr>
<td>S3</td>
<td>Combined</td>
<td>.12</td>
<td>.53</td>
<td>.94</td>
<td>S7</td>
<td>Tutor</td>
<td>-.37</td>
</tr>
<tr>
<td>(BI-3) Control</td>
<td>-</td>
<td>.27</td>
<td>.67</td>
<td>1.06</td>
<td>(AF-3) Control</td>
<td>-.63</td>
<td>.00</td>
</tr>
<tr>
<td>S4</td>
<td>Combined</td>
<td>.10</td>
<td>.52</td>
<td>.92</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(AI-4) Control</td>
<td>-</td>
<td>-.19</td>
<td>.58</td>
<td>.63</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**Note.** 90% CI’s reported for the control, tutor, and tutee conditions, and 95% CI’s reported for combined intervention; LL = Lower Limit; UL = Upper Limit; Combined = Tau-U value for tutee + tutor condition; Tutee = student only received tutee condition of the Combined intervention; Tutor = student only received tutor condition of the Combined intervention; Boldface print = Tau-U values are significant at the p < .05 level.

**The combined intervention: Visual analysis.** Visual analysis clearly indicated that more students made positive gains within the combined intervention as compared to either individual intervention. Both conditions (tutee and tutor) of the combined intervention were verified as effective, with an increase in DCPM scores for 4 students. DCPM scores also increased as a result of the tutee condition for 4 additional students, while the tutor condition yielded increased scores for 1 additional student. Finally, the performance feedback component of the combined study was verified as effective for 1 student. Indeed, the combined intervention was effective based on established guidelines for single-case research (Horner et al., 2005), as verification of effectiveness occurred eight times for the tutee condition and five for the tutor condition.
Table 17
*Tau-U Effect Sizes for the Combined Intervention Group B*

<table>
<thead>
<tr>
<th>Student/Group</th>
<th>Condition</th>
<th>90% or 95% Confidence Interval</th>
<th>p-value</th>
<th>Student/Group</th>
<th>Condition</th>
<th>90% or 95% Confidence Interval</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LL</td>
<td>Tau-U</td>
<td>UL</td>
<td></td>
<td>LL</td>
<td>Tau-U</td>
<td>UL</td>
</tr>
<tr>
<td>S8 (BI-1)</td>
<td>Combined</td>
<td>.28</td>
<td><strong>.77</strong></td>
<td>1.27</td>
<td>.002</td>
<td>S14 (BI-1)</td>
<td>Combined</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>-.15</td>
<td>.25</td>
<td>.65</td>
<td>.31</td>
<td>Control</td>
<td>-.43</td>
</tr>
<tr>
<td>S9 (BI-2)</td>
<td>Combined</td>
<td>-.20</td>
<td><strong>.29</strong></td>
<td>.79</td>
<td>.04</td>
<td>S15 (BI-2)</td>
<td>Combined</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>-.15</td>
<td>.25</td>
<td>.65</td>
<td>.31</td>
<td>Control</td>
<td>-.72</td>
</tr>
<tr>
<td>S10 (BI-3)</td>
<td>Combined</td>
<td>.15</td>
<td><strong>.65</strong></td>
<td>1.14</td>
<td>.01</td>
<td>S16 (BI-3)</td>
<td>Combined</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>.10</td>
<td><strong>.49</strong></td>
<td>.89</td>
<td>.04</td>
<td>Control</td>
<td>-.35</td>
</tr>
<tr>
<td>S11 (BI-4)</td>
<td>Combined</td>
<td>-.01</td>
<td>.50</td>
<td>1.01</td>
<td>.06</td>
<td>S17 (BI-4)</td>
<td>Combined</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>-.19</td>
<td>.22</td>
<td>.63</td>
<td>.37</td>
<td>Control</td>
<td>.35</td>
</tr>
<tr>
<td>S12 (BI-5)</td>
<td>Tutee</td>
<td>-.05</td>
<td>.50</td>
<td>.95</td>
<td>.13</td>
<td>S18 (BI-5)</td>
<td>Combined</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>-.77</td>
<td>-.27</td>
<td>.24</td>
<td>.38</td>
<td>Control</td>
<td>-.11</td>
</tr>
<tr>
<td>S13 (BI-6)</td>
<td>Tutor</td>
<td>.13</td>
<td><strong>.70</strong></td>
<td>1.27</td>
<td>.02</td>
<td>Control</td>
<td>-.43</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>-.43</td>
<td>.12</td>
<td>.36</td>
<td>.73</td>
<td>Control</td>
<td></td>
</tr>
</tbody>
</table>

Note. LL = Lower Limit; UL = Upper Limit; Combined = Tau-U value for tutee + tutor condition; Tutee = student only received tutee condition of the Combined intervention; Tutor = student only received tutor condition of the Combined intervention; 90% Cis reported for control Condition; Boldface print = Tau-U values are significant at the p < .05 level.

The combined intervention: *Tau-U analysis.* Correspondingly, analysis of effect size with the Tau-U indicated that the combined intervention was more effective than the individual interventions. The results associated with the Tau-U statistic indicated 41% (p = .00) improvement from the single interventions to the combined intervention (see Table 13). The contrasts between each individual intervention and the combined intervention...
yielded similar results. Equally significant, Tau-U values ($p = .00$) indicated improvement when comparing the performance feedback and peer tutoring interventions to the combined intervention. The Tau-U indicated 45% improvement from the performance feedback to the combined intervention, which was slightly higher than the 39% improvement that was observed from the peer tutoring to combined intervention.

According to established guidelines, moderate growth was observed across all significant comparisons (Vannest & Ninci, 2015).

It is also remarkable that DCPM scores in the control condition increased upon implementation of the combined intervention. Significant Tau-U values indicated 32% ($p = .02$) improvement from the performance feedback intervention to the combined intervention (group A), 23% ($p = .004$) improvement from the peer tutoring intervention to the combined intervention (group B) and 25% ($p = .000$) improvement from the single intervention to the combined intervention for groups A and B. This might provide evidence that the performance feedback component of the combined intervention was effective; however, a greater level of improvement was observed for group A, which received the performance feedback intervention as the single intervention. Thus, one must consider the possibility that practice effects had a positive impact on DCPM scores.

These findings were further supported when looking at individual effect size for participants (Tables 15 and 16). As previously noted, Tau-U was significant for 4 students when comparing DCPM scores from the baseline to the performance feedback and peer tutoring interventions. As can be seen in Tables 15 and 16, Tau-U analyses indicated that 4 students from group A and 7 students from group B yielded significant effect sizes ranging from moderate to very large when comparing DCPM scores from the
single interventions to those in the combined intervention. Additionally, nonsignificant Tau-U values indicating moderate improvement were observed for 4 additional students. Visual analysis indicated that much of this improvement was associated with intervention effects, but as before, several students displayed improvement that could not be attributed to intervention effects.

There was one noteworthy pattern demonstrated by student 18. Although a nonsignificant Tau-U value \( (p = .21) \) indicated that there was 33% growth observed from the peer tutoring phase to the combined phase of the intervention, visual analysis of the graph for student 18 in Figure 4 and Tau-U analysis (Table 17) indicated that student performance decreased when comparing the baseline phase to the peer tutoring intervention. Consequently, the growth observed from the peer tutoring intervention to the combined intervention was the result of scores returning to baseline levels, and not the result of additional improvement. This finding highlights the need to consider all of the data within the context of single-case research.

**The combined intervention: The t-test.** Thus far, visual analysis and Tau-U analysis indicated the combined intervention was superior to the performance feedback and peer tutoring interventions. However, in order to further understand the extent of this difference, an independent one-tailed \( t \)-test was used to assess whether or not the combined intervention was significantly better than the individual interventions. When looking at the effects of these interventions on students’ fluency scores, there was not a significant effect for type of intervention, \( t(30) = 1.09, p = .14 \), indicating the combined intervention was not significantly better than the individual interventions when they were analyzed together. This finding was replicated when comparing the combined
intervention to each individual intervention separately. There was not a significant effect when comparing the combined intervention to the performance feedback intervention \([t(9) = 0.37, p = .36]\), or when comparing the combined intervention to the peer tutoring intervention \([t(20) = 1.33, p = .10]\).

**Summary.** The fourth research question and hypothesis proposed that the combined intervention would be more effective for more students than either the performance feedback or peer tutoring intervention in isolation.

Visual analysis indicated that the combined intervention was clearly more effective as 10 students showed improvement as the result of one or more of the conditions within the combined intervention, as compared to the 6 students who showed improvement as the result of the single interventions. Tau-U analysis confirmed this finding, as 15% \((p = .06)\) improvement was observed for the peer tutoring intervention, 28% \((p = .00)\) improvement was observed for the performance feedback intervention, and 41% \((p = .000)\) improvement was observed for the combined intervention phase.

Although visual analysis and analysis of effect size with the Tau-U statistic indicated that the combined intervention was effective for more students than the individual interventions, results of the t-test indicated that the discrepancies were not statistically significant. Implications of this finding will be discussed in the final chapter.

**Comparison of the tutee, tutor, and the control conditions.** The fifth research question of this study sought to determine if students responded differently to the tutee and tutor conditions of the peer tutoring and combined interventions. It was proposed that scores attributed to the tutee condition would show greater improvement than those assigned to the tutor or control conditions.
Table 18

**Tau-U Effect Sizes for Tutee and Tutor Conditions within the Peer Tutoring Intervention: Group B**

<table>
<thead>
<tr>
<th>Student/ Group</th>
<th>Condition</th>
<th>Tau-U</th>
<th>90% CI</th>
<th>p-value</th>
<th>Student/ Group</th>
<th>Condition</th>
<th>Tau-U</th>
<th>90% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>LL</td>
<td>UL</td>
<td></td>
<td></td>
<td></td>
<td>LL</td>
<td>UL</td>
<td></td>
</tr>
<tr>
<td>S8 (BI-1)</td>
<td>Tutor</td>
<td>-.23</td>
<td>.40</td>
<td>1.03</td>
<td>.30</td>
<td>S14 (BF-1)</td>
<td>Tutor</td>
<td>-.38</td>
<td>.72</td>
</tr>
<tr>
<td></td>
<td>Tutee</td>
<td>-.79</td>
<td>-.16</td>
<td>.47</td>
<td>.68</td>
<td>Control</td>
<td>-1.45</td>
<td>-.92</td>
<td>.37</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>-.64</td>
<td>-.10</td>
<td>.44</td>
<td>.76</td>
<td>S15 (BF-2)</td>
<td>Tutor</td>
<td>-.87</td>
<td>-.20</td>
</tr>
<tr>
<td>S9 (BI-2)</td>
<td>Tutor</td>
<td>-.31</td>
<td>.32</td>
<td>.95</td>
<td>.81</td>
<td>Control</td>
<td>-1.38</td>
<td>-.80</td>
<td>.22</td>
</tr>
<tr>
<td></td>
<td>Tutee</td>
<td>-.27</td>
<td>.36</td>
<td>.99</td>
<td>.35</td>
<td>Control</td>
<td>-.57</td>
<td>.10</td>
<td>.77</td>
</tr>
<tr>
<td>S10 (BI-3)</td>
<td>Tutor</td>
<td>-.87</td>
<td>-.13</td>
<td>.60</td>
<td>.77</td>
<td>S16 (BF-3)</td>
<td>Tutor</td>
<td>-.09</td>
<td>.63</td>
</tr>
<tr>
<td></td>
<td>Tutee</td>
<td>-.14</td>
<td>.60</td>
<td>.18</td>
<td>.68</td>
<td>Control</td>
<td>-.46</td>
<td>.25</td>
<td>.96</td>
</tr>
<tr>
<td>S11 (BI-4)</td>
<td>Tutor</td>
<td>-.43</td>
<td>.20</td>
<td>.60</td>
<td>.60</td>
<td>Control</td>
<td>-.86</td>
<td>-.25</td>
<td>.36</td>
</tr>
<tr>
<td></td>
<td>Tutee</td>
<td>-.62</td>
<td>.05</td>
<td>.90</td>
<td>.90</td>
<td>Control</td>
<td>-.21</td>
<td>1.08</td>
<td>.02</td>
</tr>
<tr>
<td>S12 (BI-5)</td>
<td>Tutor</td>
<td>-.04</td>
<td>.50</td>
<td>1.04</td>
<td>.13</td>
<td>Control</td>
<td>-.62</td>
<td>-.08</td>
<td>.46</td>
</tr>
<tr>
<td></td>
<td>Tutee</td>
<td>.22</td>
<td>.61</td>
<td>1.02</td>
<td>.01</td>
<td>S18 (BF-5)</td>
<td>Tutor</td>
<td>-.104</td>
<td>-.50</td>
</tr>
<tr>
<td>S13 (BI-6)</td>
<td>Tutor</td>
<td>-.10</td>
<td>.44</td>
<td>.98</td>
<td>.18</td>
<td>Control</td>
<td>-.71</td>
<td>-.27</td>
<td>.17</td>
</tr>
<tr>
<td></td>
<td>Tutee</td>
<td>-.13</td>
<td>.31</td>
<td>.75</td>
<td>.24</td>
<td></td>
<td>-.</td>
<td>-.</td>
<td>-.</td>
</tr>
</tbody>
</table>

**Note.** LL/UL = Lower Limit/Upper Limit; B = Group B; BI = Group B, instructional range; BF = Group B, frustrational range; PT = Tutee +Tutor. Boldface print = Tau-U values are significant at the *p* < .05 level.

**Tutee, tutor, and control conditions: Visual analysis.** What follows is a comparison between students’ performance when serving as a tutee versus tutor within both intervention phases. Visual analysis indicated that both conditions were effective for multiple students, with more improvement observed within the context of the combined intervention as compared to the peer tutoring intervention. When considering student performance across both intervention phases, a total of 9 students improved as a result of the tutee condition in one or both of the intervention phases and the tutor condition was verified as effective for 6 students. While the tutee condition was found to
be beneficial for more students than the tutor condition, there were a sufficient number of students that responded well in the tutor condition to establish the condition as effective according to established guidelines (Horner et al., 2005). As differentiation was necessary for the tutee or the tutor condition to be considered effective, student scores for these conditions were higher than those observed within the control condition.

