THE EFFECTS OF FLUENCY TRAINING ON IMPLEMENTATION FIDELITY OF A
READING INTERVENTION CONDUCTED BY PARAPROFESSIONALS

by

Breda Victoria O’Keeffe

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Disability Disciplines

Approved:

Dr. Timothy A. Slocum
Major Professor

Dr. Sarah Bloom
Committee Member

Dr. Donna Gilbertson
Committee Member

Dr. Marilyn Likins
Committee Member

Dr. Charles Salzberg
Committee Member

Dr. Byron Burnham
Dean of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2009
ABSTRACT

The Effects of Fluency Training on Implementation Fidelity of a Reading Intervention Conducted by Paraprofessionals

by

Breda Victoria O’Keeffe, Doctor of Philosophy

Utah State University, 2009

Major Professor: Dr. Timothy A. Slocum
Department: Special Education and Rehabilitation

Improving educational outcomes involves many variables, including identifying effective interventions and ensuring that they are effectively implemented in schools. Within a “response to intervention” model, treatment integrity of academic interventions has become increasingly important. However, recent research has suggested that ensuring treatment integrity by instructional staff may require intensive coaching, including daily or weekly performance feedback. This system may be unsustainable in typical schools because of limited resources for supervision. Some studies have found that treatment integrity can be achieved with intense prior training that includes extensive practice followed by feedback in the training setting. Fluency-based instruction has the advantage of providing multiple practice opportunities in a relatively short amount of time. A fluency training package for paraprofessionals using the Corrective Reading: Decoding curriculum was evaluated in a multiple baseline design across individuals. The primary
dependent variables included paraprofessionals’ presentation rate and praise rate. Additional dependent variables included paraprofessionals’ accuracy in presenting error correction procedures, ratio of positive to negative comments, students’ on-task behavior, and word reading accuracy. Participants included five paraprofessionals delivering supplemental reading instruction to students in small groups, and one student from each of the paraprofessionals’ groups. We provided five hours of fluency training to paraprofessionals over five days in a group setting. Following fluency training we observed paraprofessionals during a maintenance phase. Paraprofessionals generally increased their presentation rates, praise rates, and percentage of accurate error correction steps with fluency training. Three paraprofessionals with variable positive-to-negative comments ratios decreased this variability during fluency training. We subsequently provided performance feedback if a paraprofessionals’ presentation rate or praise rate did not maintain at criterion levels. Four of the five paraprofessionals required performance feedback on at least one skill. Performance feedback had mixed effects on paraprofessionals’ skills. Most students maintained adequate word reading accuracy throughout the study, with no clear effects when interventions for paraprofessionals were introduced and withdrawn. Students’ on-task behavior was variable throughout the study, with decreases in variability for three students corresponding with fluency training for paraprofessionals.
ACKNOWLEDGMENTS

I want to thank my family, Bryan, Sofia, and Logan, for their kind patience, for keeping things fun, and for making everything an adventure. This is your work too. I appreciate the support we have received from our families. You all helped make this happen.

Dr. Tim Slocum has been an incredibly patient advisor, stretching my thinking at every turn. Thank you for your enthusiasm, expertise, humor, and being available during all but cycling and sleeping hours.

My committee members have been very helpful throughout this process. Thanks to Drs. Sarah Bloom, Donna Gilbertson, Marilyn Likins, and Charles Salzberg. Dr. Nancy Glomb also deserves thanks for her help with my proposal, and for regularly checking in on progress.

Thanks to Renee Magnusson for spending hours in front of our many screens, being flexible, dedicated, catching all the little things before they got big, and for your wonderful sense of humor. You made this project possible and enjoyable. Thanks.

I am very grateful to the paraprofessionals, students, literacy facilitators, and district coordinators who were generous with their time and energy.

I appreciate the support of everyone in the department. Co, Kris, Karl, Jozanne, and Tammy were always available and willing to help. My fellow doctoral students were happy to listen, help, encourage, brainstorm, babysit, drive, entertain, and share ideas. Ginger saved me on many occasions and has been a wonderful colleague. Alice, Alison,
Erin, Doug, Kristin, Joy, Lee, Mark, Nicky, Shawnee, Trey, Trina, and Tyra were helpful in so many ways.

Dr. Andrew Samaha helped me collect data electronically, and provided the computer programs used and his technical expertise.

The professionals in the Direct Instruction community were extremely prompt (even during the holidays!), generous with information and expertise, and supportive of my work. Molly Blakely, Doug Blancero, Anne Desjardins, Kurt Engelmann, Robert Harris, Suzy Fitch, Carol Nielsen, Tammy Pettigrew, and Carolyn Schneider provided critical information during this study.

Finally, I dedicate my graphs to Dr. Tom Higbee. I appreciate his fine instruction on single subject research and applied behavior analysis principles.

Breda Victoria O’Keeffe
# CONTENTS

<table>
<thead>
<tr>
<th>ABSTRACT</th>
<th>iii</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>x</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>I.  INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Response to Intervention</td>
<td>2</td>
</tr>
<tr>
<td>Intervention Fidelity</td>
<td>3</td>
</tr>
<tr>
<td>II. LITERATURE REVIEW</td>
<td>7</td>
</tr>
<tr>
<td>Training for Teachers in Reading Instruction</td>
<td>7</td>
</tr>
<tr>
<td>Training for Paraprofessionals in Reading Instruction</td>
<td>11</td>
</tr>
<tr>
<td>Training for Instructional Staff</td>
<td>12</td>
</tr>
<tr>
<td>Fluency Training</td>
<td>17</td>
</tr>
<tr>
<td>Direct Instruction</td>
<td>19</td>
</tr>
<tr>
<td>Direct Instruction Teaching Skills</td>
<td>22</td>
</tr>
<tr>
<td>Student Behaviors</td>
<td>25</td>
</tr>
<tr>
<td>Purpose Statement and Research Questions</td>
<td>26</td>
</tr>
<tr>
<td>III. METHOD</td>
<td>28</td>
</tr>
<tr>
<td>Setting and Context</td>
<td>28</td>
</tr>
<tr>
<td>Participants</td>
<td>31</td>
</tr>
<tr>
<td>Dependent Variables</td>
<td>36</td>
</tr>
<tr>
<td>Independent Variables</td>
<td>47</td>
</tr>
<tr>
<td>Design</td>
<td>52</td>
</tr>
<tr>
<td>Interobserver Agreement</td>
<td>54</td>
</tr>
<tr>
<td>Treatment Fidelity</td>
<td>55</td>
</tr>
</tbody>
</table>
IV. RESULTS .................................................................56
   Paraprofessionals’ Behavior ..............................................57
   Students’ Behavior .......................................................78
   Social Validity ..................................................................85

V. DISCUSSION ............................................................89
   Dependent Variables ........................................................89
   Social Validity ..................................................................98
   Independent Variables ....................................................101
   Limitations and Future Directions ..................................105
   Implications and Conclusion .......................................113

REFERENCES ...............................................................................114

APPENDICES .............................................................................135

Appendix A: Teacher Training in Reading Instruction, Review Details.............136
Appendix B: Paraprofessional Demographic Questionnaire .........................139
Appendix C: Direct Instruction Reading Instructional Rating Forms ...............144
Appendix D: Questionnaire for District Reading Coordinators .....................147
Appendix E: Questionnaire for Paraprofessionals Regarding Training ...........150
Appendix F: Outline of Fluency Training Sessions .....................................153
Appendix G: Verbal/Graphic Performance Feedback Protocol ......................164
Appendix H: Written Performance Feedback Form ......................................166
Appendix I: Fluency Training Treatment Fidelity Forms .................................168
Appendix J: Verbal/Graphic Performance Feedback Treatment Fidelity Form ..168
Appendix K: Interobserver Agreement: Computer Data Collection ...............174
Appendix L: Interobserver Agreement: Kappa ..............................................178
Appendix M: Data Analysis: Percent Nonoverlapping Data .........................181

CURRICULUM VITAE ......................................................................183
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Summary of Studies on Inservice Interventions for Reading Teachers</td>
</tr>
<tr>
<td>2</td>
<td>Demographic Information for Students</td>
</tr>
<tr>
<td>3</td>
<td>Teaching Behavior Criteria for Classroom Setting and Fluency Goals for Training Setting</td>
</tr>
<tr>
<td>4</td>
<td>Interobserver Agreement for Dependent Measures</td>
</tr>
<tr>
<td>5</td>
<td>Paraprofessionals’ Accurate Presentation Rate: Phase Means, Ranges, and Slopes</td>
</tr>
<tr>
<td>6</td>
<td>Paraprofessionals’ Praise Rate: Phase Means, Ranges, and Slopes</td>
</tr>
<tr>
<td>7</td>
<td>Paraprofessionals’ Percent of Error Correction Attempts Made for All Student Errors</td>
</tr>
<tr>
<td>8</td>
<td>Paraprofessionals’ Error Correction Accuracy: Phase Means, Ranges, and Slopes</td>
</tr>
<tr>
<td>9</td>
<td>Paraprofessionals Positive to Negative Comments: Phase Means, Ranges, and Slopes</td>
</tr>
<tr>
<td>10</td>
<td>Students’ On-task Behavior: Percentage of 10-s Intervals</td>
</tr>
<tr>
<td>11</td>
<td>Students’ Word Reading Accuracy: Percent First Time Correct</td>
</tr>
<tr>
<td>12</td>
<td>Social Validity of Paraprofessionals’ Presentation Rate, Praise Rate, and Overall Session Quality</td>
</tr>
<tr>
<td>13</td>
<td>Paraprofessionals’ Ratings of Fluency Training</td>
</tr>
<tr>
<td>14</td>
<td>Fluency Training Outline</td>
</tr>
<tr>
<td>15</td>
<td>Interobserver Agreement on Dependent Measures: Kappa</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Paraprofessionals’ accurate presentations per minute during word attack</td>
</tr>
<tr>
<td>2</td>
<td>Paraprofessionals’ praise statements per minute</td>
</tr>
<tr>
<td>3</td>
<td>Paraprofessionals’ error correction accuracy</td>
</tr>
<tr>
<td>4</td>
<td>Paraprofessionals’ positive to negative comments ratio</td>
</tr>
<tr>
<td>5</td>
<td>Students’ on-task behavior (percentage of 10-s intervals) during word attack</td>
</tr>
<tr>
<td>6</td>
<td>Students’ first time correct responses during word reading</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

Reading is a critical skill for children and adults. Students who perform poorly in reading early in school tend to perform poorly in reading in later grades (Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998; Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Juel, 1988; Stevenson & Newman, 1986; Torgesen, 1997). In addition, students who have difficulty with reading early tend to have difficulty with other subject areas that rely heavily on reading later in school (Cunningham & Stanovich, 1997; Stanovich, 1986). Older students with low reading skills tend to have lower self-esteem (Hearing on Literacy, 1997), drop out of school at higher rates, and are more likely to be unemployed than students who are proficient in reading (Snyder, Tan, & Hoffman, 2006). Adults with low reading proficiency tend to have poorer health and are much more likely to live in poverty than adults who read well (Rudd, Kirsch, & Yamamoto, 2004; Sum, Kirsch, & Yamamoto, 2004). These outcomes suggest that students who are at risk for reading failure early may be at risk for difficulties throughout school and into adulthood.

In recent history, students who were at risk for reading difficulties in first grade were taught in general education settings until they improved, or more often, until their reading deficits became very serious. When their reading level was 1.5 to 2 standard deviations below the mean, they might qualify for special education services with a learning disability in reading (Gersten & Dimino, 2006). Often, this discrepancy takes about two years to develop, so students were not given extra help until third grade or
later. Research has shown that without additional help, these students are not likely to catch up to their peers in reading, and the discrepancy is likely to persist into later grades (e.g., Francis et al., 1996). However, other studies have shown that early intervention can improve outcomes for students who are at risk for failure and reduce the chance that they might need special education services (e.g., Felton, 1993; Foorman et al., 1998; Torgesen, 1997). “Response to intervention” (RTI) is a system for providing early intervention for students at risk for failure, and for timely identification of students with learning disabilities based on how they respond to well-delivered research-based interventions.

**Response to Intervention**

Response to intervention is a multiple-tier comprehensive system for improving schoolwide learning by screening all students in critical academic skills, identifying and intervening in a timely manner with students at-risk for academic difficulties, and ultimately identifying students who may need more intensive intervention in special education (Fuchs & Fuchs, 1998; Fuchs, Mock, Morgan, & Young, 2003; Vaughn & Fuchs, 2003). The RTI system integrates intervention delivery and dynamic assessment for special education eligibility, in contrast to systems in which intervention and assessment are separate processes (VanDerHeyden, Witt, & Barnett, 2005). Although the details are operationalized differently across different RTI models, the general approach can be described in broad terms as follows (Fuchs et al., 2003):

1. Students are provided with “generally effective” instruction by their classroom teacher;
2. Their progress is monitored;
3. Those who do not respond [i.e., show adequate academic growth] get something else, or something more, from their teacher or someone else;
4. Again, their progress is monitored; and
5. Those who still do not respond either qualify for special education or for special education evaluation. (p. 159)

RTI typically includes multiple “tiers” of increasingly intensive and/or individualized instruction, ranging from general education classroom instruction, to small-group, and/or individualized instruction. Research-based practices are to be used at each level (IDEIA, 2004; Jimerson, Burns, & VanDerHeyden, 2007). Within an RTI system, learning disabilities are typically identified based on a student’s lack of adequate response to the interventions presented in earlier tiers, rather than using IQ tests and norm-referenced academic tests. Researchers have recently emphasized the need to measure intervention fidelity to confirm that research-based practices are implemented as intended at each level of the RTI system (Gansle & Noell, 2007; Noell & Gansle, 2006; VanDerHeyden et al., 2007). Assuring intervention fidelity is important for the goals of using RTI to improve outcomes for all students, efficiently identifying students who need additional services in special education, and optimizing the allocation of resources across these efforts.

**Intervention Fidelity**

Intervention fidelity (Dumas, Lynch, Laughlin, Smith, & Prinz, 2001; O’Donnell, 2008), also known as treatment fidelity (Moncher & Prinz, 1991) and treatment integrity (Gresham, 1989; Yeaton & Sechrest, 1981), is a measure of the extent to which the implementation of an intervention corresponds to the operational definition of that
intervention (Noell, 2008). Researchers have identified the importance of demonstrating high levels of intervention fidelity to increase the internal validity of experimental studies (Gersten et al., 2005; Horner et al., 2005; Moncher & Prinz; O’Donnell, 2008; Perepletchikova & Kazdin, 2005). In addition, authors have described the importance of ensuring high levels of intervention fidelity of practices implemented in schools and applied settings in order to improve outcomes for students (Fiske, 2008; Lane, Bocian, MacMillan, & Gresham, 2004; Noell, 2008; Vollmer, Sloman, & St. Peter Pipkin, 2008; Yeaton & Sechrest, 1981). Low intervention fidelity typically correlates with poorer student outcomes (Carlson & Francis, 2002; DiGennaro, Martens, & Kleinmann, 2007; DiGennaro, Martens & McIntyre, 2005; Gilbertson, Witt, Singletary, & VanDerHeyden, 2007; Greenwood, Terry, Arreaga-Mayer, & Finney, 1992; Matheson & Shriver, 2005; Noell et al., 2000; Noell et al., 2005; Sterling-Turner, Watson, & Moore, 2002; Witt, Noell, LaFleur, & Mortenson, 1997). Also, studies in which levels of fidelity are manipulated typically show that conditions with low intervention fidelity are functionally related to poorer student outcomes or less efficient learning compared to conditions with high intervention fidelity (Holcombe, Wolery, & Snyder, 1994; Noell, Gresham, & Gansle, 2002; Vollmer, Roane, Ringdahl, & Marcus, 1999; Wilder, Atwell, & Wine, 2006).

Adequate fidelity of intervention is particularly critical to the RTI model (Gansle & Noell, 2007; Noell & Gansle, 2006; VanDerHeyden et al., 2005). If practitioners do not assure adequate intervention fidelity, the validity of an RTI implementation as an alternative to previous methods for intervening with at-risk students may be undermined. Since lower intervention fidelity typically results in reduced efficacy of an intervention
for students, RTI interventions with low fidelity may be as ineffective as previous attempts to intervene with at risk students. In the identification of students with learning disabilities, due-process protections for students may be violated and measurement validity may be in question if intervention fidelity is low or not assessed. The RTI approach uses a student’s failure to benefit from generally effective instruction to infer that the student has a disability. This inference, however, requires evidence that the student experienced such instruction.

Another goal of using RTI is to optimize the allocation of resources for improving student outcomes (VanDerHeyden et al., 2005). Interventions that are not well implemented may represent a waste of resources. For example, a student may be determined to be unresponsive to a particular level of intervention and provided additional costly services. If the fidelity of the intervention was actually inadequate rather than the student’s response to intervention, resources are wasted in providing additional services. These resources might be better utilized to improve intervention fidelity. If these intervention fidelity failures occur on a regular basis, RTI may not live up to its promise and may be abandoned as an ineffective system for timely intervention and special education eligibility assessment (Gansle & Noell, 2007; Noell & Gansle, 2006). Ensuring high intervention fidelity appears to be critical to the goals of the RTI model; however, achieving this consistently within the constraints of a typical school setting remains a significant challenge (Noell, 2008; VanDerHeyden et al., 2005).

Ideally, RTI is a system that improves the provision of educational interventions to at risk students and provides for efficient identification of students who need special education services. Since researchers have found that variations in the quality of
intervention fidelity typically correspond with variations in outcomes for students, the success of an RTI implementation likely depends in part on the fidelity of intervention. Therefore, a review of the literature to determine what is known about promoting intervention fidelity and conducting additional research in this area would be worthwhile.
CHAPTER II
LITERATURE REVIEW

Intervention fidelity in schools and clinics has been addressed through a variety of approaches, including didactic training, intensive training prior to implementation, and performance feedback based on observations of implementation. Many interventions use a combination of these strategies.

Training for Teachers in Reading Instruction

Improving intervention fidelity for reading interventions within an RTI model may be informed by the literature on training teachers in reading. The National Reading Panel (NRP) reviewed the research literature on inservice training for teachers in reading instruction (National Institute of Child Health and Human Development [NICHD], 2000). Generally, they found inservice training for teachers to be effective for improving students’ reading outcomes. The authors concluded, “The set of results for these studies shows overwhelmingly that interventions in teacher education and professional development are successful. That is, teachers can learn to improve their teaching in ways that have direct effects on their students” (NICHD, pp. 5-13). However, they also observed that the literature lacked a systematic progression over time, and no single method was studied extensively. “An eclectic mix of methods was found that ranged from macro to micro in their focus” (NICHD, pp. 5-13). More recent research on teacher training in reading instruction (from 1999-2007) supports the same conclusions (see Appendix A for review procedures). A systematic search identified nine studies (see
### Table 1

**Summary of Studies on Inservice Interventions for Reading Teachers**

<table>
<thead>
<tr>
<th>Study</th>
<th>N (teachers; students)</th>
<th>Intervention Content</th>
<th>Components of Teacher Intervention</th>
<th>Components of Student Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baker &amp; Smith (1999)</td>
<td>3; 100</td>
<td>Phonemic awareness and alphabetic principle</td>
<td>Rationale, modeling, discussion (unclear)</td>
<td>PA, Phonics, Explicit, Systematic</td>
</tr>
<tr>
<td>Bos et al. (1999)</td>
<td>28; NR</td>
<td>Reading Instructional Methods of Efficacy (RIME)</td>
<td>Rationale, modeling, practice in training, setting, classroom observations, discussion</td>
<td>PA, Phonics, Fluency, Explicit, Systematic</td>
</tr>
<tr>
<td>Carreker et al. (2005)</td>
<td>NR; 518</td>
<td>Language Enrichment</td>
<td>Rationale, practice (unclear)</td>
<td>PA, Phonics, Fluency, Vocab., Comp., Explicit, Systematic</td>
</tr>
<tr>
<td>Foorman &amp; Moats (2004)</td>
<td>80; 1,400</td>
<td>NRP areas of reading instruction</td>
<td>Rationale, modeling, classroom observations (unclear)</td>
<td>PA, Phonics, Vocabulary, Comprehension, Explicit, Systematic</td>
</tr>
<tr>
<td>Jacob &amp; Lefgren (2004)</td>
<td>NR; 100,288</td>
<td>Not specified</td>
<td>Variety of professional development approaches</td>
<td>None noted</td>
</tr>
<tr>
<td>Study</td>
<td>N (teachers; students)</td>
<td>Intervention Content</td>
<td>Components of Teacher Intervention</td>
<td>Components of Student Intervention</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------------------</td>
<td>----------------------</td>
<td>-----------------------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>McCutchen et al. (2002)</td>
<td>44; 779</td>
<td>Increasing knowledge about phonemic awareness (primarily) and phonics</td>
<td>Rationale, discussion, trainee modeling of lessons (no feedback)</td>
<td>PA, Phonics</td>
</tr>
<tr>
<td>McGill-Franzen et al. (1999)</td>
<td>18; 377</td>
<td>Books in class, or books + training</td>
<td>Rationale</td>
<td>None noted</td>
</tr>
<tr>
<td>O’Connor (1999)</td>
<td>10;154</td>
<td>Professional development in <em>Ladders to Literacy</em></td>
<td>Modeling, practice, discussion meetings</td>
<td>PA</td>
</tr>
<tr>
<td>Taylor et al. (2005)</td>
<td>92; 733</td>
<td>Effective reading instruction (NRP-based)/ school reform (CIERA)</td>
<td>Rationale, modeling, discussion meetings</td>
<td>Unclear (“Balanced literacy”); contents varied across teachers and schools</td>
</tr>
</tbody>
</table>


Table 1), and more than nine different approaches to teacher training were presented in these studies (Baker & Smith, 1999; Bos, Mather, Narr, & Babur, 1999; Carreker et al., 2005; Foorman & Moats, 2004; Jacob & Lefgren, 2004; McCutchen et al., 2002; McGill-

The features of these studies reveal a variety of approaches to teacher training in reading. Most of the studies lacked operational descriptions of their teacher training procedures. None of the researchers evaluated the fidelity with which the teacher training was implemented. Many of the studies suggested that they provided the rationale for teaching students to read in a particular way. Some studies noted that teachers practiced the strategies, but the nature of this practice was not described (e.g., whether the teachers received feedback, how long they practiced, etc.). In many studies, participants met individually with researchers or colleagues during implementation, but the content of these meetings was not described. For example, Foorman and Moats (2004) noted, “During the fourth year of the project, reading coaches worked intensively with individual teacher in their classrooms” (p. 55), but did not describe these meetings further. Previously, the NRP had found that researchers presented the details on the nature of the student interventions when describing the contents of teacher training, rather than describing teacher training procedures directly. The studies reviewed here had a similar emphasis on student interventions. None of the studies compared different approaches to teacher training, or built on previous research in teacher training to evaluate a particular method of training. Adding studies that build on previous research in training for instructional staff in a more systematic way might improve our understanding of how to increase intervention fidelity of reading instruction.
Training for Paraprofessionals in Reading Instruction

Although training procedures that are effective with teachers and their students are likely to be effective for paraprofessionals, this has not been evaluated in the literature. Since paraprofessionals and teachers often have different educational attainment, a particular technique may be differentially effective for the two groups. Thus, we reviewed the literature on training for paraprofessionals in reading.

Numerous studies have found positive effects for at risk students in reading when paraprofessionals delivered the interventions (Gunn, Smolkowski, Biglan, & Black, 2002; Lane, Fletcher, Carter, Dejud, & De Lorenzo, 2007; Miller, 2003; Vadasy, Jenkins, & Pool, 2000; Vadasy & Sanders, 2008; Vadasy, Sanders, & Abbott, 2008; Vadasy, Sanders, & Peyton, 2006a; Vadasy, Sanders, & Peyton, 2006b; Vadasy, Sanders, & Tudor, 2007). In these studies, paraprofessionals typically provided supplemental reading instruction (i.e., in addition to the classroom teacher’s daily reading lesson) to individuals or small groups of students. Many of these studies provided better descriptions of training for paraprofessionals than found in the teacher training literature and included evaluations of intervention fidelity. In most of the studies, researchers provided didactic training to paraprofessionals and practice with feedback prior to implementation. This training was typically followed by weekly observations and feedback on implementation fidelity. However, similar to the literature on training teachers in reading interventions, the focus of these studies was on the efficacy of the intervention for students, while paraprofessional training was an incidental component of the study. None of the studies evaluated or compared components of paraprofessional training or systematically explored issues surrounding paraprofessional training. While it is important to show that
research-based interventions can be conducted by paraprofessionals with good fidelity and positive student outcomes, the researchers provided more training and supervision of the paraprofessionals than would be available in most schools – thus the positive outcomes obtained in the research may not reflect likely outcomes in non-research application. It is important to identify specific efficient professional development practices that can produce high fidelity of implementation and positive student outcomes.

**Training for Instructional Staff**

Since the literature on teacher and paraprofessional training in reading does not specifically and systematically address the particulars of staff training, the search for effective and efficient methods for promoting intervention fidelity must be broadened to include research on training instructional staff in general. Researchers have found that achieving high levels of intervention fidelity in applied settings is not easy. Several literature reviews on diverse topics related to training of instructional staff have found that didactic training (e.g., training outside the classroom, and typically not including practice of skills) is the least likely staff development model to result in generalization of teaching skills to the classroom or clinic setting (Joyce & Showers, 2002; Noell, 2008; Rose & Church, 1998; Scheeler, 2008). Joyce and Showers noted, “…the gradual addition of the informational, demonstration, and practice training elements does not appear to noticeably affect transfer (effect size of 0.00 for information or theory; theory plus demonstration; and theory demonstration, and feedback)” (p. 77). Researchers have found that practice of the skills outside the classroom may result in some acquisition of the skill, but trainees typically do not readily apply what has been learned to the classroom setting. These reviews identified frequent (daily or weekly) performance
feedback or peer consultation as the most effective means for increasing generalization of instructional skills to the classroom setting.

**Performance Feedback**

Performance feedback on classroom implementation of interventions has been identified as the most studied staff training method in the literature (Noell, 2008; Rose & Church, 1998). Research has suggested that providing feedback to staff based on classroom performance may be more effective than consultation only (Noell, Witt, Gilbertson, Ranier, & Freeland, 1997), didactic training (Gilbertson et al., 2007; Moore, et al., 2002; Stormont, Smith, & Lewis, 2007), commitment emphasis training (Noell et al., 2005), and practice and training prior to intervention (DiGennaro et al., 2005; Matheson & Shriver, 2005; Sterling-Turner et al., 2002). Most of the research on performance feedback has been conducted in the context of behavioral consultation on behavior intervention plans. In this setting, the consultant meets with the individual who implements the behavior intervention plan and discusses the plan prior to implementation, and sometimes models and provides practice for the individual. The studies in this area typically compared performance feedback to discussion only (e.g., Jones, Wickstrom, & Friman, 1997; Noell et al., 1997), one day of feedback in the classroom (e.g., Noell, Duhon, Gatti & Connell, 2002; Witt et al., 1997), or limited modeling and practice prior to implementation (e.g., Mortenson et al., 1998). Many of these studies documented adequate classroom performance immediately after initial training, but the fidelity of implementation subsequently declined, necessitating performance feedback. Some studies included training to a criterion prior to implementation (e.g., Gilbertson et al., 2007; DiGennaro et al., 2005, 2007), but also have
found integrity decrements during maintenance phases. The initial training in these studies seemed to be relatively low intensity (e.g., a consultant spending some time one day prompting the teacher to include steps missed). If a more robust training were used prior to implementation, the need for performance feedback may be reduced.

Unlike other methods for increasing intervention integrity, a systematic body of literature on performance feedback exists. This research suggests effective ways to provide performance feedback to promote intervention integrity. Noell and colleagues (1997) found that spoken and graphic (data displayed on a graph) performance feedback was more effective than consultation alone (meetings outside the classroom setting). Reinke, Lewis-Palmer, and Martin (2007) found that graphic performance feedback without spoken feedback was more effective than general spoken feedback during group meetings on acquisition of praising skills. However, graphic feedback alone appeared to lose its efficacy after initial increases in performance. Noell, Duhon, and colleagues (2002), and DiGennaro and colleagues (2007) found that spoken and graphic performance feedback on student behavior was less effective than spoken and graphic performance feedback on teacher behavior. Hagermoser-Sanetti, Luiselli, and Handler (2007) found that spoken and graphic feedback on teacher behavior was much more effective at maintaining treatment fidelity on a behavior support plan than spoken feedback alone. These studies suggest that graphic and spoken performance feedback based on classroom observations of teacher behavior may be an effective way to maintain implementation fidelity of interventions in classroom settings.

Despite this large body of research supporting performance feedback and suggesting specific procedures that may be most effective, the practical question of how
schools and districts should organize professional development is still perplexing. Performance feedback is costly because it requires observation and data collection. A supervisor must observe each teacher’s or paraprofessional’s instruction individually, collect data, assemble the data to present to the teacher or paraprofessional, and arrange a time to provide feedback. Paraprofessionals often are scheduled with students during all working hours. They do not usually have any prep time, so finding a time to provide feedback is not a trivial consideration. In addition, schools with numerous instructional staff members may not have the resources necessary to provide daily or even weekly observations and performance feedback. A supervisor responsible for twenty or thirty staff members may not have the time necessary to provide weekly performance feedback in addition to their other responsibilities.

Individual performance feedback based on classroom performance amounts to a type of corrective procedure; it may be a less efficient means of enhancing staff skill than implementing effective training *prior* to implementing an intervention in the classroom. If multiple staff members can be trained outside the classroom at the same time, efficiency is increased further. An efficient model of professional development for instructional staff may include an effective training prior to implementation for all individuals, followed by targeted performance feedback for those individuals who still have difficulty with implementation. Ideally, performance feedback would not be required by all individuals after the initial training if the training is effective. This model might be more efficient and feasible than performance feedback for all individuals to establish and maintain adequate intervention fidelity.
Intensive Training Prior to Implementation

Several recent studies have demonstrated that training outside the classroom with sufficient intensity and practice prior to implementing an intervention or assessment may result in adequate acquisition (Iwata et al., 2000; Roscoe & Fisher, 2008) and maintenance of these skills (Lerman, Tetreault, Hovanetz, Strobel, & Garro, 2008; Lerman, Vorndran, Addison, & Kuhn, 2004; Moore & Fisher, 2007; Slider, Noell, & Williams, 2006). In some of these studies, the need for individual feedback after initial training was minimal. Roscoe and Fisher provided training on a relatively simple skill (conducting preference assessments), so the acquisition of this skill in a brief training format outside the classroom is not surprising. However, in the other studies, participants learned a variety of relatively complex skills that required differential responding to student interactions in different conditions (e.g., functional analysis procedures; time out), and applied different instructional or behavior management skills (e.g., multiple discrete trial training procedures; providing praise appropriately). In two of these studies (Iwata et al., 2000; Roscoe & Fisher, 2008), the skills were adequately acquired in the training setting, but no assessments of the generalization of the skills to the applied settings were conducted. In the other studies, participants in relatively brief, intensive interventions in training settings acquired the skills during training and generalized these skills to the applied setting. Maintenance of these skills up to six months after training was demonstrated for most of the participants in Lerman and colleague’s (2008) study.

While these studies evaluated different training packages and various target skills, they all included features that would likely enhance generalization from the training to the classroom or clinical setting. For example, Lerman et al. (2008) included common
stimuli in the training setting and classroom setting, and provided numerous practice
opportunities with performance feedback in the training setting. Moore and Fisher
evaluated the effectiveness of several versions of video modeling and didactic training
(reading a manual) on conducting a functional analysis. They found that video modeling
that included a more complete set of exemplars was more effective in helping the
participants acquire the skills than less complete video models and didactic training.
Generalization of these skills to a clinical setting was demonstrated for all participants.
This set of findings is in contrast with the studies on performance feedback cited above
which found that training outside the classroom or clinic typically resulted in lack of
generalization or maintenance of performance in the classroom. The studies that include
intensive training suggest that with attention to features that may enhance generalization,
training outside the classroom may result in generalization and maintenance of skills
without daily or weekly performance feedback for most participants.

**Fluency Training**

Stokes and Baer (1977) recommended many strategies for promoting
generalization of skills to novel settings and maintenance of those skills over time. One
of these strategies was to introduce the skills to natural contingencies of reinforcement.
One facet of this strategy is to ensure that the individual is adequately proficient in the
skill so that the skill is reinforced in the natural setting (White et al., 1988). For example,
if a behavior is performed too slowly or not often enough, it may not produce
reinforcement in the natural setting, resulting in poor maintenance. One way to increase
the likelihood that a behavior will generalize to a new setting and maintain over time is to ensure that it can be performed fluently (i.e., automatically).

While mastery of skills (i.e., training until a high percentage of correct performance is achieved) has been shown to be more effective than training that does not monitor or ensure mastery, a high percentage of correct responses does not guarantee that those skills will be applied to new settings or maintained over time (e.g., DiGennaro et al., 2005; Gilbertson et al., 2007; Hagermoser-Sanetti et al., 2007). Research conducted with children and adults suggests that additional practice beyond high accuracy is correlated with improved academic outcomes for students (Brophy & Good, 1986), and typically enhances application of skills to new settings (Bucklin, Dickinson, & Brethower, 2000; Evans & Evans, 1985; Johnson & Layng, 1992) and maintenance of skills over time (Binder, 1996; Driskell, Willis, & Cooper, 1992; Ivarie, 1986; Peladeau, Forget, & Gagne, 2003). While authors have failed to demonstrate an added effect of fluency (i.e., a high rate of responses beyond mastery) beyond the effects of overlearning (i.e., unpaced practice of responses beyond mastery), both types of practice may be more effective than simply practicing to mastery (Binder, 1996; Peladeau et al., 2003). For example, Peladeau and colleagues compared conditions that included typical statistical training in a college level course, an unpaced practice beyond mastery condition, and a paced practice beyond mastery condition. Both conditions that included practice beyond mastery resulted in better scores on course exams than the typical classroom instruction; however, students who practiced the skills at a high rate beyond mastery did not perform better than the students who practiced the skills in an unpaced setting. While there may not be a clear added effect for practice of skills at a fast rate, fluency (paced) practice
may be more efficient than unpaced overlearning practice, since the additional practice trials are conducted quickly, by definition. Improving the efficiency of training might make it more sustainable in schools.

While performance feedback appears to be an effective way to promote intervention fidelity, it remains costly, amounts to a corrective procedure for training, and may not be practical in typical school settings for all staff who need training. Recent studies have shown that group training outside the implementation setting may be effective for most participants when it includes attention to features that promote generalization and maintenance of skills. Adding fluency practice to group training outside the implementation setting may enhance generalization and maintenance of skills, while being cost-efficient, proactive, and sustainable in school settings. If fluency training were effective for most participants, performance feedback could be added during implementation to strategically target situations and participants in which generalization or maintenance is problematic. The combination of effective up-front training and limited follow-up might strike a balance between adequate intervention fidelity and efficient resource allocation in professional development.