Table 19

*Tau-U Effect Sizes for Tutee and Tutor Conditions within the Combined Intervention: Group A*

<table>
<thead>
<tr>
<th>Student/Group</th>
<th>Condition</th>
<th>90% CI</th>
<th>p-value</th>
<th>Student/Group</th>
<th>Condition</th>
<th>90% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1 (AI-1)</td>
<td>Tutor</td>
<td>.04</td>
<td>.50</td>
<td>.96</td>
<td>.08</td>
<td>.27</td>
<td>.36</td>
</tr>
<tr>
<td></td>
<td>Tutee</td>
<td>.16</td>
<td>.73</td>
<td>1.29</td>
<td>.03</td>
<td>.42</td>
<td>.12</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>-.11</td>
<td>.31</td>
<td>.74</td>
<td>.22</td>
<td>-.16</td>
<td>.56</td>
</tr>
<tr>
<td>S2 (AI-2)</td>
<td>Tutor</td>
<td>-.03</td>
<td>.60</td>
<td>1.26</td>
<td>.23</td>
<td>-.11</td>
<td>.50</td>
</tr>
<tr>
<td></td>
<td>Tutee</td>
<td>.02</td>
<td>.56</td>
<td>1.10</td>
<td>.09</td>
<td>-.51</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>.26</td>
<td>.70</td>
<td>1.13</td>
<td>.01</td>
<td>-.47</td>
<td>.16</td>
</tr>
<tr>
<td>S3 (AI-3)</td>
<td>Tutor</td>
<td>-.22</td>
<td>.32</td>
<td>.86</td>
<td>.33</td>
<td>.63</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>Tutee</td>
<td>.27</td>
<td>.67</td>
<td>1.06</td>
<td>.01</td>
<td>-.63</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>-.17</td>
<td>.52</td>
<td>1.06</td>
<td>.11</td>
<td>-.63</td>
<td>.00</td>
</tr>
<tr>
<td>S4 (AI-4)</td>
<td>Tutor</td>
<td>.28</td>
<td>.82</td>
<td>1.36</td>
<td>.01</td>
<td>-.63</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>Tutee</td>
<td>-.19</td>
<td>.58</td>
<td>.63</td>
<td>.37</td>
<td>-.63</td>
<td>.00</td>
</tr>
</tbody>
</table>

*Note.* LL/UL = Lower Limit/Upper Limit; A = Group A; AI = Group A, instructional range; AF = Group A, frustrational range; PT = Tutee + Tutor. Boldface print = Tau-U values are significant at the p < .05 level.
Table 20

**Tau-U Effect Sizes for the Combined Intervention: Group B**

<table>
<thead>
<tr>
<th>Student/Group</th>
<th>Condition</th>
<th>90% CI</th>
<th>p-value</th>
<th>Student/Group</th>
<th>Condition</th>
<th>90% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>S8 (BI-1)</td>
<td>Tutor</td>
<td>-.03</td>
<td>.60</td>
<td>1.21</td>
<td>.12</td>
<td>S 14  (BF-1)</td>
<td>Tutee</td>
</tr>
<tr>
<td></td>
<td>Tutee</td>
<td>.38</td>
<td>.92</td>
<td>1.47</td>
<td>.005</td>
<td>Control</td>
<td>-.43</td>
</tr>
<tr>
<td></td>
<td>Tutee</td>
<td>-.79</td>
<td>-.16</td>
<td>.47</td>
<td>.67</td>
<td>Control</td>
<td>-.60</td>
</tr>
<tr>
<td>S10 (BI-3)</td>
<td>Tutor</td>
<td>.12</td>
<td>.66</td>
<td>1.19</td>
<td>.04</td>
<td>S 16  (BF-3)</td>
<td>Tutee</td>
</tr>
<tr>
<td></td>
<td>Tutee</td>
<td>.01</td>
<td>.64</td>
<td>1.27</td>
<td>.09</td>
<td>Control</td>
<td>-.27</td>
</tr>
<tr>
<td>S11 (BI-4)</td>
<td>Tutor</td>
<td>-.28</td>
<td>.26</td>
<td>.80</td>
<td>.43</td>
<td>S 17  (BF-4)</td>
<td>Control</td>
</tr>
<tr>
<td></td>
<td>Tutee</td>
<td>.13</td>
<td>.80</td>
<td>1.47</td>
<td>.05</td>
<td>Control</td>
<td>-.37</td>
</tr>
<tr>
<td>S12 (BI-5)</td>
<td>Tutor</td>
<td>-.19</td>
<td>.22</td>
<td>.63</td>
<td>.37</td>
<td>S 18  (BF-5)</td>
<td>Tutee</td>
</tr>
<tr>
<td></td>
<td>Tutee</td>
<td>-.04</td>
<td>.50</td>
<td>1.04</td>
<td>.12</td>
<td>Control</td>
<td>-.11</td>
</tr>
<tr>
<td>S13 (BI-6)</td>
<td>Tutor</td>
<td>-.77</td>
<td>-.27</td>
<td>.24</td>
<td>.38</td>
<td>Control</td>
<td>-.43</td>
</tr>
<tr>
<td></td>
<td>Tutee</td>
<td>.43</td>
<td>.70</td>
<td>1.37</td>
<td>.02</td>
<td>Control</td>
<td>-.43</td>
</tr>
</tbody>
</table>

Note. *p < .05, **p < .01, ***p < .001; S = Student; LL/UL = Lower Limit/Upper Limit; B = Group B; BI = Group B, instructional range; BF = Group B, frustrational range; PT = Tutee +Tutor. Boldface print = Tau-U values are significant at the p < .05 level.

**Tutee, tutor, and control conditions: Tau-U analysis.** Within the peer tutoring intervention, the superiority of the tutee condition as compared to the tutor condition was not evident based on analysis of effect size. Tau-U analysis indicated that 15% improvement was observed for both the tutor and tutee condition as compared to baseline scores (p = .19, p = .18 respectively). These effect sizes were insignificant and neither condition was better than the other. Further, a nonsignificant Tau-U value (p = .34)
indicated only 9% improvement from the baseline to the intervention phase within the control condition.

Within the combined intervention, the tutor, the tutee and the control conditions all yielded significant and moderate effect sizes for groups A and B. When comparing single intervention (performance feedback or peer tutoring) to the combined intervention, the tutee condition yielded the highest effect sizes as significant Tau-U values indicated 49% ($p = .000$) and 40% ($p = .001$) improvement in DCPM scores for groups A and B respectively. When considering groups A and B together, a significant Tau-U value ($p = .000$) indicated 44% improvement overall. These effect sizes are all indicative of moderate growth.

Tau-U analysis also indicated significant improvement for the tutor condition with similar rates of growth observed across groups. There was 39% ($p = .001$) growth from the performance feedback to the tutor condition of the combined intervention for group A, 38% ($p = .001$) growth from the tutor condition of the peer tutoring intervention to the same condition of the combined intervention for group B, and 39% ($p = .000$) improvement when considering both of these groups together. These numbers, while significant, were slightly lower than those generated for the tutee condition. In sum, a pattern emerged in which the tutee condition yielded higher effect sizes than tutor condition.

Finally, when considering the control condition significant Tau-U values indicated 32% ($p = .002$) improvement from the performance feedback intervention to the combined intervention, 23% ($p = .004$) improvement from the peer tutoring intervention to the combined intervention, and 25% ($p = .000$) improvement from the first intervention
phase to the second for groups A and B. Thus, the tutee condition yielded more improvement than the tutor condition, which yielded more improvement than the control condition.

As can be seen in Table 18, Tau-U analysis for individual students indicated that significant improvement was not observed for any students when measuring growth from the baseline to peer tutoring intervention with respect to the tutee and tutor conditions. However, nonsignificant Tau-U values indicated moderate and large effect sizes for several students in both conditions. Notably, two significant Tau-U values indicated large effect sizes and one significant Tau-U value indicated a moderate effect size for the control condition. Visual analysis revealed that for some students, growth could be attributed to intervention effects while for several others, it could not. Careful inspection of the graphs for students who received the peer tutoring intervention (Figure 4) indicated that growth was often inconsistent, based on relatively few data points, or began before the implementation of the peer tutoring intervention. Insignificant moderate effect sizes would be expected as a way of quantifying this type of growth.

As this research question sought to determine whether or not the tutee condition was more effective than either the control or tutor conditions, it was necessary to compare Tau-U values for each of the three conditions for each student. It was found that the control condition yielded the lowest effect size for 5 students, which was approximately half of the students who received the peer tutoring intervention. This indicates that DCPM scores for the tutee and tutor conditions improved more than DCPM scores for the control condition for these students. The tutee condition yielded the highest effect size for 3 students when compared to the tutor and control condition,
indicating that it was the most effective condition for these students. The tutor condition yielded the highest effect size for 4 students as compared to the tutee and control conditions, while the control condition also yielded the highest effect size for 4 students, indicating that these 4 students showed more growth for the control condition than for the sets targeted by the peer tutoring intervention. In sum, the data indicated that growth was evenly distributed across conditions.

When comparing DCPM scores from the single intervention phase to the combined intervention phase, Tau-U analysis indicated several significant and nonsignificant moderate, large, and very large effect sizes for the tutee and tutor conditions, and two significant effect sizes for the control condition (Tables 18 and 19). It is noteworthy that several students who displayed growth throughout the combined intervention also demonstrated improvement during the peer tutoring intervention. When comparing Tau-U values across conditions for each student, the smallest effect sizes were observed within the control condition as compared to the tutee or tutor conditions for 10 of 18 students. Interestingly, the largest effect size for each student was observed for tutee condition and the tutor condition eight times each. The control condition yielded the highest effect sizes for 2 students. Visual analysis indicated inconsistent results with regard to effect size and treatment effects for this condition, as much of the observed improvement was not attributed to intervention effects. When taken together improvement occurred for many students that was not necessarily accounted for by the tutee or tutor conditions. Further, while the control condition was not always the lowest score for students, it was also not usually the highest, as the tutee or tutor condition yielded the highest effect scores for 16 out of 18 students.
In sum, the data suggest that both the tutee and tutor conditions were useful components as several students made significant progress as the result of both conditions. However moderate growth, albeit to a lesser extent, was also observed for the control condition. Tau-U analysis confirmed the assertions made within the frame of visual analysis that considerable growth was observed for students as the result of the tutee and tutor conditions, particularly within the combined intervention phase. While Tau-U indicated a larger effect size for the tutee condition as compared to the tutor condition, further analysis was necessary to determine whether or not the two conditions were statistically different.

**The tutee, tutor, and control conditions: T-tests.** An independent two-tailed t-test was used to assess whether one condition, the tutee, the tutor, or the control, was better than the other (Table 12). When looking at the effects of these conditions on students’ fluency scores within the peer tutoring intervention, there was not a significant effect for condition, when comparing the tutee to the tutor condition \[ t(18) = 0.59, p = .56 \], the control to the tutee condition \[ t(19) = 0.18, p = .86 \], or the control to the tutor condition \[ t(16) = 0.71, p = .49 \], indicating that there was no significant difference in student performance for each of these conditions. This finding was replicated when comparing the tutor and tutee conditions within the combined intervention. There was not a significant effect when comparing the tutee condition to the tutor condition \[ t(28) = 1.35, p = .19 \], the control condition to the tutee condition \[ t(21) = 0.81, p = .42 \], or the control condition to the tutor condition \[ t(21) = 0.89, p = .38 \]. These findings were confirmed when comparing the z-scores for all conditions from both the peer tutoring and combined interventions as no significant effect was observed when comparing the tutee
and the tutor condition \( t(53) = 0.42, p = .43 \), the control condition to the tutee condition \( t(53) = .54, p = .59 \), or the control condition to the tutor condition \( t(53) = 0.42, p = .67 \).

**Summary.** In sum, both visual and Tau-U analyses indicated improvement for both the tutee and tutor conditions, with slightly more gains observed for the tutee condition as opposed to the tutor condition. However, results of the t-test indicated that the discrepancies were not statistically significant. Implications of this finding will be discussed in the final chapter.

**Instructional level as a mediating factor.** The sixth research question of this study sought to determine if instructional level at the start of the study impacted the effectiveness of the performance feedback, peer tutoring, or combined interventions.

**The impact of instructional level: Visual analysis.** Perhaps the most obvious and interesting results from this study related to students’ instructional level. Visual analysis indicated that improvement was mediated by instructional level, as students who began the study in the instructional range clearly outperformed those who began in the frustrational range when considering nearly every facet of intervention and performance.

First, visual analysis indicated that that intervention effects were attributed to 4 students in the instructional range and 2 students in the frustrational range during the single intervention phase. Furthermore, intervention effects of the combined intervention were attributed to 7 students in the instructional range and 3 in the frustrational range. Thus, the number of students from the instructional range who improved due to intervention effects was more than double those from the frustrational range. Moreover, 5 students in the instructional range, as compared to no students in the frustrational range,
showed significant improvement for at least one condition that could not be attributed to intervention effects. This means that growth occurred for these students, but there was no significant change in the data paths upon implementation of an intervention and differentiation was not present. Thus, experimental control was not demonstrated (Richards et al., 1999). When taken together, all students who began the study in the instructional range showed considerable improvement throughout the course of the study, while 3 students in the frustrational range showed little or no improvement.

**Instructional level: Tau-U analysis.** Tau-U analysis also indicated that students performing at the instructional level made greater gains than those performing at the frustrational level. Although we have found that Tau-U values were not always indicative of intervention effects, student growth has been accurately quantified throughout the course of the study. As can be seen in Tables 13 through 19, individual Tau-U values have consistently demonstrated a pattern in which larger and significant effect sizes were more frequently obtained for students performing in the instructional range as compared to those performing in the frustrational range. When considering Tau-U analysis at the group level, effect sizes for students performing in the instructional range were approximately double those of students performing in the frustrational range (Table 21).
### Table 21

*Tau-U Effect Sizes for Instructional Groups*

<table>
<thead>
<tr>
<th>Group Condition</th>
<th>LL</th>
<th>Tau-U</th>
<th>UL</th>
<th>p-value</th>
<th>95% CI</th>
<th>LL</th>
<th>Tau-U</th>
<th>UL</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI PF</td>
<td>.38</td>
<td>.54</td>
<td>.70</td>
<td>.00</td>
<td>AF PF</td>
<td>-.21</td>
<td>-.04</td>
<td>.12</td>
<td>.65</td>
</tr>
<tr>
<td>BI Tutor PT</td>
<td>-.04</td>
<td>.26</td>
<td>.55</td>
<td>.07</td>
<td>BF Tutor PT</td>
<td>-.35</td>
<td>-.02</td>
<td>.32</td>
<td>.92</td>
</tr>
<tr>
<td>BI Tutee PT</td>
<td>-.02</td>
<td>.28</td>
<td>.58</td>
<td>.06</td>
<td>BF Tutee PT</td>
<td>-.37</td>
<td>-.04</td>
<td>.29</td>
<td>.81</td>
</tr>
<tr>
<td>BI PT</td>
<td>-.03</td>
<td>.28</td>
<td>.50</td>
<td>.01</td>
<td>BF PT</td>
<td>-.26</td>
<td>-.03</td>
<td>.21</td>
<td>.81</td>
</tr>
<tr>
<td>AI Tutor C</td>
<td>.26</td>
<td>.54</td>
<td>.81</td>
<td>.000</td>
<td>AF Tutor C</td>
<td>-.27</td>
<td>.26</td>
<td>.79</td>
<td>.34</td>
</tr>
<tr>
<td>AI Tutee C</td>
<td>.28</td>
<td>.62</td>
<td>.95</td>
<td>.000</td>
<td>AF Tutee C</td>
<td>-.17</td>
<td>.31</td>
<td>.79</td>
<td>.21</td>
</tr>
<tr>
<td>AI C</td>
<td>.35</td>
<td>.60</td>
<td>.85</td>
<td>.000</td>
<td>AF Combined</td>
<td>.04</td>
<td>.36</td>
<td>.68</td>
<td>.36</td>
</tr>
<tr>
<td>BI Tutor C</td>
<td>.26</td>
<td>.54</td>
<td>.83</td>
<td>.000</td>
<td>BF Tutor C</td>
<td>-.16</td>
<td>.18</td>
<td>.51</td>
<td>.30</td>
</tr>
<tr>
<td>BI Tutee C</td>
<td>.23</td>
<td>.55</td>
<td>.87</td>
<td>.001</td>
<td>BF Tutee C</td>
<td>.06</td>
<td>.30</td>
<td>.60</td>
<td>.06</td>
</tr>
<tr>
<td>BI C</td>
<td>.33</td>
<td>.55</td>
<td>.76</td>
<td>.000</td>
<td>BF C</td>
<td>.01</td>
<td>.24</td>
<td>.47</td>
<td>.04</td>
</tr>
<tr>
<td>AI + BI Tutor C</td>
<td>.33</td>
<td>.54</td>
<td>.76</td>
<td>.000</td>
<td>AF + BF Tutor C</td>
<td>-.08</td>
<td>.20</td>
<td>.48</td>
<td>.17</td>
</tr>
<tr>
<td>AI + BI Tutee C</td>
<td>.37</td>
<td>.59</td>
<td>.81</td>
<td>.000</td>
<td>AF + BF Tutee C</td>
<td>.02</td>
<td>.28</td>
<td>.55</td>
<td>.04</td>
</tr>
<tr>
<td>AI + BI C</td>
<td>.41</td>
<td>.57</td>
<td>.72</td>
<td>.000</td>
<td>AF + BF C</td>
<td>.06</td>
<td>.25</td>
<td>.44</td>
<td>.009</td>
</tr>
</tbody>
</table>

**Note.** CI = Confidence Interval; LL = Lower Level; UL = Upper Level; A = Group A; B = Group B; I = Instructional Level; F = Frustrational Level; PT = Peer Tutoring Intervention; PF = Performance Feedback Intervention; C = Combined Intervention; Boldface Print = p < .05.