**Direct Instruction**

High intervention fidelity of research-based programs is a central component of RTI. If research-based programs are implemented with poor fidelity, efficiency and effectiveness are lost. On the other hand, if schools fail to use interventions that have a high probability of being effective for their students, increased intervention fidelity is not likely to improve outcomes for students (e.g., Foorman et al., 1998). So an important
feature of implementing RTI effectively is to include research-based practices at each
level of intervention. In fact, this component is a legal requirement set forth in IDEIA if
RTI is used to identify a student with a learning disability (IDEIA, 2004).

Direct Instruction is a research-based instructional system with programs for
teaching academic skills such as reading, writing, math, language, spelling, and English
as a second language. Direct Instruction is a complex combination of carefully
constructed and field-tested programs, effective teaching practices, and efficient
organization of instruction designed to maximize student learning (Watkins & Slocum,
2004). Features of program design include detailed content analysis, selection of
examples and wording to ensure clear communication, written formats for teachers to
follow, sequencing of skills to make learning as easy and efficient as possible, and
integrating instruction on prerequisite skills with composite skills, while providing
frequent review. Effective teaching practices include frequent, active student responding
through group unison responses, teaching until skills are mastered, immediate and direct
error corrections, and keeping students engaged through experiences of success, teacher
praise and attention. Organization of instruction includes grouping students so they are
learning appropriately challenging skills, maximizing engaged instructional time, and
continuously assessing students’ skills (Watkins & Slocum, 2004). When implemented
with high fidelity, Direct Instruction can be very effective for improving students’ skills
and helping students experience academic success.

Research on Direct Instruction programs has shown overwhelmingly positive and
educationally important results for a wide range of students across many academic
subjects. The results of Project Follow Through showed that students at the Direct
Instruction schools ended the study with much higher academic achievement, conceptual skills (e.g., reading comprehension, problem-solving), and self-esteem than students at comparison schools, and in other models of instruction (Adams & Engelmann, 1996). Multiple independent reviews of research have found Direct Instruction to be one of the most well supported instructional systems in the research literature. For example, the American Federation of Teachers (AFT) listed Direct Instruction as one of the most promising programs in reading and language arts (AFT, 1998a), remedial reading (AFT, 1999), and school reform (AFT, 1998b) based on strong research evidence. In each of these reports, only a few other programs were supported by similarly strong evidence. The American Institutes of Research (AIR) found Direct Instruction to be one of three schoolwide reform models supported by strong evidence of positive effects on academic achievement (Herman et al., 1999). The Center for Research on the Education of Students Placed at Risk also identified Direct Instruction as one of three schoolwide reform models that was supported by research (Borman, Hewes, Overman, & Brown, 2002). Direct Instruction fits well into an RTI system that uses research-based programs at all levels of instruction.

**Corrective Reading: Decoding**

*Corrective Reading: Decoding* (Engelmann et al., 1999) is a Direct Instruction reading program designed primarily for teaching reading to students in third grade or higher who have had previously received reading instruction, but continue to struggle with learning to read. Many studies support the effectiveness of *Corrective Reading: Decoding* as a remedial reading intervention. In a systematic review of the research literature on *Corrective Reading*, the authors found 28 studies, 27 of which demonstrated
positive outcomes for students with *Corrective Reading* (Przychodzin-Havis et al., 2005). For example, students with reading difficulties improved their reading skills by an average of 12 months over a three-month intervention using *Corrective Reading: Decoding*, which was significantly greater than the gains made by students in control conditions (Somerville & Leach, 1988). *Corrective Reading* has been shown to be more effective than control or comparison interventions with students with learning disabilities (Benner, Kinder, Beaudoin, Stein, & Hirschmann, 2005; Lloyd, Cullinan, Heins, & Epstein, 1980), students reading below grade level (Gregory, Hackney, & Gregory, 1982; Gunn, Biglan, Smolkowski, & Ary, 2000), and students in the juvenile corrections system (Scarlati & Asahara, 2004). Single subject studies have shown that *Corrective Reading: Decoding* is effective in improving reading skills for students with moderate intellectual disabilities (Flores, Shippen, Alberto, & Crowe, 2004) and students in the juvenile corrections system (Drakeford, 2002). *Corrective Reading: Decoding* is a research-based intervention that can have a dramatic effect on students’ reading skills.

**Direct Instruction Teaching Skills**

Promoting fidelity of certain teacher behaviors has been associated with better student outcomes in general and specifically with Direct Instruction programs. For example, frequent opportunities for students to respond, high rate of praise statements, high ratios of positive to negative comments, and accurate error corrections have been shown to improve student outcomes or have been associated with improvements in student outcomes.
**Presentation Rate**

Briskly paced instruction with frequent opportunities for students to make active academic responses has been shown to increase students’ accuracy on academic tasks and on-task behavior (Carnine, 1976; Darch & Gersten, 1985; Gilbertson, Duhon, Witt, & Dufrene, 2008; Sutherland, Alder, & Gunter, 2003). In addition, overt academic responses (i.e., saying the answer) have been found to be more effective than covert academic responses (i.e., looking at the answer quietly) for accurately reading sight words in word lists and connected text (Barbetta, Heron, & Heward, 1993), and stating and writing geography and science facts (Barbetta & Heward, 1993; Drevno et al., 1994).

**Praise Rate**

Positive feedback contingent on student performance can increase students’ correct responding (Darch & Gersten, 1985; Gable & Shores, 1980) and on-task behavior (Darch & Gersten, 1985; Ferguson & Houghton, 1992; Jones et al., 1997; Madsen, Becker, & Thomas, 1968; Sutherland, Wehby, & Copeland, 2000). Although these studies have shown that higher rates of praise are more effective than lower rates, an ideal rate of praise has not been established in the research. Providing praise or positive interactions is a relatively simple, low-cost intervention with positive outcomes in student behavior.
Error Corrections

Providing corrective feedback to students when they make academic errors is related to improved student outcomes in reading (Carlson & Francis, 2002) and other academic areas (Brophy & Good, 1986; Rosenshine & Stevens, 1986). In addition, error corrections that include active student responses after a model have been found to be more effective than those with no response (i.e., looking at the answer) after a model in learning, maintaining and generalizing sight word reading (Barbeta et al., 1993) and other academic skills (Barbeta & Heward, 1993; Drevno et al., 1994).

Positive to Negative Comments Ratio

The positive to negative comments ratio is a measure of how a teacher allots his or her attention. If a teacher provides attention (e.g., praise) to behaviors that she or he wants to increase, such as accurate reading, this is likely to result in increases in these behaviors. However, if a teacher provides a greater amount of attention (e.g., reprimands, negative corrective feedback, or response cost) for student behaviors that he or she does not want the student to engage in, then the student is likely to engage in those behaviors more, rather than less (Madsen et al., 1968; Thomas, Becker, & Armstrong, 1968; Walker, Colvin, & Ramsey, 1995). While some corrective feedback or redirection may be useful in changing behavior, if students learn that they can get a teacher’s attention primarily for inappropriate behaviors, these behaviors are likely to increase. Overall, research has shown that a higher ratio of positive comments to negative comments results in less off-task behavior with students. For example, Thomas and colleagues found that an average of 66% positive statements resulted in much lower rates of off-task behavior compared to an average of 2% positive statements for students in a general education
class. Although an exact ratio has not been confirmed in research, authors with extensive classroom and clinical experience recommend ratios of positive to negative comments of at least 3:1 (Sprick, Garrison, & Howard, 1998) or 4:1 (Walker et al., 1995).

**Student Behaviors**

A more complete model of intervention fidelity takes into account both the teacher behavior, as well as its effect on student behavior (NICHD, 2000; Perepletchikova & Kazdin, 2005). While it is important to show that teachers’ behaviors improved in response to professional development, the ultimate goal is show that students’ academic skills and behavior improved in response to the changes in teachers’ behavior.

**On Task**

On-task student behavior is a measure of the amount of time students allocate to academic tasks. Time on-task in active reading has been associated with better performance on outcome measures in reading (Leinhardt, Zigmond, & Cooley, 1981; Stallings, 1975, 1980). An additional advantage of measuring on-task behavior is its sensitivity to changes in certain teacher behaviors, such as presentation rate (Carnine, 1976; Darch & Gersten, 1985; Sutherland et al., 2003) and attention or praise (Austin & Soeda, 2008; Call, Wacker, Ringdahl, Cooper-Brown, & Boelter, 2004; Darch & Gersten, 1985; Sutherland et al., 2003; Jones et al., 1997).
**Word Reading Accuracy: Percent First-time Correct**

The percentage of correct responses on the first time each item is presented in a lesson captures the academic performance of students. It is likely to be sensitive to current presentation and praise rates, as well as error corrections and other effective teaching procedures from previous days, since effective use of these teaching skills typically results in better learning by students (Barbetta et al., 1993; Barbetta & Heward, 1993; Carnine, 1976; Darch & Gersten, 1985; Drevno et al., 1994; Gilbertson et al., 2008; Sutherland et al., 2003).

**Purpose Statement and Research Questions**

Intervention fidelity is critically important to the validity and effectiveness of RTI systems. Daily or weekly performance feedback has been studied intensively and identified as an effective method for ensuring high intervention fidelity. However, exclusive reliance on performance feedback may be unsustainable in typical schools. Some studies have found that treatment integrity can be achieved with intense prior training that includes high levels of practice and feedback prior to implementation in the classroom. Fluency-based instruction has the advantage of providing multiple practice opportunities in a relatively short amount of time and may improve the application and maintenance skills. The purpose of this study is to examine the effects of a fluency-based training on the intervention fidelity of paraprofessionals instructing students at-risk for reading failure using the *Corrective Reading: Decoding* program.
Research Questions

1. To what extent does fluency-based training in presentation rate, praise rate, error corrections, and positive to negative comments ratio affect paraprofessionals’ use of these skills, and students’ on-task and correct responding, in a classroom setting using the *Corrective Reading: Decoding* program with students at-risk for reading failure?

2. If fluency practice does not result in the demonstration of adequate skills in presenting the reading intervention, to what extent does the addition of graphic and verbal performance feedback based on classroom performance result in acquisition of these skills?

3. To what extent do these skills maintain over time after training and/or performance feedback are discontinued?
CHAPTER III

METHOD

Setting and Context

District RTI Model

This study was conducted in two public elementary schools in a rural school district in Utah. The district served more than 14,000 students across 23 schools. The district’s student population was approximately 90% Caucasian, 7.5% Hispanic, 0.65% Asian, 0.58% African American, 0.40% Pacific Islander, and 0.37% American Indian.

Paraprofessionals who participated in this study were employed by the district to provide supplemental services in tier II interventions in a 3-tier RTI model. The district had been using RTI to improve reading skills of students at risk, and as part of their process for identifying students with learning disabilities (in combination with an IQ discrepancy model). Tier I instruction was conducted in general education classrooms. Tier II interventions included supplemental reading instruction (in addition to instruction in the general education classroom at a different time during the day) in small groups with a paraprofessional using a scripted program. Supplemental instruction was conducted daily for approximately 40 to 45 minutes. Other students received enrichment (e.g., novel studies) or remedial instruction during this time. Prior to this study, students were placed into the Tier II reading groups because they had been identified by the district as needing supplemental reading instruction through the following process. Three times per year, all students are administered the benchmark screenings from the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2002). If they
score below a certain threshold ("some risk" or "at risk"), and display other difficulties in reading (e.g., low scores on the state reading test or difficulties observed by classroom teachers), an individualized literacy plan is developed and monitored by a team of school staff including the literacy facilitator, the classroom teacher, and the principal. If the team decides that the student requires additional intervention, the student is placed into the tier II intervention using *Corrective Reading: Decoding* as the primary intervention. Students are placed into the appropriate level of the *Corrective Reading: Decoding* program based on their performance on the placement test (initially) or based on mastery tests in the program if they continue to need tier II interventions. The district’s third tier is special education.

**Reading Instruction Program in Tier II**

The paraprofessionals who participated in this study implemented level B1 and/or B2 of *Corrective Reading: Decoding* (Engelmann et al., 1999). Each level consisted of 65 lessons with scripts for teacher presentations, student reading books, and student workbooks for written work. These programs featured scripted lessons including carefully designed and sequenced activities on letter and word discrimination, common word endings, sounds for combinations of letters, reading connected text, and answering literal and inferential comprehension questions. Paraprofessionals taught students strategies for applying decoding skills to text. Paraprofessionals modeled new responses, signaled for group unison responses, corrected errors, and provided praise for accurate reading and attending. Students participated in repeated reading of stories with a partner approximately one to two times per week. Reading fluency (rate and accuracy) was assessed for progress monitoring throughout the program.
Pre-baseline Training for Paraprofessionals

**District didactic training.** The district provided initial didactic training to all new paraprofessionals before the beginning of the school year in August. A literacy coordinator gave the general rationale, modeled, and provided guided practice in letter sounds, signals for group responses, and error correction formats for all paraprofessionals (1.5 hours including about 20 minutes of practice). Later, each paraprofessional attended two additional sessions on two different programs that they were likely to teach (e.g., *Early Reading Intervention, Reading Mastery, Corrective Reading, or Reading for All Learners*). During training on each program, literacy coordinators introduced components of the program and described placement test procedures for that program (1.5 hours per program; 3 hours in total). During the Corrective Reading training, no additional format practice was conducted. Thus, the total training time was four and a half hours with a maximum of 3 hours directly relevant to Corrective Reading and no more than 20 minutes of practice.

**District follow-up coaching.** Four of the paraprofessionals reported that they had not been observed or given individual feedback by a literacy facilitator or other coach during the current school year. One reported that a literacy facilitator observed one time and gave her feedback after that observation. Two paraprofessionals had participated in two small group trainings for about one hour each time during this school year, with 10 to 12 other paraprofessionals. Two had participated in one small group training during this school year. These data were based on self-reports by the paraprofessionals. No additional district training was provided during the study.
Study Settings

**Reading instruction.** Paraprofessionals and students were observed in their classrooms during group instruction. In school 1, eight groups of three to five students participated in reading instruction at the same time. Each group sat at a small table with chairs, separated from other groups on three sides by 8-foot tall dividers. In school 2, five groups of three to five students participated in reading instruction at the same time, separated from other groups by book shelves about four feet tall. In another classroom, three groups of students received instruction in language arts at the same time, separated by 8-foot tall dividers. During the performance feedback phase, observers gave the participants feedback at their table in these rooms after the students left or before they arrived.

**Fluency training.** The training at school 1 was conducted in the library of the school where the paraprofessionals worked. The library had approximately ten tables with four chairs each, with low bookshelves surrounding the training area. In school 2, training was conducted in the room where two of the paraprofessionals taught their groups. The room had five tables with five to seven chairs around each. The tables were separated by book shelves about four feet tall.

Participants

**Paraprofessionals**

Five paraprofessionals participated in this study. They taught *Corrective Reading: Decoding* level B1 and/or B2. Literacy facilitators suggested paraprofessionals who could benefit from training. From this group of paraprofessionals, we invited paraprofessionals
to participate if they were teaching five days per week, and were available before or after school for training. We then conducted informal observations prior to baseline to determine if these paraprofessionals had lower skills in at least two areas (described below). Baseline observations confirmed that the selections were appropriate (see Figures 1-4). The paraprofessionals demonstrated a need for training by showing low implementation rates of the reading curriculum in at least two of the following areas during informal and baseline observations:

1. accurate presentation rate – ten or fewer presentations per minute;
2. praise rate – three or fewer praise statements per minute;
3. error correction accuracy – 70% or less correct steps of correction procedure.

All of the paraprofessionals were Caucasian females. Three paraprofessionals had five years of experience as paraprofessionals, one had three years of experience, and one was at the end of her first year. One of the paraprofessionals with five years of experience had seven additional years of experience as a teacher. They all had taught Corrective Reading: Decoding during their entire paraprofessional experience. One paraprofessional had a bachelor’s degree in elementary education, two had attended some college without completing a degree, and two had completed high school. All of the paraprofessionals received their initial training in Direct Instruction from the school district either individually with a literacy facilitator, or in a large group setting at the beginning of the school year. We did not list this information for each paraprofessional separately, because it would be individually identifiable (see Appendix B for the questionnaire).
**Students**

We identified one student with lower reading accuracy and/or on-task rates from each of the paraprofessionals’ groups. We initially selected potential students based on DIBELS oral reading fluency scores, choosing one or two students with lower scores than the others in the same group. We then observed the groups and confirmed our choices based on the students making more word reading errors and showing more off-task behavior than other students. One student was chosen based on lower word reading accuracy, but not higher off-task behavior. Other students in that group appeared to have slightly more off-task behavior, but higher word reading accuracy.

Table 2 presents the relevant demographic information for the students. The table indicates the paraprofessional who taught each student’s group, the students’ ages at the beginning of the study, the number of school months they had been participating in tier II interventions, the oral reading fluency (ORF) benchmark score on the DIBELS assessment just prior to the study (December, 2008), and the students’ risk levels as determined by the ORF scores. DIBELS risk levels include “low risk,” “some risk” and “at risk.” At least 80% of students with scores in the “low risk” range achieve subsequent early literacy benchmarks; approximately 50% of students in the “some risk” range achieve later benchmarks; about 20% of students scoring in the “at risk” range achieve future literacy benchmarks (Good, Simmons, Kame'enui, Kaminski, & Wallin, 2002). All students in this study scored in the “some risk” or “at risk” categories prior to the beginning of the study. All participants were Caucasian.
### Table 2

**Demographic Information for Students**

<table>
<thead>
<tr>
<th>Student</th>
<th>Gender</th>
<th>Grade</th>
<th>Age (years : months)</th>
<th>Months of Instruction in Tier II</th>
<th>ORF&lt;sup&gt;a&lt;/sup&gt;</th>
<th>DIBELS Risk Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alex (Ms. Allen)</td>
<td>M</td>
<td>4</td>
<td>11 : 10</td>
<td>13</td>
<td>90</td>
<td>Some Risk</td>
</tr>
<tr>
<td>Claire (Ms. Dean)</td>
<td>F</td>
<td>3</td>
<td>9 : 1</td>
<td>12</td>
<td>44</td>
<td>At Risk</td>
</tr>
<tr>
<td>Jay (Ms. Jones)</td>
<td>M</td>
<td>3</td>
<td>9 : 4</td>
<td>6</td>
<td>67</td>
<td>Some Risk</td>
</tr>
<tr>
<td>Luke (Ms. Lewis)</td>
<td>M</td>
<td>3</td>
<td>9 : 1</td>
<td>19</td>
<td>61</td>
<td>At Risk</td>
</tr>
<tr>
<td>Todd (Ms. Tate)</td>
<td>M</td>
<td>3</td>
<td>8 : 8</td>
<td>6</td>
<td>89</td>
<td>Some Risk</td>
</tr>
</tbody>
</table>

ORF: Oral Reading Fluency

**Trainer**

The author of this study provided fluency training, and performance feedback (see descriptions below). The trainer had taught Direct Instruction reading curricula to a wide variety of students for 6 years, taught a university class on Direct Instruction reading, and supervised university practicum students, and paraprofessionals in learning to implement Direct Instruction reading curricula. The trainer had a master’s degree in special
education and had completed all requirements except a dissertation for a doctoral degree in special education.

**Research Assistant**

The research assistant was a student in the special education master’s program working on her master’s thesis. She had trained and supervised volunteer tutors in an adult literacy program for five years, and completed a supervised practicum in reading instruction and student teaching for students with disabilities. She had a bachelor’s degree in education and had completed all requirements except her thesis for her master’s degree in special education.

**Reading Facilitators**

The paraprofessionals involved in this study were supervised by “reading facilitators.” These facilitators are teachers who have had training and experience in teaching and coaching others in Direct Instruction, particularly for students who are at risk for reading failure. The district’s seven reading facilitators have received training at the national Direct Instruction Conference, some for multiple years. Each reading facilitator supervises 20-50 paraprofessionals across one or two schools in the district. They provide initial didactic training and follow-up in-class observation and coaching to paraprofessionals as needed. In addition, they oversee the assessment, placement, and progress monitoring of students in the paraprofessionals’ groups. The reading facilitators meet one time per week as a team with the district’s two reading coordinators to discuss these responsibilities.
Dependent Variables

Dependent variables included aspects of paraprofessionals’ and students’ behavior during the word attack (i.e., single word reading) portion of Corrective Reading lessons and paraprofessionals’ instructional skills on fluency training probes. During word attack, paraprofessionals modeled new words or sounds, and students read new words and previously learned words and sounds. If students made mistakes, paraprofessionals provided corrective feedback and extra practice on missed items.

Measures of Paraprofessionals’ Behavior

Presentation rate. Presentation rate was a measure of the number of correct presentations of opportunities for students to respond. A presentation for group responding was counted correct if the paraprofessional used a clear cue, pause, and signal (clearly audible or visible), and appropriate wording of the instruction or direction (i.e., did not vary from the script in a way that changed what the students should do, omit an instruction for the students that could cause an error, or include spurious prompts such as additional scaffolding not included in script). If the paraprofessional asked the students to spell a word, the signal for the first letter of the word was evaluated and scored; if the paraprofessional signaled for each subsequent letter in the word, these were not scored. Individual turns and pauses when the paraprofessional answered students’ questions or explained the meaning of a word were not counted in the rate measure. Accurate presentations were counted as events during an observation period (e.g., 5 minutes), minus the time for individual turns, and presented as a rate of accurate presentations per minute. The target presentation rate for classroom teaching was 15 or more presentations per minute. This target was set based on recommendations by several authorities on
Direct Instruction implementation (K. Engelmann, personal communication, January 6, 2009; R. Harris, personal communication, January 2, 2009; C. Schneider, personal communication, January 1, 2009).

**Praise rate.** Praise statements were defined as positive statements made by the paraprofessional and directed to one or more students contingent upon a social, behavioral or academic response. Praise was counted if it was general (“You did it!” or simply “Yes”) or specific (“Wow, you read those words just right!” or “Yes, the word is…”). Praise statements were counted as events during an observation period (e.g., 5 minutes) and presented as the rate of praise per minute. To keep observation intervals consistent across dependent variables, we did not record praise rate during individual turns – times when the paraprofessional answered student questions or stopped to explain the meaning of a word. However, one paraprofessional, Ms. Jones, almost exclusively used individual turns during baseline, so we included her praise rate during this time. Otherwise, we would not have had a measure of her praise rate during baseline. The target criterion for praise rate during classroom observations was four or more praise statements per minute (Based on C. Schneider, personal communication, January 1, 2009).

**Error correction attempts.** During observations of student responding, observers recorded which words students read correctly and incorrectly (see descriptions below). When observers evaluated paraprofessionals’ instruction, they noted which of the student errors the paraprofessionals stopped and attempted to correct, and which errors they did not correct at all. An error correction attempt was defined as the paraprofessional stopping after the student error and making any kind of corrective statement to the
student or group, regardless of the quality of the error correction. If the paraprofessional presented the next item without stopping after the error, this was not counted as a correction attempt. The percentage of error correction attempts was calculated as the number of errors the paraprofessional attempted to correct divided by the total number of student errors observed, multiplied by 100%.

**Error correction accuracy.** Of the errors that the paraprofessionals attempted to correct, the percentage of error correction steps completed accurately was calculated. Academic errors corrected accurately were counted when students made an error and the paraprofessional:

a. stopped before presenting another item (stop);

b. said the correct answer just before asking students to read the word again.

If the paraprofessional gave the definition of the word without saying the word again, this was not a correct model (model);

c. asked the student(s) to respond to the missed item, even if the paraprofessional did not model the word. For an error on a group response, another group response should be called for here (test);

d. asked the student(s) to spell the missed word (spell);

e. after the spell step, asked student(s) to read the word again (test 2);

f. provided opportunity for one or more responses on other items then asked the student(s) to respond to the missed item again. If either of these steps was missing, this step would be counted as an error. If the error was on a group response, another group response should be called for; for an error
on an individual turn, either a group response or individual response can be called for (retest).

An academic error was counted when one, some or all students answer with a response that is different than that called for in the teacher presentation book. Sometimes students provided the correct answer, but answered slightly before or after the signal, or dragged out their responses. These errors were considered “signal errors” (Lignugaris/Kraft & Marchand-Martella, 1993), and while important to address, were not counted as errors in this study (although we provided training on how to recognize and address them).

Accurate error corrections were measured as the percentage of the six steps described above that were presented accurately for each student error.

When reporting the data, each data point included at least five error correction attempts to reduce variability that would be introduced by data points based on few opportunities to correct errors. If the paraprofessional attempted to correct fewer than five errors in a session, the data from this session was combined with that from subsequent sessions until there were at least five error correction attempts. The data point was then calculated by adding the total number of correct steps, and dividing that by the number of possible steps, multiplied by 100%. However, data from sessions immediately before a phase change were not combined with data from other sessions, even if it included fewer than five error correction attempts. This exception was made to maintain the separation of data across phases. The target criterion for percent accuracy on error corrections was 95% of error correction steps completed accurately (adapted from K. Engelmann, personal communication, January 6, 2009, and Lignugaris/Kraft & Marchand-Martella, 1993).
Positive to negative comments ratio. Positive feedback was counted as an event and defined the same way as praise was defined above (i.e., positive statements to students contingent upon a social, behavioral or academic response). Negative feedback was counted as an event and defined as telling students what was wrong with a current behavior (in contrast to an error correction that modeled the correct answer and/or indicated what the student should do in the future). Examples of negative feedback included, “I didn’t hear everyone answer,” “No, that’s incorrect,” or “Stop that!” If the paraprofessional said, “I need everyone to answer,” this type of statement was not counted as negative feedback. The positive to negative comments ratio was calculated as the number of positive feedback statements (i.e., praise) divided by the total number of feedback statements (positive and negative), multiplied by 100%. Although an ideal ratio of positive to negative feedback has not been determined empirically, a ratio of 4 positive statements to 1 negative statement (80% positive statements) was set as the target criterion based on a convergence of expert opinion (K. Engelmann, personal communication, July 23, 2008; R. Harris, personal communication, January 2, 2009; Sprick et al., 1998; Walker et al., 1995). Including some negative comments would not necessarily be considered a problem, but most comments to students should be positive, as reflected in the criterion of 80% positive statements.

Paraprofessionals’ skills in training probes. During the fluency training phase, we also measured paraprofessionals’ presentation rates, praise rates, and error correction accuracy in training probes (see Independent Variables for a complete description) at the end of each training session. Those skills were coded in the same ways as described above.
Measures of Students’ Behavior

On-task. The target student in each group was observed to estimate how much time he or she participated in academic tasks. This behavior was measured using whole interval recording with 10-second intervals (Kazdin, 1982) for approximately 5 minutes. If the word attack portion of the lesson lasted less than 5 minutes, we calculated the on-task percentage based on the number of whole intervals observed during the lesson. If the word attack portion of the lesson lasted more than 5 minutes, we concluded the observation of on-task behavior after 5 minutes. The student must have demonstrated on-task behavior for the full 10 seconds for the interval to be counted as on-task. We defined on-task behavior to capture behavior that was consistent with rules observed in the groups, such as “sitting tall”, answering on signal, following directions, following under the words, etc. On-task behavior was defined as the student answering with an academic response less than three seconds after the teacher signaled for a response (i.e., minor signal errors due to early or late responses were not counted as off-task, but non-responses were), while the student was seated (i.e., bottom and/or legs in contact with the chair) with all four legs of the chair touch the floor in the instructional group. Behaviors such as singing, crying, shouting irrelevant responses, making audible comments that do not pertain to the task, looking away from the task for 3 s or more, answering 3 s before or after the teacher’s signal, failing to follow a teacher direction within 3 s, playing with objects (i.e., object off the table or manipulating object, unless following under words), holding both hands at or above one’s head for 3 s or more, pulling any part of one’s shirt above the plane of one’s chin, licking anything besides lips, leaving one’s chair (e.g., walking around, jumping), tipping one’s book off the table, laying one’s head or trunk on
the table and/or covering mouth or face so that responses could not be heard, putting any part of one’s head below the plane of the table top, hitting or audibly tapping the table or chair, stomping on the ground or table, hitting or touching another student or the teacher with the student’s body or another object (except for giving a high-five or something similar to the teacher or if allowed by the teacher) were not counted as on-task. If the student was waiting for the paraprofessional to give another instruction, on-task was counted as sitting quietly (no audible responses) in his/her chair. The student could be looking at the teacher, the book or the student with whom the paraprofessional was interacting. This measure was quantified as the percentage of whole intervals observed with on-task behavior (i.e., the number of intervals with on-task behavior divided by the total number of intervals observed, multiplied by 100%).

**Word reading accuracy: Percent first-time correct.** Correct, incorrect or no response was recorded for each academic item that the target student had not previously responded to during that day’s session (i.e., first time responses). When the teacher modeled a word, another student answered before the teacher presented a word, or the student’s mouth was not visible to the observers, the observers did not code the answer for accuracy. Correct responses were counted as answers that were the same as those in the teacher’s presentation book or pronunciations that closely matched those in the teacher’s guide (i.e., for letter sounds). An error was recorded if the response was different than in the teacher’s presentation book or pronounced differently than in the teacher’s guide, even if the student self-corrected after giving an incorrect response. No response was counted if the student did not respond within one second of the teacher’s signal. Percent first time corrects were quantified as the number of correct responses
divided by the total number of first time responses (correct, incorrect and no response), multiplied by 100%.

Reliability of Dependent Measures

At least 30% of observations across each phase (baseline, all interventions and maintenance) in both the training (for paraprofessionals) and classroom settings (for paraprofessionals and students) were independently coded by a second observer. For each measure based on event recording (i.e., presentation rate, praise rate, and negative comments) proportional agreement was determined for each 10-second interval by dividing the smaller number of observations by the larger number of observations, then multiplying by 100%. If neither observer scored anything or both observers scored the same number of observations, the interval was counted as 100% agreement. Proportional agreement for the session was then calculated by averaging the agreements across intervals for each measure. We chose this estimate of interobserver agreement because it is well-suited to agreement on relatively high rate behaviors (such as presentation rate) in which small disagreements might be expected. With proportional agreement, if one observer recorded four presentations and the other recorded five during an interval, that interval would be scored as 80% agreement. If exact agreement were used in this example, the interval would simply be counted as 0% agreement. Proportional agreement seems to be a fair estimate of interobserver agreement in this case, because it seems likely that observers might code one or two presentations incorrectly, but unlikely that all four observations that both recorded would be incorrect.

For the measure based on percentage of steps completed correctly (error corrections) inter-observer agreement was based on the number of exact agreements on
each step, divided by total number of steps (agreements plus disagreements), multiplied by 100%. For the measure based on whole interval recording (students’ on-task behavior), agreement or disagreement between the coders was determined for each interval. Interobserver agreement was calculated as intervals with agreement divided by the total number of intervals (agreements plus disagreements), multiplied by 100%. For student reading accuracy, interobserver agreement was calculated by dividing agreements on correct or incorrect word reading by the number of words read, multiplied by 100%. When agreement for any measure fell below 90%, the observers used that observation to practice rating until agreement improved (while reporting only the agreement on the first observation session). Occasionally, if observers did not reach 90% agreement even after two practice sessions, the definition for that measure was refined if needed.

**Social Validity Measures**

Three measures of social validity were used. We evaluated the perceived importance of paraprofessional teaching skill measures, the feasibility of training procedures for the district, and the acceptability of the training to the paraprofessionals.

**Paraprofessional teaching skills.** To evaluate the social validity of the measures we used, we asked experts in Direct Instruction to rate videos of the paraprofessionals. We compared the presentation rate and praise rate criterion levels used in this study with DI experts’ qualitative ratings of these skills. Based on our criteria, we rated presentation rates of 15 items per minute or more as “acceptable” and lower rates as “unacceptable.” Similarly, four or more praise statements per minute were “acceptable” and fewer were “unacceptable.” Then we rated the overall session based on presentation and praise rates. If both the presentation rate and praise rate was “acceptable”, the overall rating for that
session was “acceptable.” If either criterion was not met, the overall session was “unacceptable.” We made these ratings on four video-recorded sessions for each paraprofessional – one video from each phase in the study (i.e., baseline, fluency training, maintenance and performance feedback). For Ms. Tate, who did not receive performance feedback, a second video from baseline was used. We selected sessions to sample a range of performance for presentation and praise rates. For example, we chose sessions with some high presentation rates, some criterion-level presentation rates, and some low presentation rates.

Five experts in Direct Instruction reading rated the videos. Each expert rated the four videos from one paraprofessional. The videos were put in a random order on a DVD. We asked DI experts to rate each session individually for the quality of instruction on a 4-point likert-type scale (1= Poor to 4 = Excellent). See Appendix C for the rating forms. Experts’ ratings of “poor” and “OK” were coded as agreements with the criterion rating of “unacceptable.” Experts’ ratings of “good” and “excellent” were considered agreements with the criterion rating of “acceptable.” We assessed the agreements and disagreements between the criterion ratings and experts on presentation rate, praise rate, and overall session quality. We summarized this as percent agreement (agreements divided by agreements plus disagreements, multiplied by 100%) for each skill separately, and for all ratings combined. In addition, we analyzed correspondence between our measures and expert ratings in a 2x2 contingency table.

Feasibility of training. The two district reading coordinators completed a questionnaire together at the end of the study to determine if the procedures used for training would be feasible for the district in the future. The questionnaire included six
statements which were rated on a 5-point likert-type scale ranging from “Strongly Disagree” to “Strongly Agree.” Statements included items such as, “The training was simple enough to be easily implemented by district staff,” and “Overall, we are likely to use this training in the future.” See Appendix D for the full questionnaire.

**Acceptability of training.** At the end of fluency training, paraprofessionals completed a confidential questionnaire regarding the acceptability of the training procedures. The questionnaire included 12 statements about the training with a 5-point likert-type scale ranging from “Strongly Agree” to “Strongly Disagree.” Statements included, “I felt that the training I received helped me improve my classroom teaching,” “I feel that the training I received helps my students read better,” and “This training had little effect on my teaching in the classroom” (reverse coded). See Appendix E for the full questionnaire.

**Measurement Procedures**

Observers coded students’ reading accuracy daily live during the word attack lesson, while videotaping the session. For all other measures (paraprofessionals’ presentation rate, praise rate, negative comments, error correction accuracy, and students’ on-task), observers recorded data from video recordings of the sessions. Observers used paper and pencil recording methods for students’ reading accuracy, on-task, and paraprofessionals’ error correction accuracy. Observers used a computer program for recording data on paraprofessionals’ presentation rate, praise rate, and negative comments.
Independent Variables

Independent Variables for Paraprofessionals

Fluency training. The author and a research assistant conducted fluency training with small groups (9-13) of paraprofessionals and literacy facilitators one hour daily for five days. We conducted daily classroom observations concurrently with this training. For each group of paraprofessionals, one paraprofessional was a participant in the study, while the others only received training and acted as students during fluency practice (no data were collected on their teaching).