When considering the single intervention phase, Tau-U analysis showed that students performing in the frustrational range at the start of the study experienced no growth. In fact, Tau-U analysis indicated that scores went down for students in the frustrational group when measuring growth from the baseline to performance feedback intervention and from the baseline to the peer tutoring intervention. In fact, negative and
nonsignificant Tau-U values indicated that scores decreased upon implementation of the single interventions.

In contrast, the significant Tau-U value \( (p = .00) \) of 54\% for students in group A (AI) performing at the instructional level indicated moderate growth upon implementation of the performance feedback intervention. Although visual analysis established that the observed growth could not be attributed to intervention effects, the level of improvement for these students is noteworthy. When considering the peer tutoring intervention, a similar pattern was observed, although not as extreme. Nonsignificant Tau-U values indicated moderate growth for the instructional group B (BI) for both the tutee and tutor conditions, and when taken together a significant Tau-U value \( (p = .01) \) indicated 28\% improvement from the baseline phase to the peer tutoring phase.

Tau-U analysis indicated that students in the instructional range continued to consistently outperform those in the frustrational range when the combined intervention was introduced. Significant effect sizes were obtained for all conditions of the combined intervention across instructional groups AI and BI. Notably, a large effect size was observed when comparing the performance feedback intervention to the tutee condition of the combined intervention as evidenced by a significant Tau-U value \( (p = .000) \) which indicated 62\% growth for group AI. Improvement was such that a large effect size was also observed for the combined condition in its entirety with a significant Tau-U value \( (p = .000) \) that indicated 60\% improvement from the performance feedback intervention to the combined intervention. Moderate and significant effect sizes ranging from 54\% to
59% were noted for the instructional group for all other conditions and combinations within the combined intervention (Table 21).

Conversely, Tau-U analysis yielded nonsignificant and moderate effect sizes ranging from 20% to 31% for most conditions associated with the combined intervention for the frustrational group. There were a few exceptions to this pattern. First, frustrational group B (BF) showed less improvement when the tutor condition was compared to the baseline condition as a nonsignificant Tau-U value ($p = .30$) indicated only 18% improvement for this condition. However, some significant gains were also observed. While the moderate improvement observed for the tutee or tutor condition within the combined intervention for group BF was not significant, the combined intervention in its entirety was found to yield a significant and moderate effect size as indicated by a Tau-U value ($p = .04$) of 24%. When considering groups AF and BF together, significant Tau-U values ($p = .03; p = .007$) indicated 28% and 25% growth for the tutee condition and the combined intervention respectively. These values suggest that the groups experienced moderate growth.

When considering the Tau-U data together, an interesting pattern emerges. As previously noted, instructional group BI showed moderate growth indicating 28% improvement upon implementation of the peer tutoring intervention, while frustrational group BF showed no improvement. When the combined intervention was introduced, effect sizes became larger for both groups widening the performance gap between instructional groups AI and BI and frustrational groups AF and BF. Also of note was the fact that the Tau-U values observed for the frustrational groups for the combined intervention ranged from 18-31%, which is similar to the Tau-U values, which ranged
from 26-28%, observed for group BI during the peer tutoring intervention. That is, the frustrational group began to make improvements at the end of the study that were comparable to those observed for the instructional group at the start of the study. This finding is noteworthy, as 3 students from the frustrational group earned mean scores for the last week of the study that fell in the instructional range. Thus, students in the frustrational group began to perform in the instructional range and effect sizes paralleled those of the instructional group at the start of the study. The implications of this finding will be discussed in the next chapter.

Table 22

_Tau-U Effect Sizes for Control Condition Across Instructional Groups_

<table>
<thead>
<tr>
<th>Group/Condition</th>
<th>95% CI</th>
<th>p-value</th>
<th>Group/Condition</th>
<th>90% or 95% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI PT</td>
<td>-0.30</td>
<td>.07</td>
<td>BF PT</td>
<td>-0.30</td>
<td>.05</td>
</tr>
<tr>
<td>AI C</td>
<td>0.31</td>
<td>.000</td>
<td>AF Combined</td>
<td>-0.31</td>
<td>.00</td>
</tr>
<tr>
<td>BI C</td>
<td>0.32</td>
<td>.000</td>
<td>BF C</td>
<td>-0.08</td>
<td>.16</td>
</tr>
<tr>
<td>AI + BI C</td>
<td>0.21</td>
<td>.000</td>
<td>AF + BF C</td>
<td>-0.10</td>
<td>.09</td>
</tr>
</tbody>
</table>

_Note._ CI = Confidence Interval; LL = Lower Level; UL = Upper Level; A = Group A; B = Group B; I = Instructional Level; F = Frustrational Level; PT = Peer Tutoring Intervention; PF = Performance Feedback Intervention; C = Combined Intervention; Boldface Print = p < .05.

Another interesting pattern emerged during the control condition for those students performing at the instructional level. As evidenced by negative or very small effect sizes, Tau-U analysis indicated that little to no growth occurred for the control set across intervention phases for students performing in the frustrational range (Table 22).
In contrast, Tau-U analysis indicated moderate and significant effect sizes throughout all phases of the study for students in the instructional range with respect to the control condition. It is noteworthy that a significant Tau-U value \( (p = .000) \) indicating 57% growth from the performance feedback intervention to the combined intervention was obtained for group AI. As previously noted, 3 out of 4 students within this group made significant gains that were not attributed to intervention effects. The magnitude of the effect size for the control condition supports the theory that repeated practice was beneficial for these students and the peer tutoring component of the combined intervention, which indicated a similar rate of improvement of 60%, was not a necessary intervention component for these students. Conversely, group BI showed smaller, but still moderate effect sizes as evidenced by Tau-U values \( (p = .07; p = .02) \) indicating 23% and 26% improvement when comparing DCPM scores in the baseline to peer tutoring intervention and the peer tutoring to the combined intervention respectively.

While these effect sizes are similar to those observed for the tutee and tutor conditions within the peer tutoring intervention, effect sizes for sets targeted by the peer tutoring component in the combined intervention were twice as large. Thus, it seems that while repeated practice and performance feedback was enough for students in the instructional range to demonstrate improvement, several students progressed faster on sets targeted by the peer tutoring component within the combined intervention.

**T-tests for the instructional level.** Based on visual and Tau-U analyses, it is clear that students in the instructional range made greater gains throughout the course of this study. Independent two-tailed t-tests were used to evaluate the significance of this difference. First, there was a significant effect \[ t(25) = 2.98, p = .03 \] for instructional
level when considering student performance on the performance feedback intervention (group A). This is not surprising as we have previously noted that although growth was only attributed to intervention effects for one student, students performing in the instructional range made gains throughout the performance feedback intervention phase, while students performing in the frustrational range showed little growth.

Conversely, there was no significant effect \( t(6) = 1.79, p = .12 \) for instructional level when considering performance on the peer tutoring intervention (group B). This finding is surprising given that Tau-U analysis indicated statistically significant and moderate growth as the result of the peer tutoring intervention for students in group B1 while no significant gains were made for students in group Bf. However, as previously noted, the peer tutoring intervention was only partially verified as effective for students in the instructional range, which indicated that growth was not considerable across sets for these students. Although there was a demonstrable difference between these two groups based on visual analysis and Tau-U analysis, it was not large enough to be considered statistically significant.

However, a different pattern emerged for the combined intervention as there was a significant effect \( t(9) = 2.75, p = .02 \) for instructional level within group B, but there was no significant effect \( t(5) = 1.07, p = .33 \) for instructional level observed for group A. At first glance, this result is surprising, as visual analysis confirmed that the combined intervention was found to be effective for students in both the instructional and frustrational ranges within group B, while only students performing in the instructional range of group A responded positively to the combined intervention. However, Tau-U analysis found a larger, though nonsignificant effect size \( (p = .36) \) of 29% for students in
group AF, while a significant Tau-U value \((p = .04)\) indicated an effect size of 24% for group BF. Consequently, the difference in Tau-U values is the same when comparing group AI to AF as it was when comparing BI to BF. However, because there were fewer students in group A and visual analysis of the graphs for students 5, 6, and 7 indicated erratic performance, the effect size of 29% for AF cannot be taken at face value. It is also important to reiterate that visual analysis of the graph for student 6 indicated that DCPM scores dropped in the performance feedback range and then returned to baseline levels during the combined intervention phase. The Tau-U value and corresponding z-score would have noted this growth as though the combined intervention was effective for this student. These findings highlight the importance of visual analysis for single-case research, as once again, the statistics do not always accurately reflect intervention effects.

When comparing all students in the instructional group to all students in the frustrational group, there was a significant effect for instructional level \([t(9)=2.75, p = .02]\), with students in the instructional group outperforming those in the frustrational group on the combined intervention.

**Summary.** In sum, students in the instructional range responded better to all three interventions, although it is noteworthy that the growth observed throughout the performance feedback intervention was not attributed to intervention effects and the benefits of the peer tutoring intervention were not as generalizable as those found for the combined intervention. Further, students in the instructional range generally showed greater rates of improvement even when the improvement could not always be attributed to intervention effects. However, given that visual analysis confirmed that the tutee condition of the combined intervention was verified as effective for students performing
at the frustrational levels, nonsignificant Tau-U values indicated moderate improvement, and t-values indicated that differences between groups were not always statistically significant, there was evidence to suggest that the combined intervention was also useful for students performing in the frustrational range, albeit, to a lesser extent.

**Mastery levels of performance.** The seventh, and final, research question of this study sought to determine if instructional level at the start of the study impacted the speed at which students would achieve mastery levels of performance.

The answer is clear as the difference between the instructional and frustrational group became more obvious when considering the number of students who reached mastery levels of performance. Within the instructional range, 9 of 10 students consistently (at least 3 consecutive sessions) reached mastery levels of performance for at least one set, while only 1 of 8 students from the frustrational range earned DCPM scores in the mastery range. Not only did the instructional range impact the rate at which students could achieve mastery levels, it predicted whether they would achieve mastery levels throughout the course of the intervention. As only students from the instructional range obtained mastery levels of performance, clearly they did so quicker than students from the frustrational range.

When comparing initial baseline mean scores to average scores from the final week of the study, the mean level of improvement for students who were performing in the instructional range was 23.66 (12.09) DCPM, while the mean level of improvement for students in the frustrational range was 2.64 (5.53) DCPM. There was a ten-fold difference. While logic might dictate that students performing in the instructional range require less growth to reach mastery levels than students performing in the frustrational
range and, as a result, students performing at the instructional range would obviously meet mastery levels faster, another factor seems to be lurking in the background. These data suggest that students need a sufficient foundation in order to experience academic achievement within the context of these interventions. That is, there needs to be a minimum level of computational fluency in order to increase that fluency.

Although, as a group, students performing in the frustrational range showed significantly less improvement than did their instructional counterparts, as previously noted, all students from the frustrational group began to achieve scores in the instructional range, though sometimes inconsistently, throughout the course of this study. Simultaneously, Tau-U values began to increase, indicating greater magnitudes of growth. This finding will be further explored in the next chapter.

**Summary of Research Hypotheses.**

As was described, hypothesis 8 was supported, while hypotheses 2, 3, 4, 5, and 6 were partially supported. Each of the six hypotheses and the corresponding support and sources of data are presented in Table 23. Hypotheses 1, 6, and 7 were not supported.
### Table 23

**Summary of Research Hypotheses and Findings**

<table>
<thead>
<tr>
<th>Research Hypotheses</th>
<th>Findings</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increased fluency scores will translate to a discernible increase in student scores on an applications and concepts measure.</td>
<td>Partially Supported</td>
<td>Significant Pearson’s Product Correlation between fluency change scores and pre-post-test scores on the M-COMP. Significant correlation between fluency change scores and the M-CAP for students in the instructional group.</td>
</tr>
<tr>
<td>Students’ fluency scores will increase as a result of the performance feedback and performance feedback interventions.</td>
<td>Partially Supported</td>
<td>Varied increases in means during the intervention phase. Adequate replication of prediction and verification, varied changes in data trends, and differentiation for the tutee condition of the peer tutoring condition. Tau-U values were indicative of growth.</td>
</tr>
<tr>
<td>All students will show greater increases in fluency scores when the two interventions are combined compared to their performance as a result of either intervention in isolation.</td>
<td>Partially Supported</td>
<td>Varied increases in means during the intervention phase, differentiation established between sets for several students, adequate replication of prediction and verification, varied changes in data trends, and increased effect sizes.</td>
</tr>
<tr>
<td>Students will show greater improvements on problems included in the tutee condition than problems that are included in the control and tutor conditions.</td>
<td>Partially Supported</td>
<td>Varied increases in means during the intervention phase, differentiation established between sets for several students, adequate replication of prediction and verification, varied changes in data trends, and increased effect sizes.</td>
</tr>
<tr>
<td>Students who begin the intervention phases at the instructional range will reach mastery levels of performance sooner than students who begin the intervention at the frustrational range.</td>
<td>Supported</td>
<td>All students in the instructional range consistently met mastery levels of performance as evidenced by three consecutive scores for at least one set. No students in the frustrational range consistently met mastery levels of performance.</td>
</tr>
</tbody>
</table>
Chapter 5

Discussion

This chapter provides a detailed discussion of the results in light of current research. Educational implications of the findings, limitations of the study, and suggestions for future research are also presented.

Research has shown that academic difficulties can result from a mismatch between student skill and instructional interventions (Ardoin & Daly, 2007; Gickling & Thompson, 1985; Haring et al., 1978). The instructional hierarchy (Haring et al., 1978) is a heuristic that has been used in conjunction with the instructional delivery model (Gickling, 1977) as a guide for matching students’ instructional needs to appropriate interventions for over 35 years. However, relatively few research studies have shown how this heuristic can be applied to the selection and utility of brief, explicit, evidence-based interventions in the area of mathematics (Codding et al., 2011).

The current study is unique in several ways. First, few studies have utilized the relatively new, empirically derived fluency criteria presented by Burns et al. (2006) to ensure a match between students’ skill levels and academic interventions. And, in fact, very few studies have explored the utility of acquisition and fluency interventions when considering the impact of students’ instructional level at the start of the study (Codding et al., 2011). Fewer studies have examined the effects of a multi-component treatment that includes both acquisition and fluency intervention components. Finally, this study sought to address a gap in the research with regard to generalizability of intervention effects.

The current study sought to expand the research literature and extend our understanding of how the instructional hierarchy can be used to make instructional
decisions based on elementary-age students’ performance of basic math skills. It investigated the utility of combining acquisition and fluency interventions, which has only been done once before (Rhymer et al., 2000). In addition, this study investigated the use of performance feedback in isolation for the purpose of improving fluent responding to basic math facts. Thus, this study sought to expand our understanding how brief, explicit, and evidence-based interventions can be used in the general education setting in order to improve fluency skills of elementary-age students in the area of mathematics. The results of this study should help educators make better decisions when matching interventions to student need.

The results indicated that a majority of students in the instructional range showed considerable improvement over the course of the study, although not all improvement could be attributed to intervention effects. Conversely, only one student in the frustrating range showed considerable improvement. All but one student who began the study in the instructional range met mastery levels of performance, and the improved scores correlated with generalizability measures. A major finding was that established fluency criteria did not identify difficulties related to accuracy of performance. The implications and explanations of these findings are presented within this chapter.

**SES as a Mitigating Factor**

In order to explore the potential impact of SES, a school was chosen in which 53% of students qualify for free or reduced-price lunch (Vermont Agency of Education [VAE], 2015) for this study. Despite school-wide SES representation, when the sample was established, only 5 of 18 students, or 28% of the sample, fell into this category. Although, it is unknown why more students within the study did not qualify for free and
reduced lunch, it is possible that parents of students who qualify for free and reduced lunch were less likely to give consent for participation.

Research has found that parents of children from higher SES are more likely to give consent for participation in research studies than are parents of poverty (Heinrichs, Bertram, Kuschel, & Halweg, 2005; Spoth, Redmond, Hockaday, & Shin, 1996). Additionally, it has been found that parents of lower SES are more concerned about privacy than are parents of higher SES (Spoth et al.). Participation in this study required that parents allow the school to provide the PI with lunch eligibility information regarding each student. It is possible that parents of students who qualified for free or reduced lunch were less willing to share this information. Given the small number of students in poverty who participated in this study, conclusions could not be made regarding the impact of SES on the nature of the development of mathematical fluency.

**Treatment Acceptability**

The results of the Kids Intervention Profile (KIP) were overwhelmingly positive at both the intervention and individual-question levels. The data from the acceptability measure indicated that students found these interventions to be useful and they enjoyed participating in the procedures. This finding is important because treatment acceptability has been found to be an inherent part of successful interventions (Skinner et al., 1997). One way to define success is participants transferring research to practice in the school setting. Treatment acceptability is a tool for determining the likelihood of that transfer (Riley-Tillman et al., 2005). If students perceive an intervention as helpful or beneficial, they may apply it to other aspects of their academic experience (Skinner et al., 1989). Treatment acceptability has also been found to impact student outcomes with regard to
improved fluency scores for basic math facts (Allinder & Oats, 1997). Thus, positive opinions regarding these interventions indicates that students are more likely to engage in and respond well to similar treatments in the future.

It is noteworthy that students indicated a slightly negative response with regard to writing a math fact following a mistake within the peer tutoring intervention. This is important because this part of the peer tutoring intervention was developed from error-correction procedures in which an instructor prompts a learner to elicit a correct response following the occurrence of an error by punishing that error (Worsdell et al., 2005). Thus, a slightly negative response with regard to this procedure was to be expected and desired.

**Generalization**

Whether increased fluency scores for students could be generalized to improvement on more complex mathematical tasks is an important question because the current mathematics curriculum focuses instruction on concepts and applications and has de-emphasized the direct instruction of computational competency (Shapiro, 2011). Despite this new focus, it has been reported that relatively few acquisition interventions have investigated whether or not increased computational fluency has any impact on performance of more complex tasks (Codding et al., 2009). At the same time, it has been said that teaching academic skills through drill tasks leads to increases in student performance in other related skills (Burns, 2004).

All participants began this study performing in the frustrational and instructional ranges of the instructional delivery model (Deno & Mirkin, 1977) based on the new fluency criteria (Burns et al., 2006). While 10 students in this study improved to mastery
levels of performance, 8 remained in the frustrational or instructional ranges of performance at the completion of this study. This finding indicates that even when growth had occurred, less than half of the students earned scores that indicated they were ready to generalize single-digit multiplication skills to more complex tasks according to recommendations from the instructional hierarchy (Haring et al., 1978). Further, students who reached the mastery range of performance were just beginning to enter the generalization stage of the instructional hierarchy, indicating that they were likely to make mistakes when attempting to generalize the newly acquired skill. As such, they were likely to be in need of explicit generalization training in order to effectively transfer these skills (Haring et al., 1978).

Findings with regard to generalization of math fact fluency in the literature are mixed. Codding et al. (2009) found that increased fluency for single-digit multiplication and multi-digit addition problems did not generalize to slightly more advanced skills. Yet, when assessing for generalization between tasks that were similar in nature, Codding et al. (2010) found that increased DCPM scores for multiplication facts generalized to improved performance on a similar measure in which the practiced facts were presented in a different form. Generalization was also observed, though to a lesser degree, for word problem probes which required the participant to recognize the math facts in the context of language. These findings align the notion that generalization is more likely to occur if stimuli are similar (Alberto & Troutman, 2009). Similarly, Poncy, Duhon, Lee, and Key (2010) note that as the difference between the trained behavior (i.e. single-digit multiplication) and the generalized behavior (i.e. ability to complete applications and concepts measures) increases, so does the need for more intensive generalization training.
The current findings were not consistently aligned with existing research. When considering data from all participants, there was a significant correlation between improved fluency scores and improved scores for the M-COMP, which is a skills-based assessment. On the other hand, there was no significant correlation was observed for the M-CAP, which assesses mathematical concepts. This finding is not surprising as the skills measured on the M-COMP are more closely aligned with the skill (single-digit multiplication) targeted in this study.

When examining the correlations for the 11 students who began the study in the instructional range and who consistently met mastery levels of performance, a nonsignificant correlation was present between the increased fluency scores and increased scores on the M-COMP, while a significant positive correlation was observed between fluency scores and the M-CAP. This finding is different than the whole-group results and directly contradicts previous findings in the literature. Careful examination of individual scores did not provide any clues. Future research is required to see if this finding is replicated.

**Effectiveness of the Performance Feedback and Peer Tutoring Interventions**

Based on guidelines for interpreting single-case research (Horner et al., 2005), the tutee condition was the only intervention component that was verified as effective. Consequently, the peer tutoring intervention was partially effective. Further, only 1 student responded positively to the performance feedback intervention. Interestingly, Tau-U analysis for these interventions indicated a moderate and significant effect size when comparing DCPM scores from the baseline phase to the performance feedback phase, while no significant growth was observed with respect to the peer tutoring
intervention. When considering Tau-U analysis in conjunction with visual analysis significant effect sizes were only observed for 2 students who displayed improvement that could be attributed to intervention effects. When taken together, the data suggest that neither intervention was consistently effective for the purpose of improving students’ fluency scores.

**Performance Feedback.** There are several factors that may have contributed to the ineffectiveness of the performance feedback intervention. First, students may have had difficulty integrating the feedback they were given. Similar to the structure of feedback established by Rhymer et al. (2000), each student was given a sheet with the highest DCPM score he or she had earned during any previous assessment prior to completing assessment probes. Students in this study did not know to which set of problems to attribute their score and, as a result, did not understand the nature of the feedback they received. It has been found that student performance can be impacted by the level of the specificity attributed to performance feedback (Goodman, Wood, & Hendrickx, 2004). In this case the feedback was confusing to students, and perhaps not salient or specific enough to lend itself to improved performance.

Second, DICPM scores for 5 of 7 students who received the performance feedback intervention, were relatively high, indicating low rates of accuracy. As noted, the performance feedback condition was a fluency intervention. Within this intervention, the feedback targets the number of items completed and it does not provide information with respect to accuracy. Thus, it was important that students respond accurately before engaging in this intervention (Codding et al., 2007). Our results were consistent with previous research, as students who were unable to respond accurately did not show
improvement as the result of this intervention. Moreover, only 1 of the 2 students (student 3) who was performing with high levels of accuracy at the start of the study responded positively to this intervention.

Another interesting observation was made for two students who showed improvement that began in the baseline phase and continued into the performance feedback phase without a positive change in the data path. This means that improvement could not be attributed to intervention effects. However, it is possible that the data collection procedures were, in and of themselves, reinforcing for these students. This finding and the research associated with it will be discussed later in this chapter within the context of instructional level.

In sum, there are a few possible explanations for ineffectiveness of the performance feedback intervention. First, students’ inability to connect feedback with a specific probe was problematic. Second, the feedback in this intervention targeted fluency while students required corrective feedback. Lastly, based on the performance of two of the participants, the role of timed repeated practice on improved performance is unclear.

**Peer Tutoring.** When considering the inconsistent results associated with the peer tutoring intervention, it is possible that the intervention was lacking in dosage or intensity. For example, it has been found that students who participated in fewer peer tutoring sessions were less likely to show notable improvement with regard to fluency scores (Hawkins et al. 2009, Rhymer et al., 2000). In the current study, students only acted as the tutee and tutor for five sessions each, with the exception of 2 students. Further, students practiced for 2 minutes a day and acted as either the tutor or tutee
during each session. Although previous interventions have yielded positive results with respect to peer tutoring interventions, longer practice sessions (10 and 4 minutes respectively) were implemented and students acted as both the tutee and tutor in each session, which means they were able to benefit from both conditions (Hawkins et al.; Rhymer et al.). Further, Gersten, Beckmann, et al. (2009) note that students should devote approximately 10 minutes in each intervention for building fluent retrieval of basic arithmetic facts.

This explanation was further supported as 4 students began to earn improved scores towards the end of the peer tutoring phase, though some were not consistent enough that replication was present. However, when considering that each data point predicts the direction of the data path (Richards et al, 1999), one could infer that improvement would continue. Given that less time was devoted to practice during each session, perhaps more sessions would have been necessary to determine whether or not the peer tutoring intervention was truly effective.

Further, the type of error-correction procedure may also have been lacking in intensity, which could have negatively impacted the effectiveness of the peer tutoring intervention. Whenever students made a mistake during the tutoring session, a single-response repetition procedure was implemented. Literature has yielded mixed results with regard to the effectiveness of this type of procedure (Ferkis et al., 1997; Worsdell et al., 2005). Ferkis et al. (1997) found no difference when students had to correct an error five versus one time. In contrast, Worsdell et al. (2005) conducted multiple studies and found that increased opportunities for error-correction produced better results. Given contradicting findings, more research is needed to determine the ideal number of
repetitions for these procedures. It is possible that a single-repetition response condition was not sufficiently reinforcing to yield positive results within the context of this study.

Finally, according to the instructional hierarchy, modeling is an important component of acquisition interventions (Haring et al., 1978). In the current study, the error-correction procedure did not include a component of modeling unless an error occurred. While this type of procedure has been used in previous studies (Ferkis et al., 1997; Marvin et al., 2010; Rapp et al., 2012; Rhymer et al., 2000; Worsdell et al., 2005), other types of acquisition interventions (e.g. Cover, copy, compare, incremental rehearsal, and self-instruction) include a modeling component for every presented problem (Burns, 2005; Coddington et al., 2006; Coddington et al., 2007; Coddington et al., 2009; Coddington et al., 2010; Meichenbaum & Goodman, 1971; Skinner et al., 1997; Wood et al., 1993). Thus, it is possible that the modeling component was also lacking in peer tutoring intervention, which further contributed to a weaker dosage of intervention.

Another interesting finding is that while accuracy improved for 7 students for one or more conditions within the peer tutoring intervention, this improvement only coincided with improved fluency scores for 3 students. Thus, while the peer tutoring intervention was successful in terms of correcting errors, this improvement did not always correlate with increased fluency scores. This finding is interesting as the instructional hierarchy requires students to master a specific academic skill before progressing to the next one, and in the first stage, acquisition, students perform slowly and inaccurately while working toward understanding the skill (Haring et al., 1978). While this study sought to improve accuracy and fluency simultaneously; perhaps this was a lofty goal. This notion will explored within the context of instructional level later in this chapter.
Summary. This study found that the performance feedback and peer tutoring interventions were equally ineffective in improving student fluency scores. However, it is noteworthy that the peer tutoring intervention had an obvious positive impact on student accuracy that did not always result in increased fluency scores.

The Combined Intervention

Visual analysis indicated that intervention effects were present for one or more conditions within the combined intervention for 10 students, which is double the number of students who improved as a result of the single interventions. Further, Tau-U analysis indicated that significantly more growth occurred in the combined intervention than in the performance feedback or peer tutoring interventions. Additionally, 6 students displayed improved accuracy for one or more conditions, with most notable changes in DICPM scores observed in the tutee condition. Interestingly, improved accuracy coincided with improved fluency scores for 5 of these 6 students. This finding is noteworthy as improved accuracy within the peer tutoring intervention phase did not typically coincide with improved fluency scores.

The current results indicate that combined intervention was more effective than the either of the single interventions. Although independent t-tests indicated that the difference between the two interventions was not statistically significant, the results from the t-tests must be interpreted with caution, as they are based solely on z-scores generated as part of the Tau-U analysis, which means that results are confounded by growth that was achieved, but not attributed to intervention effects. Thus, the data suggest that the combined intervention was more effective than either the performance feedback or peer
tutoring interventions in isolation, although the difference may not be statistically significant.

The current finding is aligned with and adds to the sparse body of research about multi-component interventions. In a meta-analysis conducted by Codding et al. (2011), they found that students who were performing in the frustrational range benefited more from interventions that included multiple treatment components. Further, in a similar study, Rhymer et al. (2000) found that when they added the performance feedback intervention to the peer tutoring intervention, 3 out of 4 participants showed an increase in fluency scores with the addition of this component, indicating that performance feedback increased the effectiveness of the intervention.

However, in the current study a change in data path was not consistently observed from the peer tutoring to the combined intervention phase. Consequently, one must question the true effectiveness of the combined intervention. First, as previously noted, because the performance feedback component lacked specificity, it was possible that it was not differentiated from the naturally occurring feedback that was present within baseline procedures. Further, the peer tutoring intervention may have lacked sufficient intensity and duration and it is possible that additional sessions would have led to greater levels of growth. Thus, it is possible that the peer tutoring intervention and the combined intervention were similar enough that the improved performance within the combined intervention phase would have been observed had the peer tutoring intervention continued without the additional performance feedback component. It is important to note that performance feedback component clearly had a positive impact on scores for 1 student, but this finding was not as clear for all other students. Further research would be
necessary to determine the true value of the performance feedback condition in conjunction with the peer tutoring intervention.

When considering the data, this hypothesis was partially supported. Further research would be necessary to make a definitive statement about the improved effectiveness of the intervention due to the additional performance feedback component.

Comparison of the Tutee, Tutor, and Control Conditions.

When considering student performance across both the peer tutoring and combined intervention phases, a total of 9 students improved as a result of the tutee condition in one or both of the intervention phases, while the tutor condition was verified as effective for 6 students. Therefore, both conditions were established as effective, as there were a sufficient number of students who responded well to each condition according to established guidelines (Horner et al., 2005).

Within the peer tutoring intervention, the superiority of the tutee condition as compared to the tutor condition was not evident based on analysis of effect size, while in the combined intervention, the tutor, the tutee and the control conditions all yielded significant and moderate effect sizes for groups A and B. Tau-U values indicated the least growth was observed for the control set, with more growth experienced for the tutor set, and the most growth noted for the tutee set. Although visual and Tau-U analyses indicated that the tutee condition was slightly more effective than the tutor condition, an independent $t$-test demonstrated that the discrepancy between the two conditions was not significant. Further, neither was significantly better than the control condition.

Current research supports these findings. A literature review conducted by Robinson et al. (2005) found that positive academic outcomes have been noted for both
the tutee and tutor. Further, several studies have found that while both roles benefit students, the tutee condition generally yields better results (Menesses and Gresham, 2009; Rhymer et al., 2000; Robinson et al.). Moreover, as was observed within the current study, Rhymer et al. found that students showed more improvement on the problems assigned to the tutor and tutee problems as compared to the control problems, indicating that there was a benefit to both the tutor and tutee conditions. In another study, Menesses and Gresham noted that students showed more improvement within a reciprocal peer tutoring model than the nonreciprocal approach because twice as many students are working to develop a particular skill in the same time frame.

The peer tutoring procedures in this study imitated those described by Rhymer et al. (2000). Specifically, the tutor condition included evaluating a peer’s responses. The tutee condition included explicit timing, active responding, and immediate corrective feedback in the form of positive practice overcorrection; these components are designed to increase accuracy and fluency according to recommendations for students in the acquisition and fluency stages of the instructional hierarchy (Haring et al., 1978). Therefore, it is not surprising that the tutee condition yielded greater results than the tutor or control condition. Additionally, given the existing research, it is also not unexpected that the tutor condition led to greater improvements than the control condition (Menesses and Gresham, 2009; Rhymer et al., 2000; Robinson et al., 2005).

In sum, visual and Tau-U analyses indicated that growth rates were higher for the tutee condition than the tutor condition, while the tutor condition led to higher growth rates than the control condition. However, the independent t-test indicated that none of these differences were statistically significant.
**Instructional Level as a Mediating Factor**

**Performance feedback intervention.** With respect to the performance feedback intervention, Tau-U analysis indicated a significant and moderate effect size for the instructional group and an independent *t*-test showed that the difference between groups was significant. However, these findings were not based on values obtained from intervention effects. Consequently, the intervention was not deemed effective for either group of students. There are several factors that may have impacted the effectiveness of the performance intervention.