During fluency training, the trainer presented the rationale for the skill to be taught, modeled the skill, provided practice in role-playing situations, and provided positive and corrective verbal feedback (see Appendix F and Table 14). After the paraprofessionals correctly performed each skill, they practiced it to a fluency goal that was above what was required in the classroom (see Table 3 for fluency goals). Fluency goals for training were set above the target criteria for teaching in the classroom, to

Table 3

Teaching Criteria for Classroom Setting and Fluency Goals for Training Setting.

<table>
<thead>
<tr>
<th>Teaching Behavior</th>
<th>Classroom Target Criterion</th>
<th>Fluency Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presentation rate</td>
<td>15 or more per minute</td>
<td>20 or more per minute</td>
</tr>
<tr>
<td>Praise rate</td>
<td>4 or more per minute</td>
<td>6 or more per minute</td>
</tr>
<tr>
<td>Error corrections</td>
<td>95% accurate steps</td>
<td>95% accurate steps with target presentation rate</td>
</tr>
<tr>
<td>Positive to Negative</td>
<td>4:1; 80% positive comments</td>
<td>n/a</td>
</tr>
</tbody>
</table>
increase the likelihood that the skills would generalize to the classroom. The fluency goal 
for presentation rate in training sessions was 20 presentations per minute. The fluency 
goal for praise rate was six praise statements per minute (see Table 3). For error 
correction accuracy, the goal in training was the same as the criterion in the classroom 
(95%), but the goal in training was to be maintained with the higher presentation and 
praise rates. We did not use a fluency goal for positive to negative ratio, because 
paraprofessionals acting as students during training probes were not pretending to have 
off-task behavior, and no negative comments were observed.

The trainer gave the rationale for, modeled, and pointed out the critical features of 
effective cue, pause, and signal routines on the first day of training for presentation rate. 
These features included presenting the script as written, not answering with the students, 
and providing consistent cues, pauses, and signals. Initially, the paraprofessionals 
practiced presentations for accuracy with a simplified script (i.e., just asking “What 
word?” for each word on a list) and no other required skills (e.g., paraprofessionals acting 
as students did not make errors or add off-task behavior). The trainer pointed out that this 
practice should be used to practice a pace that might be appropriate in their groups, since 
the goal of fluency practice was not necessarily to present all items very fast, but to be 
able to adjust one’s presentation rate to meet students’ needs (i.e., slow down and speed 
up depending on student proficiency). During this “accuracy” practice, the trainer and 
research assistant provided feedback to the paraprofessionals (positive and corrective) 
and answered any questions that the participants posed. After accuracy practice, the 
paraprofessionals practiced for increased fluency. The target was 20 presentations per 
minute. The paraprofessionals graphed their performance after each round of practice.
Praise rate training started with the rationale for praising students within the context of basic behavior management strategies. The trainer explained critical features of effective praise (e.g., specific, varied, appropriate to the student, enthusiastic, and sincere). Participants then brainstormed varied student behaviors to praise. To achieve varied praise and fluent presentation of praise statements, we asked paraprofessionals to generate a praise statement for each common letter of the alphabet. They practiced saying these statements fluently (i.e., timing themselves in saying varied praise statements and increasing this rate over consecutive one-minute timings). The trainer modeled effective praise within a brief instructional presentation. Paraprofessionals then practiced the simple script with praise statements, first for accuracy, and then for fluency (target criterion = 6 praise statements per minute). Feedback was given throughout accuracy and fluency practice when needed and participants graphed their praise and presentation rates.

We introduced the procedures for academic error corrections on the third day of fluency training. First, the trainer presented the rationale for correcting errors immediately and explicitly, reviewed the steps for an academic error correction, modeled the steps and pointed out their critical features. The paraprofessionals practiced correcting errors using the simple script while the trainer and research assistant gave feedback. Then paraprofessionals practiced with a fluency goal of 15 presentations per minute, while still correcting errors on approximately 40% of words. Finally, paraprofessionals graphed the results from the fluency timings for presentation rate, praise rate and error correction accuracy on separate graphs.

During training on positive to negative comments ratio, paraprofessionals received training on a simple behavior management technique called “praise around.”
Praise around includes (a) ignoring students’ minor off-task behavior (e.g., talking during inappropriate times, laying head on table, tapping a pencil, etc.), (b) praising other students, targeting an “incompatible” behavior (e.g., “I like how Adam answered right when I signaled!”), and (c) when the student who was off-task before gets on-task, praising that student immediately. The trainer gave the rationale for this procedure and for having a high ratio of positive to negative comments. The trainer modeled this procedure, and the paraprofessionals practiced for accuracy within the context of presenting a lesson. Fluency practice followed. Paraprofessionals graphed their presentation, praise rates and error correction accuracy as before.

The trainer presented information and practice on other skills not directly measured during this study, including providing appropriate “think time” on words that might be difficult for students, and making signal error corrections (when students do not answer on signal).

Most training sessions ended with a training probe. During this five minute session, we recorded data for a measure of the paraprofessionals’ skills in the training setting. In all training probes, we provided paraprofessionals with a Corrective Decoding script from the lesson that they had most recently taught. Other paraprofessionals acted as students, with word lists indicating where to make errors, and no off-task behavior. After taking data on the training probe, the trainer presented graphic and verbal performance feedback (i.e., graphs with the paraprofessionals’ rates of praise and presentation, and percent correct of error corrections, and a line that marked the target rate, with positive verbal feedback about improvements) to the paraprofessionals individually.
**Performance feedback.** After fluency training ended, we observed participants in classrooms during maintenance observations to determine if the effects of fluency training maintained when training was discontinued. If the presentation rate and/or praise rate did not maintain at or above the target criteria for the classroom, we provided individualized performance feedback to paraprofessionals using procedures suggested by previous research (Codding, Feinberg, Dunn, & Pace, 2005; DiGennaro et al., 2005, 2007; Gilbertson et al., 2007; Hagermoser-Sanetti et al., 2007). The trainer observed the paraprofessional daily in the classroom and gave brief (approx. 5 minutes) verbal and graphic feedback after class. Graphic feedback included Excel graphs that showed the participant’s classroom performance over the course of the study on each dependent variable (presentation rate, praise rate, error correction accuracy, and positive to negative comments ratio). Each graph included a target criterion line for each instructional skill in the classroom setting. Verbal feedback included positive feedback for components implemented well (referring to the graph), pointing out components that did not meet the criterion, and suggestions for how those components could be improved. The trainer offered to answer any other questions the paraprofessional had (see Appendix G for the performance feedback protocol). Finally, the trainer gave the paraprofessional a written summary of the feedback (see Appendix H). The trainer wrote down the data for the day’s performance on each instructional skill with a brief summary of the positive feedback and suggestions.

**Training Fidelity**

**Fluency training.** An observer assessed implementation fidelity of the training components of this intervention through a checklist of targeted components (e.g.,
providing the rationale for the skill to be presented, modeling the skill accurately, providing practice on the skill, etc.). See Appendix I. For 73% of training sessions, the research assistant assessed implementation fidelity by completing the checklist. The number of components completed accurately was divided by the total number of components, and multiplied by 100% for the measure of implementation fidelity.

Interobserver agreement was evaluated for ratings on the measure of implementation fidelity of the training sessions for 53% of training sessions. For each component evaluated on the checklist, observers (the trainer and research assistant) assessed agreements and disagreements between their ratings, and divided agreements by total components (agreements plus disagreements), multiplied by 100%.

**Performance feedback.** An observer (usually the research assistant), assessed treatment fidelity of performance feedback sessions using a checklist of steps during 52% of performance feedback sessions (see Appendix J). We calculated fidelity in the same was as we did with fluency training fidelity.

We evaluated interobserver agreement for 38% of all performance feedback sessions. During these observations, the trainer self-assessed treatment fidelity while the research assistant independently assessed treatment fidelity. We calculated agreement in the same way we calculated interobserver agreement on fluency training fidelity.

**Design**

We used a multiple baseline across participants design (Kazdin, 1982) to evaluate the research questions in this study. The data presented include all sessions in which observations were conducted. Baseline conditions included observations of the
paraprofessionals and students in their classrooms. (These observations occurred well after the district didactic training and follow-up coaching.) Baseline continued until a stable or downward trend in baseline data was observed for presentation rate and praise rate for most participants. After baseline, one participant began the fluency training phase. In one case, two paraprofessionals started fluency training simultaneously because we could not stagger their training times (fluency training could only be conducted in the morning at this site). Fluency training continued for five sessions. After fluency training ended, maintenance observations began. Maintenance observations continued until a participant’s presentation and/or praise rate decreased below the target criterion for two or three days with a downward or flat trend. Sometimes this phase continued a bit longer due to scheduling constraints (e.g., a paraprofessional was not available to start performance feedback; we did not want to start performance feedback the day before spring break). When we introduced performance feedback for the skill that was below criterion (i.e., presentation or praise rate), we included performance feedback on all the paraprofessional skills (i.e., presentation rate, praise rate, positive to negative comments ratio, and error correction accuracy). So, if the paraprofessional had a low presentation rate and error corrections, but adequate praise rate and positive to negative comments ratio, we included feedback on how to improve the presentation rate and error corrections, plus positive feedback on praise rate and positive to negative comments ratio. Performance feedback continued until participants’ presentation and praise rates were above their criteria for three consecutive days, or until the end of supplemental reading instruction. Across legs of the multiple baseline, we attempted to introduce fluency training in the later legs after an effect had been seen in a previous leg, and while the later
baselines were flat or decreasing, but time constraints resulted in the introduction of fluency training on at least one occasion when a participant’s praise rate had increased in baseline.

**Interobserver Agreement**

Table 4 shows the percent of sessions in which interobserver agreement was assessed for each dependent measure and the appropriate statistical summary of results. We calculated proportional agreement between observers for presentation rate, praise rate and negative comments. We calculated exact agreement for error corrections, student on-task, and word reading accuracy. See Appendix K for a discussion of issues with computer data collection, and Appendix L (including Table 15) for a discussion of interobserver agreement using Kappa.

Table 4

**Interobserver Agreement for Dependent Measures**

<table>
<thead>
<tr>
<th>Dependent Measure</th>
<th>Percent of sessions</th>
<th>Agreement</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Proportional</td>
<td></td>
</tr>
<tr>
<td>Presentation Rate</td>
<td>32</td>
<td>93.04</td>
<td>76.36-100.00</td>
</tr>
<tr>
<td>Praise Rate</td>
<td>32</td>
<td>93.62</td>
<td>78.47-100.00</td>
</tr>
<tr>
<td>Negative Comments</td>
<td>33</td>
<td>98.61</td>
<td>83.33-100.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exact</td>
<td></td>
</tr>
<tr>
<td>Error Corrections</td>
<td>32</td>
<td>95.99</td>
<td>83.33-100.00</td>
</tr>
<tr>
<td>Student On-Task</td>
<td>34</td>
<td>87.62</td>
<td>41.67-100.00</td>
</tr>
<tr>
<td>Reading Accuracy</td>
<td>47</td>
<td>90.07</td>
<td>64.58-100.00</td>
</tr>
</tbody>
</table>
Treatment Fidelity

Fluency Training

The research assistant assessed treatment fidelity during 73% (11/15) of fluency training sessions. The trainer completed 88% of the steps during training fidelity sessions, including three sessions when a paraprofessional could not complete the training probe due to illness. When those sessions are removed from the calculation, treatment fidelity reached an average of 95%. If a step was missed (typically intentionally skipped due to time constraints), we completed that step during the next training session. The trainer and research assistant independently rated sessions for 73% of fidelity sessions (i.e., 53% of all fluency training sessions). Interobserver agreement on treatment fidelity ratings was an average of 99%.

Performance Feedback

During performance feedback sessions, the trainer or second observer assessed treatment fidelity during 48% of feedback sessions (10/21). The trainer completed 98% of steps correctly. Occasionally, the trainer did not ask the paraprofessional if she had any questions on a particular skill. Interobserver agreement was assessed for 80% of treatment fidelity sessions (8/10), which was 38% of all performance feedback sessions (8/21). Average interobserver agreement was 99% for all treatment fidelity ratings on performance feedback.
CHAPTER IV
RESULTS

The purpose of this study was to evaluate the effects of fluency training on paraprofessionals’ instructional skills and students’ responding in a Direct Instruction reading intervention. This study addressed the following research questions:

1. To what extent does fluency-based training in presentation rate, praise rate, error corrections, and positive to negative ratio affect paraprofessionals’ use of these skills, and students’ on-task behavior and word reading accuracy, in a classroom setting using the *Corrective Reading: Decoding* program with students at-risk for reading failure?

2. If fluency training does not result in the demonstration of adequate skills in presenting the reading intervention, to what extent does the addition of graphic and verbal performance feedback based on classroom performance result in acquisition of these skills?

3. To what extent do these skills maintain over time after training and/or performance feedback are discontinued?

First, we present the data on paraprofessionals’ behavior, then on students’ behavior, summarizing the levels, trends, variability and latency of change across phases (Kazdin, 1982; Horner et al., 2005). See Appendix M for a discussion of other possible approaches to presenting the data. Figures 1-6 show graphs of each dependent variable. Tables 3-9 present the means, ranges and slopes for each dependent variable for each phase across participants.
**Paraprofessionals’ Behavior**

**Presentation Rate**

Figure 1 shows paraprofessionals’ presentation rate in classroom observations throughout the study and in training probes during the training phase. The solid horizontal line represents the target criterion in the classroom: 15 presentations per minute. The dashed horizontal line represents the fluency goal of 20 presentations per minute during training probes. Table 5 includes the statistical summary of each phase across participants.

During baseline, Ms. Allen’s accurate presentation rate was low ($M = 7.97$; range, 3.44 – 12.32), but had an increasing slope (1.08). In training probes, Ms. Allen’s presentation rate was high, at or above the fluency goal (20 presentations per minute) on four of five days. At the same time, her classroom presentation rate increased ($M = 16.04$; range, 13.74 – 18.38) over baseline, but the slope decreased (-0.80). The slope continued to be negative during maintenance (-0.63), decreasing to a level below the criterion for five of the last six days. During performance feedback Ms. Allen’s presentation rate immediately increased and maintained above the criterion during this phase ($M = 23.14$; range, 18.36 – 28.71). During the second maintenance phase, Ms. Allen’s presentation rate remained above the criterion ($M = 25.81$; range, 21.69 – 31.30), but with a decreasing slope again (-0.45).
Figure 1. Paraprofessionals’ accurate presentations per minute during word attack
Table 5

Paraprofessionals’ Accurate Presentation Rate: Phase Means, Ranges, and Slopes

<table>
<thead>
<tr>
<th>Paraprofessional</th>
<th>Baseline Mean (Range)</th>
<th>Slope</th>
<th>Fluency Training Mean (Range)</th>
<th>Slope</th>
<th>Maintenance Mean (Range)</th>
<th>Slope</th>
<th>Performance Feedback Mean (Range)</th>
<th>Slope</th>
<th>Maintenance 2 Mean (Range)</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ms. Allen</td>
<td>7.97 (3.44-12.32)</td>
<td>1.08</td>
<td>16.04 (13.74-14.38)</td>
<td>-0.80</td>
<td>19.99 (8.05-16.02)</td>
<td>-0.63</td>
<td>23.14 (13.36-28.71)</td>
<td>0.27</td>
<td>25.81 (21.69-31.30)</td>
<td>-0.45</td>
</tr>
<tr>
<td>Ms. Dean</td>
<td>0.35 (0.00-1.71)</td>
<td>0.00</td>
<td>12.90 (9.68-14.60)</td>
<td>-0.61</td>
<td>14.82 (11.67-18.37)</td>
<td>0.25</td>
<td>15.50 (13.95-17.27)</td>
<td>1.11</td>
<td>15.51 (15.34-15.79)</td>
<td>0.23</td>
</tr>
<tr>
<td>Ms. Jones</td>
<td>0.03 (0.00-0.29)</td>
<td>-0.01</td>
<td>13.2† (11.43-14.30)</td>
<td>-1.35</td>
<td>15.94 (0.00-21.89)</td>
<td>2.68</td>
<td>19.33 (17.30-21.35)</td>
<td>4.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ms. Lewis</td>
<td>12.11 (6.13-18.25)</td>
<td>-0.07</td>
<td>13.59 (11.51-15.30)</td>
<td>0.97</td>
<td>14.01 (12.18-15.76)</td>
<td>-0.76</td>
<td>16.30 (13.27-18.00)</td>
<td>0.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ms. Tate</td>
<td>19.98 (14.15-26.13)</td>
<td>0.39</td>
<td>20.26 (16.87-23.71)</td>
<td>-1.76</td>
<td>16.96 (15.66-18.59)</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Ms. Dean answered with students on every word reading response during initial observations. Prior to baseline observations, the author of this study requested that Ms. Dean not answer with the students, so that we could assess students’ accuracy in reading. Ms. Dean complied. During baseline observations, Ms. Dean presented items inaccurately (i.e., due to an inaudible or inconsistent signal), resulting in an average presentation rate of 0.35 items per minute (range, 0 – 1.71) with a flat slope. On fluency training probes, Ms. Dean presented items more quickly and accurately, with two of five data points above the fluency goal. Her classroom presentation rate during fluency training increased over baseline ($M = 12.90$; range, 9.68 – 14.60), but did not reach the classroom criterion. During maintenance, Ms. Dean’s presentation rate hovered around the classroom criterion ($M = 14.82$; range, 11.67 – 18.37), with a slightly increasing slope (0.25). We initiated performance feedback for Ms. Dean because of low praise rates, but we also provided feedback on presentation rate. Her presentation rate remained near the classroom criterion with an increasing slope during performance feedback. During the second maintenance phase, this pattern continued.