Current research suggests that fluency scores are a reliable criterion for choosing instructional interventions. For example, Codding et al. (2007) found that students performing in the instructional range responded better to a fluency intervention than they did to an acquisition intervention. However, it has also been suggested that accuracy be considered when choosing appropriate academic interventions. More specifically, two studies found fluency interventions were effective for increasing students’ speed of performance, but only when students presented with high rates of accuracy at the start of the study (Rhymer & Morgan, 2005; VanHouten & Thompson, 1976). Similarly, it was noted that students receiving fluency interventions must be performing with high rates of accuracy before beginning the intervention (Codding et al., 2007).

Given that the groups were determined based on Burns et al. (2006) fluency criteria, accuracy rates were not taken into account. When considering the error rates for students in the instructional range who received the performance feedback intervention, DICPM scores were relatively high for all students with the exception of 2 of 7 students, who were both performing in the instructional range. As these two students were the
only ones appropriately matched to the performance feedback intervention, it is not surprising that the performance feedback intervention was not consistently successful for this group of students.

Second, based on existing literature, it was predicted that students performing in the frustrational range would display little or no improvement as a result of the performance feedback intervention. This prediction was confirmed. As students performing in the frustrational range require acquisition interventions that provide immediate corrective feedback and modeling, a fluency intervention would not be a good instructional match as students would repeatedly practice mistakes (Haring et al., 1978). Research has supported this claim as Codding et al. (2007) found that a fluency intervention was not as successful as an acquisition intervention for students performing in the frustrational range. Further, Burns et al. (2010) conducted a meta-analysis in which it was found that fluency interventions produce small to moderate effect sizes when they are applied to students in the frustrational range of performance.

**Peer tutoring intervention.** When considering group performance, 2 students in each instructional range showed intervention effects for the tutee condition and 1 student from each instructional range showed intervention effects for the tutor condition. Tau-U analysis indicated that more growth occurred for the instructional group as compared to the frustrational group. However, an independent *t*-test indicated that the difference between groups was not significant.

Once again, these findings contradict existing research. Burns et al. (2010), in a meta-analysis, found that acquisition interventions yielded large effect sizes for students performing at the frustrational level, while small to moderate effect sizes were observed
for students in the instructional range. Additionally, as previously noted, Codding et al. (2007) found that students performing in the frustrational range responded better as compared to students performing in the instructional range to an acquisition intervention.

However, upon closer inspection of the data, only 1 student was consistently performing with high rates of accuracy, indicating that all students in group B were in need of the immediate corrective feedback that was provided within the peer tutoring intervention. Students who were performing in the instructional range were already working faster than those students in the frustrational range. Thus, as accuracy improved upon implementation of the peer tutoring intervention, DCPM scores began to increase at a faster rate than was observed for students in the frustrational range. This finding is consistent with Codding et al. (2007)

Although intervention effects were only attributed to 3 of 6 students in the instructional group, two additional students showed significant growth that was not attributed to intervention effects. Thus, only one student in the instructional group failed to make consistent gains in fluency rates, although improved accuracy was observed for this student. Several implications arise from these findings. First, the peer tutoring intervention with the embedded correct response repetition component, was found to be effective for increasing accuracy, and sometimes fluency, scores for several students in the instructional group. This finding is consistent with existing research, as improved accuracy and fluency scores have been observed as the result of correct response repetition interventions in the areas of reading (e.g., Ferkis et al., 1997; Marvin et al., 2010; Worsdell et al., 2005) and mathematics (Rapp et al., 2012; Rhymer et al., 2000).
Conversely, only two students in the frustrational group showed minimal improvement and no growth was observed for the rest of the frustrational group. While intervention effects were present for these students due to differentiation, no change in data path observed.

There are several possible explanations for these findings. First, as previously noted, the intensity of the peer tutoring intervention may not have been sufficient for students performing in the frustrational range. Additionally, Codding et al. (2007) found that students who were performing in the frustrational range responded poorly to explicit timing procedures. As explicit timing component was embedded in the tutoring procedures (each student was tutored for two minutes) and in the data collection procedures (three probes for two minutes each), perhaps these students who lack automaticity were frustrated or flustered by this treatment component. Further research would be necessary to explore this finding.

**Combined intervention.** Visual analysis for the combined intervention indicated that one or more conditions within the combined intervention were verified as effective for 6 students who were performing in the instructional range and for 3 students who were performing in the frustrational range. Tau-U analysis confirmed that gains for the instructional group were approximately double those observed for the frustrational group. When comparing the instructional group to the frustrational group, an independent t-test indicated that the difference between the groups was significant.

Although it has been suggested that combined interventions including both acquisition and fluency interventions may be beneficial for students, relatively little research exists to support this statement (Burns et al., 2010; Codding et al., 2011). As a
result, it was difficult to predict how students from either instructional group would respond to the combined intervention.

Two students from the instructional group who had participated previously in the performance feedback intervention showed immediate improvement with regard to accuracy and fluency for sets targeted by the peer tutoring component of the combined intervention, particularly with regard to the tutee condition. These findings highlight the importance of using acquisition interventions for students who struggle to perform accurately as growth for both students had largely stalled and error rates remained high until the introduction of the immediate corrective feedback within the peer tutoring intervention component.

It is noteworthy that a third student also began to show improved scores as a result of the peer tutoring component of the intervention, even though he was already completing the probes with accuracy. This finding contradicts the notion that students performing in the instructional range do not respond as well to acquisition interventions (Burns et al., 2010, Codding et al., 2007). Perhaps the inclusion of fluency intervention components within the combined intervention mitigated these negative effects that have been reported in the literature, as speed of performance continued to be targeted. Further research would be necessary to support this claim.

Interestingly, students performing in the instructional range continued to perform at rates similar to those observed during the peer tutoring phase or growth plateaued as students met mastery levels of performance. Additionally, only 2 students showed notable improvement with regard to the control condition. Thus, little evidence exists to suggest that the added component of performance feedback impacted student
performance. Rhymer et al.’s (2000) research supports this finding as they noted that because the performance feedback condition was introduced after, and in conjunction with, the peer tutoring intervention, performance within the combined intervention was likely confounded by sequence effects. Based on the current data it is impossible to determine if the added component of performance feedback led to the combined intervention being more effective than the peer tutoring intervention, or if the peer tutoring intervention would have continued to be effective as it was.

An interesting pattern emerged with respect to the frustrational group. Although consistent growth was only observed for 2 students, 5 additional students experienced small and inconsistent gains. Further, Tau-U analysis indicated moderate and significant growth for students performing in the frustrational range with respect to the combined intervention. This finding is noteworthy, because the Tau-U value for students in the frustrational range within combined intervention phase was larger than the growth observed within the peer tutoring intervention phase for students who were performing in the instructional range at the start of the study. Further, 3 students who began the study in the frustrational range, earned mean scores for the last week of the study that fell in the instructional range. Given that students in the frustrational range began to display improvement similar to students in the instructional range at the start of the study, we can hypothesize that improvements would continue over time. This finding raises further questions regarding whether or not students in the frustrational range simply require a higher intensity intervention, which could include additional sessions or more time per session, to practice the target skill in order to demonstrate faster rates of improvement (Coddington et al., 2011).
**Patterns across groups.** Four students in the instructional group demonstrated patterns of growth in which improvement began in the baseline phase and continued on a similar trajectory into the single intervention phase. Although growth slowed for 2 of these students, 2 students quickly reached mastery levels of performance. As this growth cannot be attributed to treatment effects, another possibility is worth noting.

Baseline and data collection procedures are similar to explicit timing procedures. As noted in chapter 2, explicit timing is a fluency intervention in which students are repeatedly asked to complete a timed academic task (Rhymer et al., 2002). Codding et al. (2007) report that performance feedback occurs naturally within this procedure as students receive feedback regarding how many problems they have completed, simply by looking at their worksheets. As a result, they may be intrinsically motivated to perform more quickly on each assessment. This procedure has been shown to be effective in numerous studies (e.g., Codding et al., 2007; Rhymer et al., 2002; Van Houten & Thompson, 1976).

Thus, these students may have benefited from the repeated practice and been motivated by the inherent feedback that was part of the data collection procedures. As 4 students in the instructional range replicated this pattern, it is a noteworthy finding. Further research is necessary to explore the types of performance feedback that could lead to improved fluency scores.

**Summary.** The performance feedback intervention was not effective for either instructional group, while students in the instructional range outperformed students in the frustrational range within the peer tutoring phase.
With regard to the combined intervention, students in the instructional group continued to outperform students in the frustrational group both in terms of the number of students who displayed improvement and with regard to the magnitude of improvement observed. However, as students in the frustrational group began to show minimal improvements, one must question whether students in the frustrational range simply require a higher intensity intervention to practice the target skill in order to demonstrate faster rates of improvement (Codding et al., 2011).

Data indicated that 4 students in the instructional group made gains that began during the baseline phase and continued. As baseline procedures are similar to explicit timing procedures, it is possible that the inherent performance feedback, timed repeated practice, and intrinsic motivation may have been sufficient for these students to make initial gains, thus confounding treatment effects.

In sum, these findings indicate that ensuring student performance is matched to appropriate interventions may be more complicated than the proponents of fluency criteria suggest (Burns et al., 2006; Daly & Martens, 1994; Deno & Mirkin, 1977; Vanderheyden & Burns, 2005). By relying solely on established fluency criteria (Burns et al.) for determining instructional interventions, students’ level of accuracy is overlooked. Therefore, to ensure interventions are carefully matched to students’ needs, it may be necessary to consider both fluency and accuracy of performance.

**Mastery Levels of Performance**

Most students in the instructional range consistently in this study met mastery levels of performance, while only one student in the frustrational range earned any scores in the mastery range. While the gap between the instructional and mastery ranges is
smaller than the gap between the frustrational and mastery ranges, the magnitude of improvement is a more striking difference between the two groups. This finding further suggests that the peer tutoring and combined interventions were not intense enough to meet the needs of students performing in the frustrational range. Although these interventions included carefully selected evidenced-based components such as peer tutoring (Hawkins et al., 2009; Kunsch et al., 2007; Meneses & Gresham, 2009; Robinson, et al., 2005), correct response repetition (Rapp et al., 2012; Rhymer et al., 2000), and the use of a multi-component intervention (Codding, 2011; Rhymer et al., 2000) that have been shown to be beneficial for students performing in the frustrational range, it is clear that they were not effective for these students.

As noted several times in this chapter, there are several possible explanations for these findings. First, the intensity of the peer tutoring intervention may not have been enough for students performing in the frustrational range (Hawkins et al., 2009; Rhymer et al, 2000). Second, the use of a single-response repetition procedure instead of a multiple-response procedure has been found to yield less improvement (Worsdell et al., 2005).

It is also possible that students performing in the frustrational range, who lack automaticity and accuracy, dislike being timed (Codding et al., 2007). As explicit timing was a major feature within this study, perhaps students in the frustrational group became frustrated or flustered by this treatment component. Further research would be necessary to explore these findings.
Educational Implications

This study adds to the literature to support the use of brief, evidenced-based math interventions for building students’ fluency when completing basic math facts in several ways. Perhaps one of the most important findings in this study relates to the ineffectiveness of the peer tutoring and combined interventions for the frustrational group of students. Although the interventions were carefully chosen to meet the needs of these students, it seems that the treatment dosage or intensity was not enough to yield positive results. This finding implies that researchers and educators must ensure that in addition to choosing an intervention that is aligned with the recommendations of the instructional hierarchy, the duration of an intervention must be long enough to be effective for these students. Further, it seems that the treatment dosage required to meet the needs of students performing in the instructional range is lower than that required for students in the frustrational range, even when both groups are struggling with accuracy.

A second finding, with important implications for practice, was with regard to generalizability. Increased fluency for single-digit multiplication correlated with improved scores on a more advanced computational measure and on a concepts and applications measure. Consequently, the use of brief interventions that target student fluency scores for basic math skills remains important even though the current mathematics curricula focus instruction on concepts and applications and have de-emphasized the direct instruction of computational competency.

A fourth finding is that students were successfully able to provide immediate corrective feedback to one another, which minimizes the educational resources necessary
to provide individualized interventions for students. This makes the implementation of acquisition interventions more feasible in the classroom setting.

A fifth finding is that fluency criteria may not have been sufficient for identifying students’ needs, as most students performing in the instructional range were also struggling to perform accurately. This finding implies that educators and researchers may want to consider both fluency and accuracy rates when choosing appropriate interventions.

Finally, students’ ratings on the treatment acceptability scale were relatively high, indicating that they enjoyed the intervention procedures and they found them to be helpful. This is important for practice, as students are more likely to actively engage in an academic intervention, rather than attempt to avoid or escape the intervention, if they find that the procedure is not aversive or useless in nature.

In sum, the findings of the current study indicate that brief evidence-based interventions, within the context of peer tutoring, are useful for improving students’ accuracy and fluency of performance when the intervention components are carefully matched to students’ instructional level. These improved scores were found to be associated with improvements on generalizability measures. Finally, these types of interventions are found to be generally liked and helpful by participants, indicating a high level of engagement.

Limitations

One limitation inherent to single-case research is generalization, which threatens the external validity of results. Due to the nature of using students from a single grade in a single school with specific instructional needs, it is difficult to predict how well other
students from other schools or grade levels would respond to the same intervention. As Richards et al. (1999) note, external validity is enhanced when systematic replication occurs. While several findings within this study were aligned with existing research, others clearly require further investigation.

There were also other limitations. First, rather than asking students to serve as both the tutee and tutor during every treatment session, the conditions were alternated on a weekly basis. This was done in order to alleviate the possibility of multiple treatment interference. However, this design was problematic for several reasons. As students were introduced to the intervention phases in a step-wise fashion, some students only acted as the tutee or the tutor within the context of the combined intervention, meaning that it was impossible to draw conclusions about which condition was more effective for these students. It also limited the ability to make comparisons with other students in the study and made it difficult to assess the effectiveness of the combined intervention as a whole. Further, it often happened that the tutor and tutee conditions were not evenly distributed within an intervention phase. This made it difficult to compare the conditions. Finally, by splitting up the peer tutoring intervention, the dosage of the intervention was reduced, which may have had a negative impact the effectiveness of this intervention for students who were performing in the frustrational range.

Another limitation of this study is with regard to the sample. The sample did not include an adequate number of students who received free and reduced lunch, which made it impossible to assess whether or not this is a mitigating factor with regard to the effectiveness of these interventions. Further, due to attrition, groups were not evenly distributed across interventions, as more students received the peer tutoring intervention
than the performance feedback intervention. This also impacted the stepwise introduction of the interventions within the multiple baseline across subjects design, which resulted in minimal variability in the length of treatment phases as initially planned.

Further, when comparing different interventions within the context of single-case research it is necessary that these treatments are sufficiently different from one another (Richards et al., 1999). Given the data of the current study, questions arose regarding the extent to which students were able to differentiate between the different procedures. As these conditions may have been too similar, it was difficult to draw strong conclusions regarding the addition of the performance feedback condition. This issue was further complicated when students were clearly confused by the nature of the performance feedback. As it has been found that student performance can be impacted by the level of the specificity attributed to performance feedback (Goodman et al., 2004), this may have been a contributing factor to the lack of intervention effects attributed to the performance feedback component.

Another limitation is with regard to the experimental design. Within an alternating treatments design, researchers typically semi-randomly assign the treatment condition to a session with no more than 2 or 3 consecutive sessions of one condition given (Richards et al., 1999). This practice enhances the experimental control of the independent variable. As the current procedures did not follow this recommendation, the ability to make comparisons across treatments was compromised.

A limitation with regard to the Tau-U effect sizes must also be noted. Parker et al. (2005) noted that autocorrelation violates the assumptions of most statistical
techniques and it often exists at problematic levels within single-case research. Although it has been noted that the Tau-U can be used when autocorrelation is present (Brossart et al., 2014), the problem of autocorrelation was not directly addressed within the context of the current study. Thus, the possible effects of autocorrelation on the Tau-U values is unknown.

An additional limitation of this study is that teacher acceptability could not assessed. This limits the external validity as the extent to which a teacher is willing or able to implement these procedures in the classroom setting is unknown.

Finally, as the number of research assistants desired to implement this study were not obtained and formal fidelity checks with regard to students’ adherence to intervention procedures were not obtained.

**Implications for Future Research**

As this study is one of the only studies to examine the effects of these types of intervention with regard to students’ instructional level (Burns et al., 2010) further research is necessary to understand how to best match students’ instructional level to appropriate interventions within the context of the instructional hierarchy. Previous research has suggested the use of fluency criteria (e.g. Burns et al., 2006; Daly & Martens, 1994; Deno & Mirkin, 1977; Vanderheyden & Burns, 2005) is superior to accuracy criteria when attempting to match student performance to appropriate interventions. However, in the current study it was found that high fluency rates were not always indicative of accurate responding, which indicates that further research is necessary to examine whether or not the use of fluency and accuracy criteria may be the best option for determining students’ instructional needs.
Furthermore, this is one of the only studies that has investigated the effects of a multicomponent intervention that includes both acquisition and fluency measures (Burns et al., 2010; Codding et al, 2011; Rhymer et al., 2000). While the current study found that students in the instructional range responded better than students in the frustrational range to these interventions, further research is necessary to replicate or refute this finding.

Additionally, intervention effects were not present for students performing in the frustrational range, even though interventions were developed to meet the needs of this group of students based on the instructional hierarchy (Haring et al., 1978). As this finding is contradictory to other research findings (Burns et al., 2010; Codding et al., 2007) further research is necessary to understand why students did not respond to this intervention package. It has been hypothesized that treatment dosage or intensity was not enough for these students to find success. As relatively little research exists with regard to treatment length and intensity, this is another area in need of further research (Codding et al., 2011). This is particularly true when considering that the intensity of treatment required for positive effects seems to differ depending on instructional levels of performance.

Similarly, this study was one of the only studies to investigate the use of overcorrection procedures with regard to mathematics. As current results are mixed with respect to the use of a multiple-response repetition as compared to single-response repetition procedures (Ferkis et al., 1997; Worsdell et al., 2005) and the single-response repetition was not successful for students performing in the frustrational range, this area is in need of further exploration.
Further research is also necessary to investigate the generalizability of these procedures to educational practice. As the teachers were not involved with these procedures, it is not known if these intervention procedures would be accepted and/or implemented with fidelity in the classroom setting.

Finally, this study sought to address the gap in literature that exists with regard to the effectiveness of evidenced-based treatments for students of poverty. As it failed in this pursuit, further research is necessary to ensure that interventions are effective for all students.

**Conclusion**

The current findings of this study indicate that a multicomponent treatment delivered within the context of a reciprocal peer tutoring intervention was effective for increasing fluency scores for single-digit multiplication for students who began in the study in the instructional range of performance. Further, these improvements correlated with generalizability measures, indicating that addressing basic math skills can generalize to improvements in more complex tasks within the scope of mathematics.

While there were several limitations to this study, this research provides valuable insight into the importance of matching student needs to academic instruction. More specifically, the current study presented a brief evidenced-based intervention that was found to be effective for students who presented with a particular skill set. Furthermore, it was demonstrated that students enjoyed a reciprocal peer tutoring intervention that allowed for students to receive the necessary immediate corrective feedback from each other, which allowed for individualized instruction or intervention to take place simultaneously with multiple students. Finally, it was found that both accuracy and
fluency may need to be considered within the context of the instructional hierarchy
(Haring et al., 1978) when choosing interventions.
Appendix A

CITY UNIVERSITY OF NEW YORK

The Graduate Center
Educational Psychology Program

PARENTAL/LEGAL GUARDIAN PERMISSION FORM FOR CHILD’S PARTICIPATION IN RESEARCH

Project Title: Examination of the Impact of Students' Skill Levels on the Effectiveness of Evidence-Based Interventions for Improving Mathematics Fluency

Principal Investigator: Erica Fanning
Graduate Student
CUNY, Graduate Center
365 Fifth Avenue
New York, NY 10016 USA
518-281-9688

Faculty Advisor: Yung-Chi Chen, Ph.D.
Visiting Assistant Professor
Office Number: 3204.04
CUNY, Graduate Center
365 Fifth Avenue
New York, NY 10016 USA
(212) 817-8288

Site where study is to be conducted: Library at Monument Elementary School, 66 Main Street
Bennington, VT 05201

Introduction/Purpose: Your child is invited to participate in a research study. The study is conducted by Erica Fanning, who is a doctoral student at the CUNY Graduate Center. The purpose of this research study is to better understand how brief mathematics interventions can be used for students of varying abilities and backgrounds to help improve fluent completion of single-digit multiplication problems. The study will also assess whether improvement in this specific area of mathematics is related to the improvement of student scores on more difficult tasks in mathematics. The results of this study may provide teachers with strategies for improving student performance in the area of mathematics.

Procedures: Approximately sixteen to twenty third and fourth grade students are expected to participate in this study. Each child will complete pre- and post-intervention
assessments. These assessments will take sixteen minutes to complete at the beginning and end of the intervention period. Initially, all students will complete 3, 2-minute assessments as a screening procedure. Students who have already mastered the target skill of single-digit multiplication according to specified criteria will be ineligible for participation in the study, as they are not in need of intervention in this area. All eligible participants will begin the initial phase in which they will complete 3, 2-minute assessments for 5 consecutive sessions in the absence of any intervention. Following this initial phase, all participants will be randomly assigned to one of two groups. Each group will receive a different part of a brief, multi-component, evidence-based intervention. Each part of the intervention will take two or less minutes to complete during each session. One group of students will be encouraged to beat their prior high score on the 2-minute assessment. The other group of students will act as peer tutors and tutees, as they will use flashcards to practice multiplication facts for two minutes. In the last intervention phase, all students will receive both of these intervention components as a combined intervention. The time commitment of each participant is expected to be less than fifteen minutes for each session. This includes transition and instruction. There will be thirty sessions for this study, and they are expected to take place over the course of 6-8 weeks. Each session will take place at Library at Monument Elementary School, which is located at 66 Main Street in Bennington, VT.

Possible Discomforts and Risks: Your child’s participation in this study may involve limited risk. Some students may become anxious when completing timed tasks or when encouraged to beat a prior score. To minimize these risks students will be assured that the assessments are not graded and they cannot fail. They will be provided with support throughout the intervention and encouraged to share any feelings of discomfort with a trusted adult or researcher. Additionally, students will be assured that their scores will not be shared with anyone other than the researchers. This includes parents and teachers. When scores are used for research purposes, they will be shared under code names and will not be linked to individual students in any way. If your child is bothered at all as a result of this study you should immediately contact his/her classroom teacher, the school principal, Mrs. Cauley, or the principal investigator, Erica Fanning. School personnel can be reached at 802-447-7979 and Erica Fanning can be reached by phone at 518-281-9688 or by email at ericafanning@gmail.com.

Benefits: Direct benefits of this study may include increased ability to complete basic multiplication problems. Research has shown that fluent responding to these types of problems is inherent for student success. Additionally, repeated practice and the use of the proposed interventions have been shown to increase fluency scores for students in the past. Indirect benefits may also be present as the general results of this study will help provide educators with information regarding the use of such practices in the school setting.

Voluntary Participation: Your child’s participation in this study is voluntary, and you may decide to withdraw your child from participation without prejudice, penalty, or loss of benefits to which you are otherwise entitled. If you decide to remove your child from
the study, please contact the principal investigator, Erica Fanning, to inform her of your decision.

**Financial Considerations:** Participation in this study will involve no cost to you or your child. For your child’s participation in this study he/she will receive small tokens of thanks following the completion of random intervention sessions. This will occur approximately once weekly. These tokens might include pencils, erasers, stickers, and/or small toys.

**Confidentiality:** The information obtained from your child will be collected in two ways. First, because we want to be sure that these interventions are equally effective for all students, we will asking for information regarding lunch status. By signing this form, you are giving consent for the school to share this information with the principal investigator. Additionally, students will regularly complete paper and pencil math assessments that will be collected, scored, and stored by the principal investigator or research assistants. All records will be stored under lock and key by the principal investigator or the research assistant. Students will only be asked to write their first names on these assessments. The collected records will be accessible only to the principal investigator and her advisor, Dr. Yung-Chi Chen, at the CUNY Graduate Center.

**Contact Questions/Persons:** If you or your child have any questions about the research now or in the future, you should contact the Principal Investigator, Erica Fanning, by phone at 518-281-9688 or via email at ericafanning@gmail.com, or her advisor, Dr. Yung-Chi Chen, at ychen8@gc.cuny.edu or 212-817-8288. If you have questions about your rights as a research participant, or you have comments or concerns that you would like to discuss with someone other than the researchers, please call the CUNY Research Compliance Administrator at 646-664-8918. Alternately, you can write to:

CUNY Office of the Vice Chancellor for Research  
Attn: Research Compliance Administrator  
205 East 42nd Street  
New York, NY 10017

**Statement of Consent:**

“I have read the above description of this research and I understand it. I have been informed of the risks and benefits involved, and all my questions have been answered to my satisfaction. Furthermore, I have been assured that any future questions that I may have will also be answered by the principal investigator of the research study. I voluntary agree to allow my child to participate in this study.

By signing this form I have not waived any of my legal rights to which my child would otherwise be entitled.

I will be given a copy of this statement.”
<table>
<thead>
<tr>
<th>Printed Name of Subject’s Legal Guardian</th>
<th>Signature of Subject’s Legal Guardian</th>
<th>Date Signed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Printed Name of the Child-subject</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Printed Name of Person Explaining Form</td>
<td>Signature of Person Explaining Form</td>
<td>Date Signed</td>
</tr>
<tr>
<td>Printed Name of Investigator</td>
<td>Signature of Investigator</td>
<td>Date Signed</td>
</tr>
</tbody>
</table>
ASSENT TO PARTICPATE IN A RESEARCH PROJECT

**Project Title:** Examination of the Impact of Students' Skill Levels on the Effectiveness of Evidence-Based Interventions for Improving Mathematics Fluency

**Principal Investigator:** Erica Fanning

**Faculty Advisor:** Dr. Yung-Chi Chen

---

**Child’s Name:**

You are invited to participate in Ms. Fanning’s research study. The reason for this study is to examine different ways in which we can help you learn multiplication facts.

**What will happen to me in this study?**

For the next 7 weeks we are going to meet 5 times a week in the library. While you are here, you will be asked to complete math work sheets that contain multiplication facts. On some days you will also be working with a partner to practice those math facts. Sometimes you will act as a tutor and show them flash cards, while other times they will tutor you with flashcards. The amount of time we spend together in the library will vary from about ten to fifteen minutes each day. At the start and end of the 7 weeks, you will also be asked to complete two math worksheets that contain other kinds of math problems. The first work sheet will ask you questions about things like patterns, measurement, and geometry, while the second worksheet will have some problems that might include decimals,
fractions, and other math facts. During our first and last sessions, you may be with us a little longer so that you can complete these worksheets.

**Will I get hurt?**

You may experience some stress because the math worksheets are timed and some students worry about completing everything before the time limit is up. To make you as comfortable as possible, we want to promise you that we will not tell your teachers what your scores are on these worksheets. Your scores will not impact your grade and you cannot fail these worksheets. All we ask is that you do your best. Also, sometimes, you may find the worksheets to be boring. Math isn’t always fun! So, if you are feeling stressed out or unhappy as a result of this study you should tell me, Ms. Parmenter, your parent/guardian, or someone else you know right away.

**Will anything good happen to me?**

The more time you spend practicing your math facts, the easier they should be. In school, you need to learn to complete your multiplication facts quickly. This study should help you do that. For your participation in this study you will receive small tokens of thanks following the completion of some intervention sessions. This will occur approximately once weekly. These tokens might include pencils, erasers, stickers, and/or small toys.

**What if I do not want to do this?**

You don’t have to be in this study. No one will be mad at you if you don’t want to do this. If you don’t want to be in this study, just tell us. If you want to be in this study, just tell us. Remember, it is ok to say yes now and change your mind later. Nothing will happen to you if you decide to stop.

**Will anyone know I was involved?**

Your name will be kept confidential. This means that we will not share your name with anyone. Other people may ask why you are going to the library, and you can tell them that you volunteered to spend extra time working on math. Or you can tell them that you agreed to participate in a research study. Only my research assistant, Ms. Parmenter, and I will know your scores on anything you do with us and we will keep your scores and individual progress secret. When we share information, we will not use your names, so no one will know anything about how you did during this study.
Who can I talk to about this study?
You can ask questions any time. You can ask now. You can ask later. You can talk to me or someone else, like your teacher, or Mrs. Cauley (the principal), or Ms. Parmenter.

Do you want to participate in this study? □ Yes □ No

PERSON CONDUCTING ASSENT

I have explained the study to ______________________________ (name of child) in language he/she understands, and he/she has agreed to be in the study.

_______________________________  ______________________________
Name of Person Conducting Assent (print)  Signature Person Conducting Assent
Date Signed

_______________________________  ______________________________
Name of Investigator (print)  Signature Person Investigator
Date Signed
Appendix C

Sample Single Skill Assessment Probe

Peer Tutoring Probe 29 CAB

<table>
<thead>
<tr>
<th>4</th>
<th>9</th>
<th>6</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>x3</td>
<td>x6</td>
<td>x6</td>
<td>x2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2</th>
<th>3</th>
<th>7</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>x5</td>
<td>x4</td>
<td>x4</td>
<td>x5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6</th>
<th>9</th>
<th>3</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>x6</td>
<td>x3</td>
<td>x4</td>
<td>x9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>7</th>
<th>5</th>
<th>5</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>x5</td>
<td>x5</td>
<td>x7</td>
<td>x3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>7</th>
<th>5</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>x8</td>
<td>x5</td>
<td>x2</td>
<td>x6</td>
</tr>
<tr>
<td>9 \times 3</td>
<td>6 \times 9</td>
<td>6 \times 8</td>
<td>7 \times 5</td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>7 \times 8</td>
<td>8 \times 6</td>
<td>2 \times 5</td>
<td>6 \times 8</td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>8 \times 7</td>
<td>4 \times 7</td>
<td>7 \times 4</td>
<td>6 \times 9</td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>8 \times 7</td>
<td>5 \times 5</td>
<td>4 \times 7</td>
<td>9 \times 6</td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>6 \times 6</td>
<td>8 \times 6</td>
<td>3 \times 9</td>
<td>5 \times 7</td>
</tr>
</tbody>
</table>
3

STOP
3 x5 | 6 x5 | 2 x7 | 6 x2 |

9 x8 | 9 x5 | 6 x2 | 7 x7 |

8 x9 | 7 x2 | 5 x6 | 3 x5 |

4 x9 | 3 x8 | 9 x5 | 2 x7
6

STOP
<table>
<thead>
<tr>
<th>4</th>
<th>8</th>
<th>7</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>x8</td>
<td>x4</td>
<td>x6</td>
<td>x7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5</th>
<th>4</th>
<th>9</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>x8</td>
<td>x4</td>
<td>x7</td>
<td>x4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6</th>
<th>9</th>
<th>7</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>x7</td>
<td>x2</td>
<td>x9</td>
<td>x8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6</th>
<th>9</th>
<th>5</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>x3</td>
<td>x9</td>
<td>x4</td>
<td>x8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>9</th>
<th>7</th>
<th>8</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>x9</td>
<td>x9</td>
<td>x4</td>
<td>x4</td>
</tr>
</tbody>
</table>
Appendix D

Teacher Acceptability Scales

D-1: Teacher Acceptability Scale – Performance Feedback

1. Teachers are likely to use the performance feedback intervention because it requires little or no technical skill.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

2. Teachers are likely to use the performance feedback intervention because it requires little training to implement effectively.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

3. Most teachers would find the use of the performance feedback intervention suitable for teaching math fluency in their classroom.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

4. Most teachers would find the use of the performance feedback intervention appropriate for teaching similar academic skills in their classroom (e.g., spelling, sight-word fluency).