Ms. Jones presented almost no items correctly during baseline, with all but one session at zero. Ms. Jones participated in training for 5 days, but health considerations prevented her from completing training probes on four days, and classroom observations on 2 days. On the single training probe she was able to complete, Ms. Jones included 1.5 correct presentations per minute. In classroom observations during fluency training, Ms. Jones’ presentation rate increased dramatically, but did not quite reach the classroom criterion ($M = 13.28$; range, 11.43 – 14.30). The first day after training ended, Ms. Jones’ presentation rate fell to zero, but it increased during the rest of the maintenance phase to
levels above the classroom criterion ($M = 15.94$; range, 0.00 – 21.89). All of the last five data points of this phase were above the classroom criterion. When we introduced performance feedback to Ms. Jones due to low praise rates, her presentation rate remained above criterion for those two days. We were not able to continue performance feedback or add another maintenance session, because tier II groups ended.

Ms. Lewis presented a mean of 12.11 items correctly per minute during baseline with a flat slope and high variability (range, 6.13 – 18.25) near the target criterion line. On training probes, her presentation rate was at or above the fluency goal (20 presentations per minute) on the last three days of training. In the same phase, on classroom observations, Ms. Lewis’ presentation rate was increasing (0.97), with much less variability than during baseline (range, 11.51 – 15.30). Ms. Lewis met the classroom criterion on the last day of training. During maintenance, the variability in Ms. Lewis’ presentation rate continued to be low (range, 12.18 – 15.76), but the slope was decreasing (-0.76). Since she met the classroom criterion for presentation rate on only one day during maintenance, we initiated performance feedback with Ms. Lewis. During performance feedback, variability remained lower than baseline, and her presentation rate exceeded the classroom criterion on two of three days. We were not able to complete the performance feedback phase due to the end of tier II groups at the school.

Ms. Tate presented items quickly and accurately, with a mean of 19.98 presentations per minute during baseline, with a slightly increasing slope and high variability primarily above the criterion line (range, 14.15 – 26.13). When she participated in fluency training probes, she presented items at rates higher than the fluency goal on three consecutive days. Her presentation rate remained above the
classroom criterion during fluency training and maintenance. During maintenance, the variability in her presentation rate (range, 15.66 – 18.59) decreased dramatically compared to baseline.

Across the baseline phases, three paraprofessionals demonstrated baseline presentation rates well below the target criterion of 15 presentations per minute in the classroom. The other two paraprofessionals had highly variable baseline patterns around or above the target criterion line. Three of the five paraprofessionals demonstrated higher presentation rates during fluency training. The other two paraprofessionals reduced the variability in their presentation rates during fluency training. When we introduced fluency training to some legs, the other legs remained unaffected. For example, in the first phase change, Ms. Allen and Ms. Dean were simultaneously introduced to fluency training and the data patterns in the remaining legs of the multiple baseline for presentation rate remained flat (Ms. Jones and Ms. Tate) or decreased (Ms. Lewis) during this phase.

Praise Rate

Figure 2 shows the number of praise statements paraprofessionals made per minute during classroom observations and training probes. The target criteria for praise rate in the classroom was six per minute (shown by a solid horizontal line) and the fluency goal during training sessions was six per minute (shown by dashed line in fluency training phase). The vertical dashed line in the training phase is placed just before the day
Figure 2. Paraprofessionals’ praise statements per minute.
<table>
<thead>
<tr>
<th>Paraprofessional</th>
<th>Baseline</th>
<th>Fluency Training</th>
<th>Maintenance</th>
<th>Performance Feedback</th>
<th>Maintenance 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (Range)</td>
<td>Slope</td>
<td>Mean (Range)</td>
<td>Slope</td>
<td>Mean (Range)</td>
</tr>
<tr>
<td>Ms. Allen</td>
<td>2.40 (0.94-3.77)</td>
<td>0.23</td>
<td>4.31 (2.66-5.91)</td>
<td>0.71</td>
<td>8.49 (5.70-13.14)</td>
</tr>
<tr>
<td>Ms. Dean</td>
<td>1.86 (0.78-3.14)</td>
<td>-0.13</td>
<td>3.46 (1.89-4.95)</td>
<td>0.34</td>
<td>3.63 (1.54-5.51)</td>
</tr>
<tr>
<td>Ms. Jones</td>
<td>0.35 (0.00-1.68)</td>
<td>0.01</td>
<td>1.60 (0.59-2.54)</td>
<td>0.98</td>
<td>2.39 (0.98-3.48)</td>
</tr>
<tr>
<td>Ms. Lewis</td>
<td>2.56 (0.88-4.33)</td>
<td>-0.02</td>
<td>6.10 (2.94-8.47)</td>
<td>1.20</td>
<td>6.00 (3.39-7.19)</td>
</tr>
<tr>
<td>Ms. Tate</td>
<td>0.40 (0.00-1.3c)</td>
<td>-0.01</td>
<td>3.19 (1.45-4.15)</td>
<td>1.26</td>
<td>5.42 (3.55-8.15)</td>
</tr>
</tbody>
</table>
that paraprofessionals began practicing praise (on the first day of training, we discussed the rationale for and features of appropriate praise, but provided no fluency practice on it). Table 6 gives the mean praise rates, ranges and slopes for each paraprofessional across the phases of the study.

Ms. Allen praised students a mean of 2.40 times per minute during baseline (range, 0.94 – 3.77), with a slightly increasing slope (0.23). On training probes, her praise rate increased over time, with three of five data points above the fluency goal (six praise statements per minute). During the first two days of training, Ms. Allen did not increase her classroom praise rate over baseline; however, during the last two days of training, her praise rate exceeded her previous performance and the classroom criterion (four praise statements per minute). Her praise continued to increase during maintenance ($M = 8.49$; range, 5.70 – 13.14) with increased variability. We initiated performance feedback with Ms. Allen due to low accuracy in presentations, and included feedback on all instructional skills (including praise). Ms. Allen’s praise rate decreased dramatically ($M = 3.57$; range, 1.42 – 5.17) compared to maintenance; however, by the end of the phase, she praised at rates above the classroom criterion on four consecutive days. During the second maintenance phase, Ms. Allen’s praise was somewhat more variable around the criterion with a slight decreasing trend (-0.39).

Ms. Dean praised at low rates during baseline ($M = 1.86$; range, 0.78 – 3.14), with a slightly decreasing slope (-0.13). In training probes, she gradually increased her praise, meeting the fluency goal on the last day of training. Her praise rate in the classroom during this time followed the same pattern, with an increasing trend (0.34), and the final data point above the classroom criterion. During the first eight days of maintenance, her
praise rate gradually increased; but the on the last four of five days, her praise rate was below the classroom criterion. Ms. Dean received performance feedback on all teaching skills. During this time, the slope of her praise rate increased dramatically (1.48), with three consecutive days of praise well above the classroom criterion. During the final maintenance phase, her praise remained above the criterion for all three days.

Ms. Jones praised students infrequently during baseline ($M = 0.35$; range, 0.00 – 1.68). On the single fluency training probe that she completed, her praise was similarly low. Over the course of fluency training, her praise rate had an increasing slope (0.98), but the level remained below the classroom criterion ($M = 1.60$; range, 0.59 – 2.54). On the first two days of maintenance, Ms. Jones’ praise rate dropped to baseline levels. However, over the course of the maintenance phase, the slope of her praise rate increased gradually (0.30). The level remained below the classroom criterion ($M = 2.30$; range, 0.98 – 3.48), so we initiated performance feedback on all skills with Ms. Jones. During performance feedback, the level of her praise decreased slightly ($M = 1.93$; range, 1.12 – 2.73), with an increasing slope (1.61). We concluded performance feedback when tier II groups were discontinued at the school. Ms. Jones never reached the classroom performance target for praise rate.

Ms. Lewis had a somewhat variable praise rate during baseline with all but two data points below the classroom criterion ($M = 2.56$; range, 0.88 – 4.33). During all training probes, Ms. Lewis’ praise rate was above the fluency goal (six praise statements per minute). With the introduction of practice in praise rate on the second day of fluency training, her praise rate increased immediately to levels above the classroom criterion (four per minute). The slope of her praise rate increased during this phase also (1.20).
During maintenance, Ms. Lewis’ praise rate remained at or above the training criterion $(M = 6.00; \text{range}, 3.99 – 7.19)$, but with a decreasing slope (-0.28). We gave Ms. Lewis performance feedback because her presentation rate was below the classroom criterion. During this phase, Ms. Lewis’ praise rate decreased, but remained near or above the classroom criterion $(M = 4.12; \text{range}, 3.75 – 4.49)$ with a slightly increasing slope (0.19).

We discontinued performance feedback when tier II groups ended.

Ms. Tate praised students infrequently during baseline $(M = 0.40; \text{range}, 0.00 – 1.34)$. During training probes, her praise rate increased, with the last two days at the fluency goal. On the first day of fluency training, Ms. Tate’s praise rate did not increase much in the classroom; however, once fluency practice on praise was initiated, her praise rate increased immediately and met the classroom criterion. The slope for her praise rate increased from -0.01 in baseline to 1.26 during fluency training. Her praise rate continued to increase during maintenance (0.25), with six of seven days above the classroom criterion $(M = 5.42; \text{range}, 3.55 – 8.15)$.

All paraprofessionals demonstrated levels of praise below the classroom criterion during baseline (all means below 2.6 praise statements per minute). Ms. Allen’s praise rate was slightly increasing, while the others had flat or slightly decreasing slopes in baseline. Ms. Allen, Ms. Dean and Ms. Lewis had variable data patterns in baseline, while Ms. Jones and Ms. Tate had very low variability in their praise rates. During fluency training, all paraprofessionals increased the level and slope of their praise rates. Three of the five paraprofessionals praised at rates higher than the classroom criterion on at least two of the fluency training days. The same paraprofessionals also maintained praise rates above the classroom criterion for most days of maintenance.
Paraprofessionals’ praise rates maintained baseline levels when fluency training was introduced in prior legs, except for Ms. Lewis’. When Ms. Jones began the fluency training phase, Ms. Lewis’ praise rate increased, but with a decreasing slope on the last three days of baseline.

**Error Corrections**

Figure 3 shows the percentage of error correction steps that paraprofessionals accurately completed when students made word attack errors. The target criterion for error correction accuracy (shown by horizontal dashed lines) was 95% for both the classroom and training settings. The vertical dashed line in each fluency training phase is just before the session when specific training on error corrections was introduced to paraprofessionals. Dark circles represent data points in which five or more errors were corrected, sometimes over more than one session. Open circles depict data points based on fewer than five errors; this occurred just before phase changes because we did not aggregate data across phase changes. Dark or open circles with a white plus sign indicate overlapping data points for training and classroom observations. Table 7 shows the percentage of error correction attempts (i.e., the percentage of all student errors that paraprofessionals started to correct) for each individual across each phase. Table 8 shows the means, ranges and slopes of error correction accuracy (i.e., of the error correction attempts, the percentage of correction steps completed accurately). Ms. Allen attempted to correct 85.71% of student errors during baseline. Of these error correction attempts, she accurately completed 53.64% of the steps (range, 33.33% – 69.44%). Within the
Figure 3. Paraprofessionals’ error correction accuracy.
Table 7

*Paraprofessionals’ Percentage of Error Correction Attempts Made for All Student Errors*

<table>
<thead>
<tr>
<th>Paraprofessional</th>
<th>Baseline</th>
<th>Fluency</th>
<th>Performance</th>
<th>Maintenance</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ms. Allen</td>
<td>85.71</td>
<td>94.44</td>
<td>72.22</td>
<td>80.65</td>
<td>94.44</td>
</tr>
<tr>
<td>Ms. Dean</td>
<td>73.68</td>
<td>78.95</td>
<td>82.54</td>
<td>93.75</td>
<td>100.00</td>
</tr>
<tr>
<td>Ms. Jones</td>
<td>70.00</td>
<td>63.64</td>
<td>50.00</td>
<td>71.43</td>
<td></td>
</tr>
<tr>
<td>Ms. Lewis</td>
<td>76.67</td>
<td>93.33</td>
<td>94.21</td>
<td>91.30</td>
<td></td>
</tr>
<tr>
<td>Ms. Tate</td>
<td>87.72</td>
<td>92.86</td>
<td>87.50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Entire fluency training phase, Ms. Allen attempted to correct 94.44% of student errors. During training probes prior to the introduction of error correction training, Ms. Allen’s error correction accuracy increased in level, but decreased in slope. Before error correction training was introduced, Ms. Allen’s error correction accuracy in the classroom increased slightly over baseline. When error correction training began, Ms. Allen’s error correction accuracy in training probes decreased slightly. The rate then increased and was above the training criterion for the last two days of training. After error correction training started, Ms. Allen’s classroom error correction accuracy increased in level and slope, with the last day of training above the fluency goal. During maintenance, Ms. Allen attempted to correct 72.22% of student errors. When correcting these errors, Ms. Allen’s error correction accuracy returned to baseline levels ($M = 57.14\%$; range, $50.00\% – 64.29\%$).
Table 8

Paraprofessionals’ Error Correction Accuracy: Phase Means, Ranges, and Slopes

<table>
<thead>
<tr>
<th>Paraprofessional</th>
<th>Baseline</th>
<th>Fluency Training</th>
<th>Maintenance</th>
<th>Performance Feedback</th>
<th>Maintenance 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (Range)</td>
<td>Slope</td>
<td>Mean (Range)</td>
<td>Slope</td>
<td>Mean (Range)</td>
</tr>
<tr>
<td>Ms Allen</td>
<td>53.64 (33.33-69.44)</td>
<td>1.08</td>
<td>81.40 (70.83-99.00)</td>
<td>6.96</td>
<td>57.14 (50.00-64.29)</td>
</tr>
<tr>
<td>Ms Dean</td>
<td>40.00 (36.67-44.44)</td>
<td>-2.62</td>
<td>68.47 (43.33-88.89)</td>
<td>16.17</td>
<td>67.53 (59.26-77.75)</td>
</tr>
<tr>
<td>Ms Jones</td>
<td>19.58 (12.50-23.81)</td>
<td>0.90</td>
<td>80.83 (70.00-91.67)</td>
<td>-11.83</td>
<td>63.33 (50.00-83.33)</td>
</tr>
<tr>
<td>Ms Lewis</td>
<td>49.40 (47.62-50.00)</td>
<td>0.24</td>
<td>67.62 (50.00-76.67)</td>
<td>7.55</td>
<td>78.82 (72.22-85.42)</td>
</tr>
<tr>
<td>Ms Tate</td>
<td>62.08 (53.66-66.67)</td>
<td>-0.19</td>
<td>83.33 (56.67-99.00)</td>
<td>33.33</td>
<td>92.37 (73.33-100.00)</td>
</tr>
</tbody>
</table>

*Indicates only one data point in the phase.
During performance feedback, Ms. Allen attempted to correct 80.65% of student errors. During this time, Ms. Allen increased her error correction accuracy ($M = 91.67\%$; range, $86.67\% - 96.67\%$), meeting the criterion on the last two days. During the second maintenance phase, Ms. Allen increased the percent of errors she attempted to correct ($M = 94.44\%$). Ms. Allen’s error correction accuracy decreased somewhat during the second maintenance phase, but remained well above baseline levels ($M = 88.89\%$; range, $83.33\% - 100\%$).

Ms. Dean attempted to correct a mean of 73.68% of student errors during baseline. Of these errors, Ms. Dean’s mean error correction accuracy was 40.00% (range, $36.67\% - 44.44\%$) with a decreasing slope (-2.62). During the entire fluency training phase, Ms. Dean attempted to correct 78.95% of student errors. Before error corrections were introduced in fluency training, Ms. Dean’s error correction accuracy in training probes was below 70%. In the classroom before error correction training was introduced, Ms. Dean’s error correction accuracy remained below 60% correct. After error corrections were introduced in training, Ms. Dean’s training probe accuracy was above 80%, with two data points at 100%. In classroom observations during error correction training, Ms. Dean completed more than 80% of steps accurately for two days. During maintenance, Ms. Dean attempted to correct 82.54% of student errors. When maintenance started, Ms. Dean’s error correction accuracy fell to a mean of 67.53% correct (range, $59.26\% - 77.78\%$), which was still above all baseline points. During performance feedback, Ms. Dean attempted to correct 93.75% of student errors. When correcting these errors, Ms. Dean’s error correction accuracy increased slightly ($M = 73.15\%$; range, $63.89\% - 83.33\%$), with an increasing slope (3.77). During the second
maintenance phase, Ms. Dean attempted to correct all student errors. Ms. Dean accurately presented about 85% of steps correctly during this phase with almost no trend. Throughout the study, Ms. Dean’s error correction accuracy did not meet the classroom criterion; however, all data points after the introduction of fluency training were above the highest baseline data point.

During baseline, Ms. Jones attempted to correct a mean of 70.00% of student errors. Ms. Jones’ mean error correction accuracy was 19.58% (range, 12.50% – 23.81%), with a slightly increasing trend (0.90) during baseline. During the fluency training phase, Ms. Jones attempted to correct 63.64% of student errors. Prior to error correction training within the fluency training phase, Ms. Jones’ single training probe error correction accuracy was 50.00% correct. During a classroom observation the next day, Ms. Jones accurately completed 91.67% of steps correctly. After error correction training was introduced, Ms. Jones’ error correction accuracy dropped to 70.00% during one observation. She was unable to complete any training probes during this part of the phase. During maintenance, Ms. Jones attempted to correct 50.00% of student errors. Of these error correction attempts, Ms. Jones’ mean error correction accuracy was 63.33% (range, 50.00% – 83.33%), with an increasing slope (5.00). All maintenance data points were higher than all baseline data points. During performance feedback, Ms. Jones attempted to correct 71.43% of errors. Ms. Jones accurately presented 70.00% of error correction steps during this phase.

Ms. Lewis attempted to correct 76.67% of student errors during baseline. For the errors that she attempted to correct, she accurately completed a mean of 49.40% error correction steps (range, 47.62% - 50.00%), with almost no trend. During the fluency
training phase, Ms. Lewis attempted to correct 93.33% of student errors. During the first training probe, before error corrections were introduced, Ms. Lewis’ error correction accuracy was 50.00%, as was the first classroom observation during training. After error correction training was introduced, training probes increased dramatically, with all data points above 75% correct. During error correction training, Ms. Lewis’ error correction accuracy in classroom observations increased to 76%. During maintenance, Ms. Lewis attempted to correct 94.21% of student errors. Of these attempts, Ms. Lewis’ error correction accuracy remained above baseline levels \((M = 78.82\%; \text{range, } 72.22\% – 85.42\%)\) with a decreasing slope (-13.20). Ms. Lewis attempted to correct 91.30% of student errors during performance feedback. She accurately completed a mean of 86.34% error correction steps (range, 69.44% – 95.83%), with an increasing slope (12.16) during this phase.

Ms. Tate attempted to correct 87.72% of student errors during baseline. In this phase, Ms. Tate’s error correction accuracy was stable and flat \((M = 62.08\%; \text{range, } 53.66\% – 66.67\%)\). During the fluency training phase, Ms. Tate attempted to correct 92.86% of student errors. When fluency training was introduced, but before error corrections were covered, Ms. Tate’s error correction accuracy was 66.67% during the training probe and the classroom observation. After error corrections were introduced in fluency training, Ms. Tate’s error correction accuracy in training probes was 100%. The single classroom observation during this phase was also at 100% correct. During maintenance, Ms. Tate attempted to correct 87.50% of student errors. In this phase, the first error correction accuracy data point was lower than during training, but Ms. Tate’s
error correction accuracy increased to the classroom criterion (95%) and remained above it for the final three data points.

When students made an error, paraprofessionals corrected less than 88% (range, 70.00 to 87.72%) of those errors during baseline (Table 5). Of these error correction attempts, all paraprofessionals had low accuracy (below 70% correct). Four of five paraprofessionals had flat or decreasing baselines for error correction accuracy. When fluency training was introduced, all but one paraprofessional increased the percentage of errors that they attempted to correct. During the first two sessions of fluency training, error correction training was not specifically addressed. However, two of the paraprofessionals experienced a small increase in the level (Ms. Allen) or slope (Ms. Dean) of their error correction accuracy. Ms. Lewis and Ms. Tate continued to correct errors with similar accuracy as during baseline sessions. When error correction training was introduced on the third day of fluency training, four of five paraprofessionals showed a distinct level or slope increase in error correction accuracy. All of these data points were at 75% accuracy or higher. Two paraprofessionals maintained levels of error correction accuracy after fluency training ended. One paraprofessional met the classroom criterion (95% accuracy) on three consecutive days. All paraprofessionals showed substantially higher error correction accuracy during their final maintenance phase than they had during the initial baseline.
Figure 4. Paraprofessionals’ positive to negative comments ratio (percent positive).
### Table 9

**Paraprofessionals Positive to Negative Comments: Phase Means, Ranges, and Slope**

<table>
<thead>
<tr>
<th>Paraprofessional</th>
<th>Baseline</th>
<th>Fluency Training</th>
<th>Maintenance</th>
<th>Performance Feedback</th>
<th>Maintenance 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>(Range)</td>
<td>(Range)</td>
<td>(Range)</td>
<td>(Range)</td>
<td>(Range)</td>
</tr>
<tr>
<td></td>
<td>Slope</td>
<td>Slope</td>
<td>Slope</td>
<td>Slope</td>
<td>Slope</td>
</tr>
<tr>
<td>Ms. Allen</td>
<td>97.13</td>
<td>97.02</td>
<td>99.59</td>
<td>97.52</td>
<td>99.07</td>
</tr>
<tr>
<td></td>
<td>(90.48-100.00)</td>
<td>(93.33-100.00)</td>
<td>(97.56-100.00)</td>
<td>(87.59-100.00)</td>
<td>(94.44-100.00)</td>
</tr>
<tr>
<td>Ms. Dean</td>
<td>90.08</td>
<td>98.33</td>
<td>97.59</td>
<td>97.46</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>(57.4-100.00)</td>
<td>(93.33-100.00)</td>
<td>-0.34</td>
<td>1.23</td>
<td>0.00</td>
</tr>
<tr>
<td>Ms. Jones</td>
<td>55.33</td>
<td>95.24</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>(0.00-100.00)</td>
<td>(0.00-100.00)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Ms. Lewis</td>
<td>95.11</td>
<td>99.31</td>
<td>96.63</td>
<td>93.28</td>
<td>-3.92</td>
</tr>
<tr>
<td></td>
<td>(84.62-100.00)</td>
<td>(96.55-100.00)</td>
<td>-0.46</td>
<td>(94.74-97.73)</td>
<td>(88.00-96.00)</td>
</tr>
<tr>
<td>Ms. Tete</td>
<td>47.22</td>
<td>106.00</td>
<td>97.30</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00-100.00)</td>
<td>(94.12-100.00)</td>
<td>0.00</td>
<td>0.28</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Data represent percentage of all comments that were positive.
Positive to Negative Comments Ratio

Figure 4 presents the percentage of positive comments (out of all positive and negative comments) paraprofessionals made to students in the classroom setting during word attack. The target criterion for this ratio was 80% positive statements. Table 9 shows the summary statistics for each phase.

Three of the five paraprofessionals demonstrated high ratios of positive comments to students throughout all phases. Ms. Allen, Ms. Dean, and Ms. Lewis maintained average levels of positive feedback above 80% throughout the study. Across these three paraprofessionals, there was only a single data point below 80%.

Ms. Jones and Ms. Tate had highly variable baseline levels of positive comments to students; both ranged from 0% to 100%. During fluency training, both paraprofessionals increased their positive comments ratio to above the criterion, while decreasing the variability, which they maintained for the rest of the study. Ms. Jones’ ratio ranged from 86% to 100% positive through the end of the study. Ms. Tate’s positive comments ratio ranged from 94% to 100% positive during fluency training until the end of the study. This decrease in variability appeared to be due to increases in praise as well as decreases in negative comments to students.

Students’ Behavior

On Task

Figure 5 shows the percentage of 10-second intervals in which students displayed on-task behavior for the whole interval. Table 10 shows the summary data for each phase.
Figure 5. Students’ on-task behavior (percentage of 10-s intervals) during word attack.
Table 10

*Students’ On-task Behavior: Percentage of 10-s Intervals*

<table>
<thead>
<tr>
<th>Student</th>
<th>Baseline Mean (Range)</th>
<th>Baseline Slope</th>
<th>Fluency Training Mean (Range)</th>
<th>Fluency Training Slope</th>
<th>Maintenance Mean (Range)</th>
<th>Maintenance Slope</th>
<th>Performance Feedback Mean (Range)</th>
<th>Performance Feedback Slope</th>
<th>Maintenance 2 Mean (Range)</th>
<th>Maintenance 2 Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alex</td>
<td>19.89 (-3.55)</td>
<td>15.56 (-13.33)</td>
<td>20.00 (6.32)</td>
<td></td>
<td>21.74 (3.50)</td>
<td></td>
<td>12.62 (-4.00)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Claire</td>
<td>81.14 (-8.04)</td>
<td>92.65 (5.49)</td>
<td>83.39 (-0.11)</td>
<td></td>
<td>93.28 (0.98)</td>
<td></td>
<td>29.06 (-11.45)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jay</td>
<td>31.34 (-5.61)</td>
<td>.876* n/a</td>
<td>14.96 (4.70)</td>
<td></td>
<td>52.98 (51.29)</td>
<td></td>
<td>23.33-34.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luke</td>
<td>27.50 (-9.36)</td>
<td>53.50 (1.31)</td>
<td>37.48 (-9.30)</td>
<td></td>
<td>35.55 (-20.00)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Todd</td>
<td>27.60 (0.39)</td>
<td>57.86 (4.39)</td>
<td>28.33 (-12.97)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Indicates only one data point in the phase.
Alex’s on-task behavior was highly variable throughout the phases with many sessions below 30%. During baseline, Alex’s on-task behavior was highly variable (range, 0.00% – 69.23%). When his teacher, Ms. Allen, entered fluency training, Alex’s had a similarly large range (0.00% – 46.67%). During maintenance, the pattern continued, except for a high point on the last day of the phase. When we provided Ms. Allen performance feedback, Alex’s on-task behavior increased gradually (3.50), but remained low and variable (\(M = 21.14\%\); range, 0.00% – 76.00%). During the second maintenance phase, Alex’s on-task behavior decreased again (-4.00).

Claire’s on-task behavior was generally high throughout the study. In baseline observations, on-task behavior started high and decreased on the last two days (\(M = 81.14\%\); range, 45.45% – 100%). When Ms. Dean participated in fluency training, Claire’s on-task behavior increased again with less variability (\(M = 92.65\%\); range, 77.27% – 100%). During maintenance, Claire’s on-task behavior was fairly high with increased variability (\(M = 83.39\%\); range, 63.33% – 100%). During performance feedback for Ms. Dean, the variability in Claire’s on-task behavior decreased again (\(M = 93.28\%\); range, 88.00% – 96.15%). With the second maintenance phase, Claire’s on-task behavior decreased dramatically (\(M = 29.06\%\); 23.33% – 34.78%).

Jay’s on-task behavior was fairly low and variable during the study. Over the course of the baseline phase, Jay’s on-task behavior was highly variable (\(M = 31.34\%\); 0.00% – 91.67%). When his teacher, Ms. Jones, participated in fluency training, Jay’s single on-task behavior data point was low (18.76%). During maintenance, Jay’s on-task behavior was low (\(M = 14.96\%\); range, 0.00% – 44.44%), but increasing in the last two
data points. During the two sessions in which we gave Ms. Jones performance feedback, Jay’s on-task behavior increased in slope (51.29).

Luke’s on-task behavior was low and variable during baseline observations ($M = 27.50\%$; range, 0.00% – 78.95%). During Ms. Lewis’ fluency training, Luke’s on-task behavior increased in level ($M = 63.50$) and decreased substantially in variability (range, 55.56% – 68.42%). Luke’s on-task behavior decreased during maintenance (-9.30). During performance feedback for Ms. Lewis, Luke’s on-task behavior decreased again (-20.00).

Todd’s on-task behavior was variable throughout the study. During baseline, his on-task behavior was fairly low and highly variable ($M = 27.60\%$; range, 0.00% – 65.22%). When Ms. Tate participated in fluency training, Todd’s on-task behavior increased in level and decreased in variability ($M = 67.86\%$; range, 54.55% – 85.71%). During maintenance, Todd’s on-task behavior started very high and decreased dramatically ($M = 28.33\%$; range, 0.00 – 92.31; slope, -12.97).

All five students had highly variable on-task behavior during baseline observations. When fluency training was introduced to paraprofessionals, Claire, Luke and Todd’s on-task behavior increased in level and slope, while decreasing in variability. These effects did not maintain in later phases.
Figure 6. Students’ first time correct responses during word reading.
Table 11

Students’ Word Reading Accuracy: Percentage First Time Correct

<table>
<thead>
<tr>
<th>Student</th>
<th>Baseline Mean (Range)</th>
<th>Baseline Slope</th>
<th>Fluency Training Mean (Range)</th>
<th>Fluency Training Slope</th>
<th>Maintenance Mean (Range)</th>
<th>Maintenance Slope</th>
<th>Performance Feedback Mean (Range)</th>
<th>Performance Feedback Slope</th>
<th>Maintenance 2 Mean (Range)</th>
<th>Maintenance 2 Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alex</td>
<td>83.48 (86.67-93.33)</td>
<td>-3.34</td>
<td>79.58 (75.93-84.09)</td>
<td>-2.40</td>
<td>88.52 (75.07-97.14)</td>
<td>0.27</td>
<td>88.76 (32.05-96.00)</td>
<td>0.45</td>
<td>92.17 (86.67-97.83)</td>
<td>-0.58</td>
</tr>
<tr>
<td>Clare</td>
<td>78.27 (63.04-90.57)</td>
<td>4.71</td>
<td>82.27 (77.78-85.42)</td>
<td>-0.91</td>
<td>84.46 (75.00-91.89)</td>
<td>0.03</td>
<td>91.38 (33.87-97.50)</td>
<td>0.52</td>
<td>83.78* (n/a)</td>
<td>n/a</td>
</tr>
<tr>
<td>Jay</td>
<td>36.73 (80.00-93.33)</td>
<td>-0.32</td>
<td>92.31* (72.50-90.50)</td>
<td>n/a</td>
<td>82.34 (72.50-90.50)</td>
<td>0.03</td>
<td>88.28 (37.88-88.08)</td>
<td>0.80</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Luke</td>
<td>78.30 (25.10-100.00)</td>
<td>0.22</td>
<td>90.58 (78.57-100.00)</td>
<td>-0.90</td>
<td>77.46 (76.73-87.50)</td>
<td>5.43</td>
<td>69.32 (58.82-78.00)</td>
<td>-3.3</td>
<td>(59.38-89.74)</td>
<td>(75.55-87.50)</td>
</tr>
<tr>
<td>Tocd</td>
<td>31.38 (59.38-89.74)</td>
<td>0.38</td>
<td>85.65 (73.55-92.50)</td>
<td>-8.69</td>
<td>84.39 (75.55-87.50)</td>
<td>-0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Indicates only one data point in the phase.
Word Reading Accuracy: First-time Correct

Figure 6 shows the students’ accuracy on reading words the first time each word was encountered in a particular lesson. Table 11 shows the summary data for each phase. We did not detect any patterns in the data corresponding to the introduction of different interventions for the paraprofessionals. For four of the five students, the means for each phase were roughly between 80 and 90% correct with little to moderate variability. During baseline, Luke’s accuracy was variable (range, 25.00% - 100.00%) with a slightly increasing slope. During fluency training, variability decreased. During maintenance, the slope increased, followed by a decreasing slope in performance feedback.

Social Validity

Paraprofessional Teaching Skills

The DI experts’ ratings were compared to the criteria used in this study for paraprofessionals’ presentation rates, praise rates and overall session quality (based on presentation rate and praise rate criteria only) for select sessions. Table 12 shows a 2x2 contingency table making this comparison. Percent agreement for presentation rate was 75% (15/20) of ratings. Disagreements for presentation rate occurred in 5 cases in which the paraprofessional had a presentation rate of 15 presentations per minute or higher, but the expert rated the presentation rate as “OK” or “poor.” There were no cases in which the paraprofessional’s presentation fell below the criterion and the expert rated it as “good” or “excellent.” For praise rate, agreement was 80% (16/20). Disagreements
Table 12

*Social Validity of Paraprofessionals’ Presentation Rate, Praise Rate, and Overall Session Quality*

<table>
<thead>
<tr>
<th>Paraprofessional Skill</th>
<th>Criterion Categories</th>
<th>DI Expert Ratings</th>
<th>Percent Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Excellent or Good</td>
<td>OK or Poor</td>
</tr>
<tr>
<td>Presentation Rate</td>
<td>≥ 15 per min</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>&lt; 15 per min</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Praise Rate</td>
<td>≥ 4 per min</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>&lt; 4 per min</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Overall Session</td>
<td>Presentation ≥ 15 And Praise ≥ 4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Presentation &lt; 15 Or Praise &lt; 4</td>
<td>0</td>
<td>14</td>
</tr>
</tbody>
</table>

occurred in the same pattern as with presentation rate: four cases in which the paraprofessionals’ praise rate was at or above the criterion were rated as “OK” or “poor” by the expert. No cases in which the paraprofessional praised less than four times per minute were judged as “good” or “excellent” by the expert. This pattern also maintained for the overall session ratings. Agreement on the overall session ratings was 90% (18/20).
On the two disagreements, experts rated sessions that included presentation and praise rates at or above the criterion as “OK” or “poor.” No sessions with praise or presentation rates below the criterion were rated as “good” or “excellent” by the experts.

Feasibility of Training

The two district reading coordinators completed their questionnaire together. They rated the feasibility of the training highly, agreeing strongly (the highest rating) with five statements (amount of time for training was reasonable; amount of money required was reasonable; the training was simple enough to be used by district staff; paraprofessionals enjoyed training; we are likely to use the training in the future). The coordinators agreed with the sixth statement, “We would be able to provide more than 5 hours of this type of training to paraprofessionals in the future.” This statement was added when the effects of the training were apparent. To the open-ended question, “Are there any barriers to the use of this training in the future?” they answered, “The only barrier is fitting everything in. It is valuable enough that it should take priority over many things that fill our days. We plan to make this type of training a priority.”