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

5. I would suggest the use of the performance feedback intervention to other teachers.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>
6. Deficits in math fact fluency are severe enough to warrant the use of the performance feedback intervention in my classroom.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

7. Use of this intervention is appropriate for a variety of children.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

8. My use of the performance feedback intervention is consistent with other procedures I have used in classroom setting.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

9. The performance feedback intervention would be appropriate for use before making a referral.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

10. The performance feedback intervention would not be difficult to implement in a classroom with 20 other students.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

11. The performance feedback intervention is practical in the amount of time required for record keeping.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>
12. The performance feedback intervention is practical in the amount of out-of-school time required for implementation.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

13. It would not be difficult to use the performance feedback intervention and still meet the needs of all children in the classroom.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

14. I liked the procedures used in the performance feedback intervention.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

15. The use of the performance feedback intervention did prove effective in changing this child's math performance.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

16. The use of the performance feedback intervention was an acceptable process for this child's math performance.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

17. The use of the performance feedback intervention did not result in negative side effects for this child.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>
18. The use of the performance feedback intervention would not be considered a "last resort."

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Somewhat Disagree</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

19. Overall, the performance feedback intervention was beneficial for my students.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Somewhat Disagree</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

20. I would be willing to use the performance feedback intervention in the classroom setting.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Somewhat Disagree</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

Please provide any comments, questions, or feedback you may have with regards to the performance feedback intervention:

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
D-2: Teacher Acceptability Scale – Peer tutoring

1. Teachers are likely to use the peer tutoring intervention because it requires little or no technical skill.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

2. Teachers are likely to use the peer tutoring intervention because it requires little training to implement effectively.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

3. Most teachers would find the use of the peer tutoring intervention suitable for teaching math fluency in their classroom.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

4. Most teachers would find the use of the peer tutoring intervention appropriate for teaching similar academic skills in their classroom (e.g., spelling, sight-word fluency).

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

5. I would suggest the use of the peer tutoring intervention to other teachers.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

6. Deficits in math fact fluency are severe enough to warrant the use of the peer tutoring intervention in my classroom.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>
7. Use of this intervention is appropriate for a variety of children.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

8. My use of the peer tutoring intervention is consistent with other procedures I have used in classroom setting.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

9. The peer tutoring intervention would be appropriate for use before making a referral.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

10. The peer tutoring intervention would not be difficult to implement in a classroom with 20 other students.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

11. The peer tutoring intervention is practical in the amount of time required for record keeping.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

12. The peer tutoring intervention is practical in the amount of out-of-school time required for implementation.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>
13. It would not be difficult to use the peer tutoring intervention and still meet the needs of all children in the classroom.

14. I liked the procedures used in the peer tutoring intervention.

15. The use of the peer tutoring intervention did prove effective in changing this child's math performance.

16. The use of the peer tutoring intervention was an acceptable process for this child's math performance.

17. The use of the peer tutoring intervention did not result in negative side effects for this child.

18. The use of the peer tutoring intervention would not be considered a "last resort."

19. Overall, the peer tutoring intervention was beneficial for my students.
20. I would be willing to use the peer tutoring intervention in the classroom setting.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

Please provide any comments, questions, or feedback you may have with regards to the peer tutoring intervention:

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
D-3: Teacher Acceptability Scale – Combined

1. Teachers are likely to use the combined intervention because it requires little or no technical skill.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

2. Teachers are likely to use the combined intervention because it requires little training to implement effectively.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

3. Most teachers would find the use of the combined intervention suitable for teaching math fluency in their classroom.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

4. Most teachers would find the use of the combined intervention appropriate for teaching similar academic skills in their classroom (e.g., spelling, sight-word fluency).

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

5. I would suggest the use of the combined intervention to other teachers.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

6. Deficits in math fact fluency are severe enough to warrant the use of the combined intervention in my classroom.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>
7. Use of this intervention is appropriate for a variety of children.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

8. My use of the combined intervention is consistent with other procedures I have used in classroom setting.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

9. The combined intervention would be appropriate for use before making a referral.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

10. The combined intervention would not be difficult to implement in a classroom with 20 other students.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

11. The combined intervention is practical in the amount of time required for record keeping.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

12. The combined intervention is practical in the amount of out-of-school time required for implementation.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>
13. It would not be difficult to use the combined intervention and still meet the needs of all children in the classroom.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

14. I liked the procedures used in the combined intervention.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

15. The use of the combined intervention did prove effective in changing this child's math performance.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

16. The use of the combined intervention was an acceptable process for this child's math performance.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

17. The use of the combined intervention did not result in negative side effects for this child.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

18. The use of the combined intervention would not be considered a "last resort."

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

19. Overall, the combined intervention was beneficial for my students.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>
20. I would be willing to use the combined intervention in the classroom setting.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Disagree Somewhat</th>
<th>Agree Somewhat</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

Please provide any comments, questions, or feedback you may have with regards to the combined intervention:

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
D-4: Acceptability – Overview

1. The intervention that was **most effective** was: _______________________
   I believe this because:
   
   ___________________________________________________________________
   ___________________________________________________________________
   ___________________________________________________________________

2. The intervention I would be **most likely** use in my classroom is:
   ______________________
   I would be most likely to use this intervention because:
   
   ___________________________________________________________________
   ___________________________________________________________________
   ___________________________________________________________________

3. The intervention I found to be **least effective** was: __________________
   I believe this because:
   
   ___________________________________________________________________
   ___________________________________________________________________
   ___________________________________________________________________

4. The intervention I would be **least likely** use in my classroom is:
   ______________________
   I would be **least likely** to use this intervention because:
   
   ___________________________________________________________________
   ___________________________________________________________________
   ___________________________________________________________________

Please feel free to add any comments/suggestions/thoughts you have about these intervention procedures:

_____________________________________________________________________
_____________________________________________________________________
_____________________________________________________________________
_____________________________________________________________________
Appendix E

Student Acceptability Scales

E-1: KIP – Performance Feedback Intervention

Question #1
How much did you like practicing multiplication?

- Not at all
- A little bit
- Some
- A lot
- Very, very much

Question #2
How much do you like hearing what your best score is?

- Not at all
- A little bit
- Some
- A lot
- Very, very much
Question #3
Were there times when you didn’t want to practice math this way?

Not at all  A little bit  Some  A lot  Very, very much

Question #4
Were there any times when you wished you could work more your multiplication facts?

Not at all  A little bit  Some  A lot  Very, very much
Question #5
How much do you like being asked to beat your score each week?

- [ ] Not at all
- [ ] A little bit
- [ ] Some
- [ ] A lot
- [ ] Very, very much

Question #6
How much do you think it helps you when you were asked to beat your score?

- [ ] Not at all
- [ ] A little bit
- [ ] Some
- [ ] A lot
- [ ] Very, very much
**Question #7**
Do you think your multiplication skills have improved?

- Not at all
- A little bit
- Some
- A lot
- Very, very much

**Question #8**
Do you think your multiplication skills have gotten worse?

- Not at all
- A little bit
- Some
- A lot
- Very, very much
Question #1
How much did you like practicing multiplication?

[Diagram with options: Not at all, A little bit, Some, A lot, Very, very much]

Question #2
How much do you like being the tutor?

[Diagram with options: Not at all, A little bit, Some, A lot, Very, very much]
Question #3
How much do you like being the tutee?

Not at all  A little bit  Some  A lot  Very, very much

Question #4
How much do you like writing down the problems after you made a mistake?

Not at all  A little bit  Some  A lot  Very, very much
Question #5
How much do you practicing your math facts after the tutoring intervention?

Question #6
Were there times when you didn’t want to practice math this way?
Question #7
Were there any times when you wished you could work more your multiplication facts this way?

☐ Not at all  ☐ A little bit  ☐ Some  ☐ A lot  ☐ Very, very much

Question #8
How much do you like the peer tutoring intervention?

☐ Not at all  ☐ A little bit  ☐ Some  ☐ A lot  ☐ Very, very much
Question #9
How much do you think it helped to practice with peer tutoring?

Not at all  A little bit  Some  A lot  Very, very much

Question #10
Do you think your multiplication skills have improved?

Not at all  A little bit  Some  A lot  Very, very much
Question #11
Do you think your multiplication skills have gotten worse?

Not at all  A little bit  Some  A lot  Very, very much
E-3: KIP – Combined Intervention

**Question #1**
How much did you like practicing multiplication?

- Not at all
- A little bit
- Some
- A lot
- Very, very much

**Question #2**
How much do you like being the tutor?

- Not at all
- A little bit
- Some
- A lot
- Very, very much
Question #3
How much do you like being the tutee?

Not at all  A little bit  Some  A lot  Very, very much

Question #4
How much do you like writing down the problems after you made a mistake?

Not at all  A little bit  Some  A lot  Very, very much
Question #5
How much do you practicing your math facts after the tutoring intervention?

[Options]
- Not at all
- A little bit
- Some
- A lot
- Very, very much

Question #6
How much do you like trying to beat your best score each day?

[Options]
- Not at all
- A little bit
- Some
- A lot
- Very, very much
Question #7
Were there times when you didn’t want to practice math this way?

- Not at all
- A little bit
- Some
- A lot
- Very, very much

Question #8
Were there any times when you wished you could work more your multiplication facts this way?

- Not at all
- A little bit
- Some
- A lot
- Very, very much
Question #9
How much do you like the peer tutoring intervention?

Not at all  A little bit  Some  A lot  Very, very much

Question #10
How much do you think it helped to practice with peer tutoring?

Not at all  A little bit  Some  A lot  Very, very much
Question #11
Do you think your multiplication skills have improved?

- [ ] Not at all
- [ ] A little bit
- [ ] Some
- [ ] A lot
- [ ] Very, very much

Question #12
Do you think your multiplication skills have gotten worse?

- [ ] Not at all
- [ ] A little bit
- [ ] Some
- [ ] A lot
- [ ] Very, very much
Appendix F

**Research Assistant Training Procedures**

1) Fingerprinting and background check to ensure she is able to enter the school. (Required by the district.)

2) Undergraduate Citi Human Subjects Training Course will be completed by the research assistant.

3) A series of four 1 hour training sessions will take place, with reading and practice assignments for the RA to complete between each one:

   1) The first training session will focus the research questions and design. The RA will be given copies of the flow charts and procedures to review after our discussion. Interscorer reliability and treatment integrity will be discussed in general terms. During this session, pre/post-test measures will be discussed and shared for review.

   2) The second session will focus on the creation of the single skill progress monitoring probes. These directions are included at the end of this document. We will create one probe together, with me leading the way. She will then lead me through the creation of a second probe. The RA will also be provided with time to ask any clarifying questions at this time.

   3) The RA will then receive the instructions for scoring single skill probes. She will be asked to score 3 practice probes to ensure she understands the procedure. She will also receive formulas for calculating treatment integrity and interscorer agreement.

   4) We will meet a third time to go over scoring procedures and I will answer any questions she has. At this time, the RA will practice using the formulas.

   5) We will meet a fourth and final time to go over the intervention procedures. We will read through them together and ensure that we have all the materials we need to implement the intervention readily available.

4) At the start of the study, the RA will observe me implementing the baseline procedures for a minimum of 3 sessions. She will then be observed implementing these procedures on at least 2 occasions.

5) At the start of each intervention phase, the RA will observe me implementing the intervention at least twice. She will be observed at least once before carrying out any procedures on her own.

*Any time treatment integrity scores are lower than 100% additional training will be provided.*
*If the RA has any questions or concerns additional training will be given.

**Directions for Creating Single Skill Probes**

Step 1: Use template taken from www.interventioncentral.org

Step 2: Go to http://randomnumbergenerator.intemodino.com/en/

Step 3: Generate 40 random numbers from 1-40. Select the option to not repeat numbers.

Step 4: Look at the chart for the designated set of problems. Put the corresponding problem to the randomly generated number into the worksheet template.

- Do not use the same problem twice in a row. If the same problem is designated to go twice in a row, add the problem to the end of the worksheet.

Step 5: Save the worksheet as “Set _ Worksheet _”. For example, “Set A Worksheet 1”. There will be 36 probes per set generated

Set A:

<table>
<thead>
<tr>
<th>Problem 1</th>
<th>Problem 2</th>
<th>Problem 3</th>
<th>Problem 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 x 6 = 12</td>
<td>1, 21</td>
<td>6 x 2 = 12</td>
<td>11, 31</td>
</tr>
<tr>
<td>2 x 7 = 14</td>
<td>2, 22</td>
<td>7 x 2 = 14</td>
<td>12, 32</td>
</tr>
<tr>
<td>3 x 5 = 15</td>
<td>3, 23</td>
<td>5 x 3 = 15</td>
<td>13, 33</td>
</tr>
<tr>
<td>3 x 8 = 24</td>
<td>4, 24</td>
<td>3 x 8 = 24</td>
<td>14, 34</td>
</tr>
<tr>
<td>4 x 6 = 24</td>
<td>5, 25</td>
<td>4 x 6 = 24</td>
<td>15, 35</td>
</tr>
<tr>
<td>4 x 9 = 36</td>
<td>6, 26</td>
<td>4 x 9 = 36</td>
<td>16, 36</td>
</tr>
<tr>
<td>5 x 6 = 30</td>
<td>7, 27</td>
<td>6 x 5 = 30</td>
<td>17, 37</td>
</tr>
<tr>
<td>5 x 9 = 45</td>
<td>8, 28</td>
<td>9 x 5 = 45</td>
<td>18, 38</td>
</tr>
<tr>
<td>7 x 7 = 49</td>
<td>9, 29</td>
<td>7 x 7 = 49</td>
<td>19, 39</td>
</tr>
<tr>
<td>8 x 9 = 72</td>
<td>10, 30</td>
<td>9 x 8 = 72</td>
<td>20, 40</td>
</tr>
</tbody>
</table>
### Set B:

<table>
<thead>
<tr>
<th>Equation</th>
<th>Result</th>
<th>Equation</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 x 9 = 18</td>
<td>1, 21</td>
<td>9 x 2 = 18</td>
<td>11, 31</td>
</tr>
<tr>
<td>3 x 6 = 18</td>
<td>2, 22</td>
<td>6 x 3 = 18</td>
<td>12, 32</td>
</tr>
<tr>
<td>3 x 7 = 21</td>
<td>3, 23</td>
<td>7 x 3 = 21</td>
<td>13, 33</td>
</tr>
<tr>
<td>4 x 4 = 16</td>
<td>4, 24</td>
<td>4 x 4 = 16</td>
<td>14, 34</td>
</tr>
<tr>
<td>4 x 5 = 20</td>
<td>5, 25</td>
<td>5 x 4 = 20</td>
<td>15, 35</td>
</tr>
<tr>
<td>4 x 8 = 32</td>
<td>6, 26</td>
<td>8 x 4 = 32</td>
<td>16, 36</td>
</tr>
<tr>
<td>5 x 8 = 40</td>
<td>7, 27</td>
<td>5 x 8 = 40</td>
<td>17, 37</td>
</tr>
<tr>
<td>6 x 7 = 42</td>
<td>8, 28</td>
<td>7 x 6 = 42</td>
<td>18, 38</td>
</tr>
<tr>
<td>7 x 9 = 63</td>
<td>9, 29</td>
<td>9 x 7 = 63</td>
<td>19, 39</td>
</tr>
<tr>
<td>9 x 9 = 81</td>
<td>10, 30</td>
<td>9 x 9 = 81</td>
<td>20, 40</td>
</tr>
</tbody>
</table>

### Set C:

<table>
<thead>
<tr>
<th>Equation</th>
<th>Result</th>
<th>Equation</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 x 5 = 10</td>
<td>1, 21</td>
<td>5 x 2 = 10</td>
<td>11, 31</td>
</tr>
<tr>
<td>3 x 4 = 12</td>
<td>2, 22</td>
<td>4 x 3 = 12</td>
<td>12, 32</td>
</tr>
<tr>
<td>3 x 9 = 27</td>
<td>3, 23</td>
<td>9 x 3 = 27</td>
<td>13, 33</td>
</tr>
<tr>
<td>4 x 7 = 28</td>
<td>4, 24</td>
<td>7 x 4 = 28</td>
<td>14, 34</td>
</tr>
<tr>
<td>5 x 5 = 25</td>
<td>5, 25</td>
<td>5 x 5 = 25</td>
<td>15, 35</td>
</tr>
<tr>
<td>5 x 7 = 35</td>
<td>6, 26</td>
<td>7 x 5 = 35</td>
<td>16, 36</td>
</tr>
<tr>
<td>6 x 6 = 36</td>
<td>7, 27</td>
<td>6 x 6 = 36</td>
<td>17, 37</td>
</tr>
<tr>
<td>6 x 8 = 48</td>
<td>8, 28</td>
<td>8 x 6 = 48</td>
<td>18, 38</td>
</tr>
<tr>
<td>6 x 9 = 54</td>
<td>9, 29</td>
<td>9 x 6 = 54</td>
<td>19, 39</td>
</tr>
<tr>
<td>7 x 8 = 56</td>
<td>10, 30</td>
<td>8 x 7 = 56</td>
<td>20, 40</td>
</tr>
</tbody>
</table>
**Procedural Integrity Formula**

\[
\left( \frac{\text{# of steps checked by independent observer}}{\text{# of steps listed for procedure}} \right) \times 100 = \text{procedural integrity percentage}
\]

**Interscorer Agreement Formula**

\[
\left( \frac{\text{# of digits scored the same}}{\text{# of total digits scored}} \right) \times 100 = \text{interscorer agreement percentage}
\]
Appendix G
Procedures and Integrity Checklists

G-1: Baseline Procedures

Materials:

☐ Packet of Probes  ☐ Stopwatch  ☐ Pencils

Procedures:

☐ Individually distribute the materials to the students, with the front page facing down to each student.

Read over the following instructions to the students as the material is being distributed:
“Hi everyone. Please do not turn over the worksheets.”

Say to the students:

☐ “Is everybody ready to begin the worksheet?”
☐ “The sheets on your desk contain math facts. All of the problems are multiplication facts.”
☐ When I say ‘begin’, turn the worksheet over and begin answering the problems. Start on the first problem on the left on the top row [point]. Work across and then go to the next row. If you can’t answer a problem, make an ‘X’ on it and go to the next one. If you finish the page go on to the next page. Are there any questions?”

☐ Say, “Begin” and start the timer.
☐ After 2 minutes have passed say, “Stop. Please put down your pencils.”
☐ Say, “Now we are going to complete another worksheet. When I say ‘begin’, please turn the page and begin answering the problems. Start on the first problem on the left of the top row [point]. Work across and then go on to the next row. If you can’t answer a problem, make an ‘X’ on it and go on to the next one. If you finish the page go on to the next page. Are there any questions?”

☐ Say, “Begin” and start the timer.
☐ After 2 minutes have passed say, “Stop. Please put down your pencils.”
☐ Say, “Now we are going to complete one last worksheet in the same way. When I say begin, please turn the page and begin answering the problems. Are there any questions?”

☐ Say, “Begin” and start the timer.
☐ After 2 minutes have passed say, “Stop. Please put down your pencils.”
☐ Collect the packets.
☐ Praise students for cooperating.
G-2: Performance Feedback Procedures

Materials:
- Packet of Probes
- Stopwatch
- Pencils
- Post-it with previous high score.

Procedures:
- Individually distribute the materials to the students, with the front page facing down to each student. A post it containing the highest DCPM score for that student should be on the back of each packet.

Read over the following instructions to the students as the material is being distributed:

“Hi everyone. Please do not turn over the worksheets.”

Say to the students:
- “The number on the post it on the back of your worksheet is the highest number of digits correct per minute you have completed so far. Today I would like everyone to try to beat their score. Please remove this post it and put it on your desk where you can see it. Please do your best to work quickly and carefully.”
- “Is everybody ready to begin the worksheet?”
- “The sheets on your desk contain math facts. All of the problems are multiplication facts.”
- When I say 'begin', turn the worksheet over and begin answering the problems. Start on the first problem on the left on the top row [point]. Work across and then go to the next row. If you can't answer a problem, make an 'X' on it and go to the next one. If you finish the page go on to the next page. Are there any questions?"

Say, “Begin” and start the timer.
- After 2 minutes have passed say, “Stop. Please put down your pencils.”
- Say, “Now we are going to complete another worksheet. Please remember that we are trying to beat our best scores. When I say ‘begin’, please turn the page and begin answering the problems. Start on the first problem on the left of the top row [point]. Work across the page and then go on to the next row. If you can’t answer a problem, make an ‘X’ on it and go on to the next one. If you finish the page go on to the next page. Are there any questions?”

Say, “Begin” and start the timer.
- After 2 minutes have passed say, “Stop. Please put down your pencils.”
- Say, “Now we are going to complete one last worksheet in the same way. Remember that we are trying to beat our high score today. Please work as quickly and carefully as you can. When I say begin, please turn the page and begin answering the problems. Are there any questions? When I say begin, please turn the page and begin working.”

Say, “Begin” and start the timer.
☐ After 2 minutes have passed say, “Stop. Please put down your pencils.”
☐ Collect the packets.
☐ Praise students for cooperating.
G-3: Peer Tutoring Procedures

Materials:
- Packet of Probes
- Tutor and tutee flashcards for each student.
- Stopwatch
- Pencils
- Paper.
- Student directions for tutoring procedures.

Procedures:
- Individually distribute the materials to the students. Ensure that each student receives only the flashcards for which he/she will act as tutor for. Place flashcards and packets face down.

Read over the following instructions to the students as the material is being distributed:

“Hi everyone. Please do not turn over the materials.”

Peer Tutoring:
Say to the students:

- “Today we are going to practice multiplication. To do this, you are going to quiz each other using flashcards.”
- “Please raise your hand if you are tutoring first today.” [Ensure that one student from each dyad has raised his/her hand].
- “When I say ‘begin’ you will show the first card to your partner. If your partner answers correctly, say ‘correct’ and move on to the next card. If your partner answers incorrectly, say “incorrect, the answer is ______.”
- “Tutees, when you get the answer incorrect, please write the number sentence, including the answer on your paper.”
- “At the end of two minutes I would like all tutors to count the number of flash cards your partners have answered correctly. Then say, ‘Great job, you answered ___ math facts correctly today’. All tutees will then write that number on their papers.”
- “Does anybody have any questions?”
- “When I say ‘begin’ tutors pick up your flashcards and begin.”

Say, “Begin” and start the timer.

After 2 minutes have passed say, “Stop. Please put down your pencils and count the flashcards that were answered correctly. Remember to write that number on your paper.”

Assessment:
Say to the students:

- “Is everybody ready to begin the worksheet?”
- “The sheets on your desk contain the problems you just practiced.”
- “When I say 'begin', turn the worksheet over and begin answering the problems. Start on the first problem on the left on the top row [point]. Work across and then go to the next row. If you can't answer a problem, make an 'X' on it and go to the next one. If you finish the page go on to the next page. Are there any questions?”
Say, “Begin” and start the timer.

After 2 minutes have passed say, “Stop. Please put down your pencils and turn over your packets.”

Peer Tutoring:
Say to the students:

- “Now please raise your hand if you are going to be the tutor.” [Ensure that one student from each dyad has raised their hand.]
- “When I say ‘begin’ you will show the first card to your partner. If your partner answers correctly, say ‘correct” and move on to the next card. If your partner answers incorrectly, say “incorrect, the answer is ____.”
- “Tutees, when you get the answer incorrect, please write the number sentence, including the answer on your paper.”
- “At the end of two minutes I would like all tutors to count the number of flash cards your partners have answered correctly. Then say, ‘Great job, you answered ____ math facts correctly today’. All tutees will then write that number on their papers.”
- “Does anybody have any questions?”
- “When I say ‘begin’ tutors pick up your flashcards and begin.”

Say, “Begin” and start the timer.

After 2 minutes have passed say, “Stop. Please put down your pencils and count the flashcards that were answered correctly. Remember to write that number on your paper.”

Assessment:

Say, “Now we are going to complete another worksheet. This worksheet contains the problems we just practiced. Please put your packets back in front of you.” [Wait for students to have the packets ready.]

When I say ‘begin’, please turn the page and begin answering the problems. Start on the first problem on the left of the top row [point]. Work across the page and then go on to the next row. If you can’t answer a problem, make an ‘X’ on it and go on to the next one. If you finish the page go on to the next page. Are there any questions?”

Say, “Begin” and start the timer.

After 2 minutes have passed say, “Stop. Please put down your pencils.”

Say, “Now we are going to complete one last worksheet in the same way. This worksheet contains math facts that we have not practiced today. When I say begin, please turn the page and begin answering the problems. Are there any questions?”

Say, “Begin” and start the timer.

After 2 minutes have passed say, “Stop. Please put down your pencils.”

Collect the flashcards, packets, and papers.

Praise students for cooperating.
Materials:
- Packet of Probes
- Tutor and tutee flashcards for each student.
- Stopwatch
- Pencils
- Paper.

Procedures:
- Individually distribute the materials to the students. Ensure that each student receives only the flashcards for which he/she will act as tutor for. Place flash cards and packets face down.
- Read over the following instructions to the students as the material is being distributed: “Hi everyone. Please do not turn over the materials.”

Peer Tutoring:
Say to the students:

- “Today we are going to practice multiplication. To do this, you are going to quiz each other using flashcards.”
- “Please raise your hand if you are tutoring first today.” [Ensure that one student from each dyad has raised his/her hand].
- “When I say ‘begin’ you will show the first card to your partner. If your partner answers correctly, say ‘correct’ and move on to the next card. If your partner answers incorrectly, say “incorrect, the answer is ______.”
- “Tutees, when you get the answer incorrect, please to write the number sentence, including the answer on your paper.”
- “At the end of two minutes I would like all tutors to count the number of flash cards your partners have answered correctly. Then say, ‘Great job, you answered ____ math facts correctly today’.”
- “Does anybody have any questions?”
- “When I say ‘begin’ tutors pick up your flashcards and begin.”

Say, “Begin” and start the timer.
After 2 minutes have passed say, “Stop. Please put down your pencils and count the flashcards that were answered correctly. Remember to write that number on your paper.”

Assessment and Performance Feedback:
Say to the students:

- “Is everybody ready to begin the worksheet?”
- “The worksheets on your desk contain the problems you just practiced.”
- “The number on the post it on the back of your worksheet is the highest number of digits correct per minute you have completed so far. Today I would like everyone to try to beat their score. Please remove this post it and put it on your desk where you can see it. Please do your best to work quickly and carefully.”
- “When I say 'begin', turn the worksheet over and begin answering the problems. Start on the first problem on the left on the top row [point]. Work across and then go to the next row. If you can’t answer a problem,
make an 'X' on it and go to the next one. If you finish the page go on to the next page. Are there any questions?"

☐ Say, “Begin” and start the timer.
☐ After 2 minutes have passed say, “Stop. Please put down your pencils and turn over your packets.”

Peer Tutoring:
Say to the students:

☐ “Now please raise your hand if you are going to be the tutor.” [Ensure that one student from each dyad has raised their hand.]
☐ “When I say ‘begin’ you will show the first card to your partner. If your partner answers correctly, say ‘correct’ and move on to the next card. If your partner answers incorrectly, say “incorrect, the answer is ______.”
☐ “Tutees, when you get the answer incorrect, please to write the number sentence, including the answer on your paper.”
☐ “At the end of two minutes I would like all tutors to count the number of flash cards your partners have answered correctly. Then say, ‘Great job, you answered ____ math facts correctly today’.”
☐ “Does anybody have any questions?”
☐ “When I say ‘begin’ tutors pick up your flashcards and begin.”
☐ Say, “Begin” and start the timer.
☐ After 2 minutes have passed say, “Stop. Please put down your pencils and count the flashcards that were answered correctly. Remember to write that number on your paper.”

Assessment and Performance Feedback:

☐ Say, “Now we are going to complete another worksheet. This worksheet contains the problems we just practiced. Please put your packets back in front of you. Remember that we are trying to beat our high score today. Please work as quickly and carefully as you can.” [Wait for students to have the packets ready.]
☐ When I say ‘begin’, please turn the page and begin answering the problems. Start on the first problem on the left of the top row [point]. Work across the page and then go on to the next row. If you can’t answer a problem, make an ‘X’ on it and go on to the next one. If you finish the page go on to the next page. Are there any questions?”
☐ Say, “Begin” and start the timer.
☐ After 2 minutes have passed say, “Stop. Please put down your pencils.”
☐ Say, “Begin” and start the timer.
☐ After 2 minutes have passed say, “Stop. Please put down your pencils.”
☐ Collect the flashcards, packets, and papers.
☐ Praise students for cooperating.
G-5: Peer Tutor Student Procedures

Peer Tutoring:

☐ Tutor picks up flashcards and displays the first one when the experimenter says begin.
☐ Tutor says “correct” when the tutee responds correctly.
☐ Tutor says “incorrect, the answer is _____” when the tutee responds incorrectly.
☐ The tutee writes the number sentence three times following an incorrect response.
☐ The tutoring procedures stop when the experimenter says “stop.”
☐ The tutor counts the number of correctly answered flashcards and says, “Great job, you answered _____ math facts correctly today.”
☐ The tutee writes that number on his/her worksheet.

Assessment:

☐ Students begin assessment as directed.
☐ Students stop assessment as directed.

Peer Tutoring:

☐ Students swap roles when directed.
☐ Tutor picks up flashcards and displays the first one when the experimenter says begin.
☐ Tutor says “correct” when the tutee responds correctly.
☐ Tutor says “incorrect, the answer is _____” when the tutee responds incorrectly.
☐ The tutee writes the number sentence three times following an incorrect response.
☐ The tutoring procedures stop when the experimenter says “stop.”
☐ The tutor counts the number of correctly answered flashcards and says, “Great job, you answered _____ math facts correctly today.”
☐ The tutee writes that number on his/her worksheet.

Assessment:

☐ Students begin second assessment as directed.
☐ Students stop assessment as directed.
☐ Students begin third assessment as directed.
☐ Students stop third assessment as directed.

Completion:

☐ Students turn in flashcards as directed.
☐ Students turn in papers as directed.
☐ Student turn in CBM packets a directed.
G-6: Combined Student Procedures

Peer Tutoring:
- Tutor picks up flashcards and displays the first one when the experimenter says “begin”.
- Tutor says “correct” when the tutee responds correctly.
- Tutor says “incorrect, the answer is ______” when the tutee responds incorrectly.
- The tutee writes the number sentence three times following an incorrect response.
- The tutoring procedures stop when the experimenter says “stop.”
- The tutor counts the number of correctly answered flashcards and says, “Great job, you answered ____ math facts correctly today.”
- The tutee writes that number on his/her worksheet.

Assessment:
- Student remove post-it as directed.
- Students begin assessment as directed.
- Students stop assessment as directed.

Peer Tutoring:
- Students swap roles when directed.
- Tutor picks up flashcards and displays the first one when the experimenter says begin.
- Tutor says “correct” when the tutee responds correctly.
- Tutor says “incorrect, the answer is ______” when the tutee responds incorrectly.
- The tutee writes the number sentence three times following an incorrect response.
- The tutoring procedures stop when the experimenter says “stop.”
- The tutor counts the number of correctly answered flashcards and says, “Great job, you answered ____ math facts correctly today.”
- The tutee writes that number on his/her worksheet.

Assessment:
- Students begin second assessment as directed.
- Students stop assessment as directed.
- Students begin third assessment as directed.
- Students stop third assessment as directed.

Completion:
- Students turn in flashcards as directed.
- Students turn in papers as directed.
- Student turn in CBM packets a directed.