Acceptability of Training

The paraprofessional participants rated the training on a likert scale from 1 (strongly disagree) to 5 (strongly agree). See table 13. The mean rating for all questions was 4.45 (range, 2 to 5), favoring approval of the training. One participant rated the reverse-coded item (training had little effect on my teaching) as “agree”, which accounted for the single rating of 2. This rating was not consistent with her other answers, which rated the intervention highly (all 4 or 5).
### Paraprofessionals’ Ratings of Fluency Training

<table>
<thead>
<tr>
<th>Question Topic</th>
<th>Mean Rating (Range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classroom teaching improved</td>
<td>4.6 (4-5)</td>
</tr>
<tr>
<td>I enjoyed the training</td>
<td>4.6 (4-5)</td>
</tr>
<tr>
<td>Students read better</td>
<td>4.4 (4-5)</td>
</tr>
<tr>
<td>Students behaved better</td>
<td>4.4 (4-5)</td>
</tr>
<tr>
<td>I presented items faster and more effectively</td>
<td>4.4 (4-5)</td>
</tr>
<tr>
<td>I praised more</td>
<td>4.8 (4-5)</td>
</tr>
<tr>
<td>I provided more positives than negatives</td>
<td>4.4 (4-5)</td>
</tr>
<tr>
<td>Error corrections were more accurate</td>
<td>4.4 (4-5)</td>
</tr>
<tr>
<td>Practicing at a fast rate helped</td>
<td>4.6 (4-5)</td>
</tr>
<tr>
<td>I would recommend training to others</td>
<td>4.8 (4-5)</td>
</tr>
<tr>
<td>Training had little effect (reverse coded)</td>
<td>3.6 (2-5)</td>
</tr>
<tr>
<td>Overall, training was worthwhile</td>
<td>4.4 (4-5)</td>
</tr>
</tbody>
</table>
CHAPTER V

DISCUSSION

The primary purpose of this study was to evaluate the effects fluency training in a Direct Instruction reading program had on paraprofessionals’ intervention fidelity, including presentation rate, praise rate, error correction accuracy, and positive to negative comments ratio. A secondary purpose of this study was to evaluate the effects of this fluency training on the students’ on-task and word reading accuracy. In general, paraprofessionals increased their praise rates, error correction accuracy and decreased variability in their positive to negative comments ratio with fluency training. The effects were good for some paraprofessionals’ presentation rates, but there were not enough baselines that were low enough for it to be tested across all participants.

Dependent Variables

Paraprofessional Behaviors

We observed four paraprofessional behaviors, which may be considered central to intervention fidelity for Direct Instruction programs. Accurate presentation rate, praise rate, error correction accuracy and positive to negative comments ratio were assessed during baseline, fluency training, and maintenance for all participants. Four participants received performance feedback, and we observed two participants during a second maintenance phase. An examination of the experimental control for each behavior follows.
In single subject research experimental control is assessed by examining data patterns over time and across replications (e.g., legs of a multiple baseline design) (Kazdin, 1982; Horner et al., 2005). A clear change in the pattern of the behavior should occur at three points in the data, with at least five data points per phase. Level, slope, variability, and latency of effect should be considered.

For presentation rate, three participants had levels lower than the criterion during baseline (Ms. Allen, Ms. Dean, Ms. Jones). However, Ms. Allen’s baseline was increasing in the direction of the predicted change for the fluency intervention, which makes the level change in that direction less convincing. In addition, we introduced fluency training to Ms. Allen and Ms. Dean at the same time, so they should not be considered completely independent legs of the multiple baseline. Experimental control can be identified in two legs of the multiple baseline (Ms. Dean and Ms. Jones) for the introduction of fluency training. For Ms. Dean and Ms. Jones, fluency training had a dramatic effect on their presentation rates. These effects were maintained for the rest of the study. For individuals with low presentation rates, it is important to show that they can improve their rates to the criterion with a relatively brief intervention. Ms. Allen’s presentation rate improved with fluency training, with three of four data points at or above the criterion during fluency training. However, the increase in her baseline made the change less dramatic.

We were unable to definitively test the effects of performance feedback on presentation rate, because only one paraprofessional demonstrated a clear need for it in maintenance (Ms. Allen). When Ms. Allen received performance feedback, her presentation rate increased immediately and dramatically. Ms. Lewis’ presentation rate
was near the criterion during maintenance, so we started performance feedback with her. Her presentation rate improved some with performance feedback, but we did not have enough time to see the effects over many days. With additional performance feedback, her presentation rate may have increased in level and stability. To our knowledge, no previous studies have tested the effects of training and performance feedback on instructor presentation rate. This study suggests that this combination may be effective, but more research is needed.

Experimental control for praise rate is demonstrated in comparing the slopes and levels across phases. On the first day of fluency training, praise was discussed, but participants did not practice providing praise. In the classroom, after the first day of training, each paraprofessional praised at similar rates to baseline. On the second and following days of fluency training, we included fluency practice on praise. On subsequent days in the training phase, participants’ classroom praise rates increased over time. All paraprofessionals increased the slope of their praise in fluency training. These increasing slopes lead to distinctly higher maintenance levels for three of the participants (Ms. Allen, Ms. Lewis, Ms. Tate), and somewhat higher maintenance levels for another participant (Ms. Jones). If we had elided the training phase and just presented baseline and maintenance data, as some studies do with typical training (e.g., Bingham, 2005; Sarokoff & Sturmey, 2004; Schepis, Ownbey, Parsons, & Reid, 2000), we would see a large level change in praise rates for Ms. Allen, Ms. Lewis, and Ms. Tate. However, the gradual increase in praise rates observed from baseline to training suggests that this skill increases steadily but gradually during training, rather than with a large immediate level change. When training sessions were introduced to Ms. Jones, Ms. Lewis’ baseline praise
rate increased to near criterion levels, albeit with a decreasing slope in the last three data points. Although we attempted to avoid this, Ms. Lewis might have heard that training included a focus on praise, since Ms. Lewis and Ms. Jones taught in the same room (they could not see or hear each other during teaching). At the start of the baseline phase, Ms. Lewis praised students at rates near the criterion. If she heard that we emphasized praise, this may have been motivation enough to increase her praise rate during baseline.

However, the dramatic level and slope change that occurred when we introduced fluency training on praise for Ms. Lewis suggests that the fluency training was effective in increasing her praise rate. The data on fluency training for praise rate are promising insofar as all participants increased their praise rate slopes from baseline to fluency training. Two participants received performance feedback primarily due to lower praise rates (Ms. Dean and Ms. Jones). Performance feedback appeared to be effective for Ms. Dean but not for Ms. Jones, although this phase was cut short due to scheduling constraints. Thus, the effects of performance feedback on praise rates could not be strongly tested in this study.

Paraprofessionals showed overall low rates of error correction accuracy in baseline, with increases during fluency training. Experimental control for error correction accuracy was demonstrated when taking into account the introduction of error correction accuracy on the third day of during fluency training. After error correction training was introduced during fluency training, all participants experienced a large level change from baseline. When error correction training was introduced to Ms. Allen and Ms. Dean, the error correction data remained low and stable across the other baselines. This pattern maintained when we introduced error correction training to each subsequent participant.
However, the small number of data points (one or two) for the error correction portion of fluency training weakens this demonstration of control slightly. Four of the five paraprofessionals maintained levels above baseline when fluency training was discontinued, although two had lower accuracy than during training. Despite the general improvements for all paraprofessionals, only Ms. Tate achieved the criterion of three consecutive data points above 95% accuracy during classroom observations. Possibly, the criterion of 95% accuracy on error correction steps is too high, or more intense training would be needed to help paraprofessionals’ achieve the criterion. Researchers have not determined an optimal accuracy level for error corrections, although error correction accuracy appears to be important to student outcomes (Carlson & Francis, 2002).

Fluency training appeared to reduce variability in positive to negative comments ratios for those participants who showed high variability in baseline. Ms. Lewis had high ratios of positive to negative comments with some variability during baseline while Ms. Jones and Ms. Tate had extremely variable ratios of positive to negative comments during baseline. When fluency training was introduced to each of these three paraprofessionals, variability decreased dramatically. Further, when training was introduced to any one of the paraprofessionals, performance of the others was unaffected. Low variability continued into maintenance for Ms. Lewis and Ms. Tate, and into performance feedback for Ms. Lewis. Experimental control was demonstrated across Ms. Jones, Ms. Lewis and Ms. Tate. Different variables may have contributed to this decrease in variability, such as general increases in praise, which may have displaced negative comments, or awareness from the training that negative comments may maintain problem behavior. In our
observations, paraprofessionals’ decreases in variability were due to a combination of increases in praise and decreases in negative comments to students.

**Student Behaviors**

Consistent patterns were difficult to detect in student on-task behavior. For example, Alex’s on-task behavior was variable and low throughout the study. Claire’s variability decreased during the paraprofessionals’ fluency training, increased again in maintenance and decreased during performance feedback. The variability in Jay’s on-task behavior during baseline decreased during maintenance, but was at a low level. His on-task behavior increased dramatically at the end of maintenance and during performance feedback. Luke’s and Todd’s behavior had the most distinct level changes from baseline to fluency training, but there was still a certain amount of overlap between the phases, weakening this conclusion. In addition, both students’ on-task behavior returned to low baseline levels during maintenance. When fluency training was introduced to Ms. Lewis, Luke’s level increased from baseline, while his variability decreased. Todd’s level and slope of on-task behavior increased with fluency training for Ms. Tate, also with overlap in the data patterns. For these students, the on-task behavior patterns achieved during fluency training did not maintain during the maintenance phase. No conclusions can be made about the effects of performance feedback on student on-task behavior.

All students except Jay ended the study with low and decreasing amounts of on-task behavior. Since we observed groups up until the last day of reading instruction for tier II, this decrease may have been partly due to the end of the groups.
Experimental control was not established for any of the legs of the study for students’ word reading accuracy. All students had relatively high levels of word reading accuracy across the study. Luke had lower levels during maintenance and performance feedback than baseline and fluency training, but it is not clear what contributed to this change.

Overall, we were not able to detect clear changes in student behavior patterns after training for instructional staff, although this is certainly the end goal. Changes in student behavior may be more temporally remote and occur after a longer delay with more sustained use of the techniques taught in this training. Also, the intensity of the training may have been inadequate to improve paraprofessional behavior enough to see corresponding changes in student behavior. Studies on intervention fidelity that have shown stronger effects on instructional behavior have sometimes shown more diffuse or variable effects on student behavior (Holcombe et al., 1994; Jones et al., 1997; Noell et al., 1997; Sterling-Turner et al., 2002). Increasing intervention fidelity frequently leads to changes in student behavior, but not always in an unequivocal manner (Perepletchikova & Kazdin, 2005). Future research is warranted to isolate the variables that make it more likely that student effects will be observed when intervention fidelity is improved.

In previous research, larger changes in instructional behaviors such as praise and presentation rates typically corresponded with larger changes in student behavior such as on-task and word reading accuracy (e.g., Carnine, 1976; Thomas et al., 1968). In addition, changes in paraprofessional behaviors seen in this study may not have reflected changes in the critical features that affect students’ behaviors. For example, Carnine found that students’ correct responding increased and off-task behavior decreased with a
faster presentation rate during reading instruction. To control the pace of instruction during the slower presentation, the instructors added a 5 s pause after each reading item, resulting in a presentation rate of about four items per minute. During the phase with faster presentation, the instructors minimized pauses between items, with an average presentation rate of 12 items per minute. In the current study, few paraprofessionals actually introduced pauses in the instruction. Lower presentation rates were often due to incorrect presentations (e.g., presenting an item simply by tapping with no cue; modeling the word, then asking students to read it when not indicated in the script; answering with the students during word reading) embedded within the instruction. This difference may have contributed to the relatively high and stable word reading accuracy. This result is consistent with previous research that suggests that certain treatment integrity failures may not result in degradations of performance for certain student behaviors (Gansle & McMahon, 1997; Holcombe et al., 1994; Rhymer, Evans-Hampton, McCurdy, & Watson, 2002; Vollmer et al., 1999).

Since we defined instructional behaviors topographically rather than functionally, different topographies of behavior may have resulted in the same function. For example, Ms. Dean and Ms. Jones had near-zero baselines for presentation rate; however, Ms. Dean was answering with students, while Ms. Jones gave almost all individual turns. Ms. Allen, Ms. Lewis and Ms. Tate had variable baselines, all for different reasons. Ms. Allen’s lower data points were typically a result of not presenting a “cue” before signaling for a response (i.e., she frequently had a higher presentation rate, but it consisted of simply tapping for each word, which were incorrect presentations and not counted in the rate measure). Ms. Lewis frequently modeled words before students
responded, reducing her accurate presentations in some sessions during baseline. Ms.
Tate presented items quickly, but with variable rates during baseline. Some days, she
stopped to interact with students, while some days she did not. Since we did not observe
systematic differences in student on-task behavior or reading accuracy, these variations in
presenting items may have all functioned as adequate group responses for students.
Future research would be needed to determine a threshold for when different types of
intervention fidelity failures relate functionally to student behavior and reading accuracy
in Direct Instruction implementations. For example, it is possible that a high rate of
presentations is functionally related to student outcomes, rather than accurate
presentation signals (i.e., a consistent cue, pause, signal). On the other hand, the quality
of praise (i.e., specific, varied, enthusiastic, appropriate to the student, sincere) may be
more related to student outcomes than an absolute rate of praise.

We defined on-task behavior in a way that reflected classroom rules and problem
behavior that we observed. This definition was stricter than many definitions in previous
literature. We did not see strong effects on student on-task behavior possibly because the
definition we used did not focus on behaviors that are sensitive to changes in praise and
instructional pace. For example, this definition did not distinguish between relatively
minor off-task behavior (e.g., playing with an object) and more dramatic off-task
behavior (e.g., leaving one’s seat). For the first example, the student could still be reading
while playing with the object, but for the other example, the student would not be able to
read if he or she left the area. Although on-task behavior has been found to relate to
achievement in reading, it may be a problematic measure of student productivity. In this
study a student may have been “off-task,” but still able to read and benefit from
instruction. Off-task behavior may have diminished in intensity, but this would not be reflected by the definition used in this study. In addition, whole interval recording of on-task behavior may be insensitive to changes in the behavior. Students may be decreasing off-task behavior (e.g., from nine seconds of off-task behavior per ten-second interval to one second per interval), but the measurement procedure might not detect it.

Interobserver agreement on teacher behaviors was relatively high (i.e., 93% to 98% agreement); however, interobserver agreement on student behaviors was slightly lower (i.e., 87% and 90%). Some things that may have contributed to lower interobserver agreement for student behaviors include the difficulty in defining on- and off-task generally enough to account for many novel student behaviors. In addition, the students in these groups did not display many high-intensity off-task behaviors (such as yelling, leaving their seats, hitting others, etc.), so the differences between on- and off-task were sometimes subtle. Student reading accuracy was evaluated during live observations, where it was more difficult to see or hear students’ responses. During some phases, when instruction was less consistent, students dragged out responses or did not consistently answer on signal. In these situations, we found it more difficult to achieve high interobserver agreement on students’ reading accuracy. The difficulty in measuring student reading accuracy may have also contributed to the limited effects observed for students’ reading. Improving data collection procedures might increase the chances of detecting changes in students’ reading as a result of improved intervention fidelity on the part of paraprofessionals.
Social Validity

We assessed whether our measures reflected important aspects of instruction by surveying Direct Instruction experts. We also asked participants and their supervisors whether or not fluency training was feasible and useful to them. These measures shed light on whether or not fluency training might be used successfully in the future.

Paraprofessional Teaching Skills

When DI experts rated videos of paraprofessionals teaching, we found moderate to good agreement between the raters and the criterion measures (range, 75% to 90%). This level of agreement suggests that our measures captured important features of intervention fidelity for teaching Corrective Reading: Decoding. The experts never rated a video with “good” or “excellent” when the participant’s presentation or praise rate was below the criterion. All disagreements occurred when the participant’s presentation or praise rate was above the criterion, but experts rated the skills as “poor” or “OK.” This distribution of the ratings provides strong evidence that performance above our rates for presentation and praise are necessary, but not sufficient to obtain a “good” or “excellent” rating from the experts. Future studies could attempt to capture the qualities of presentation and praise that better correspond with “good” and “excellent” ratings from experts in the field.

Feasibility of Training

The two administrators in the district who would oversee the allocation of time and resources to this type of training rated the feasibility of the fluency training highly. The primary concern would be finding the time to implement it. This might be a concern,
since the results of this study suggest that additional training time might be needed to reduce the number of individuals who require performance feedback after the training is discontinued. Future research could explore ways to keep the training relatively short, while increasing its efficacy.

**Acceptability of Training**

Paraprofessionals rated the training favorably on a questionnaire about the training and its perceived effects on their teaching and students. This assessment was informal. Although it was anonymous, paraprofessionals may have rated the training higher due to perceived expectations of the researchers. In addition, paraprofessionals indicated that the training helped their students read and behave better. This result did not correspond with our observations of student behavior and reading accuracy, in general.

Paraprofessionals could have discontinued their participation in the study at any time without negative repercussions. However, all participants completed the study, lending further support to the acceptability of this training to participants.

Anecdotally, we offered the training to all paraprofessionals teaching tier II groups. Almost all of the paraprofessionals involved in tier II reading instruction at both schools volunteered for training. Paraprofessionals from a school not involved in the training came as well. Two paraprofessionals working with students in the special education program at the second school also came to the training. Five district literacy facilitators also participated in the training. In each of three training sessions, 9 to 13 individuals participated voluntarily (they were paid for their time by the district). District literacy facilitators have requested training and observation materials from us for use next
year. These events suggest that the training was acceptable to participants and supervisors in this school district.

Independent Variables

Fluency Training

Fluency training lasted approximately one hour per day for five days. The training included didactic training, brief modeling of the skills by the trainer, practice to increase accuracy, fluency practice, and feedback during practice. Trainers conducted training probes at the end of each training session to monitor the acquisition of the skills in the training setting. The trainers also observed paraprofessionals in their classroom settings during this training period to observe any generalization to the classroom setting. Some researchers have reported data on the acquisition of skills during the training with simultaneous monitoring of how participants apply the skills to the classroom setting (e.g., DiGennaro et al., 2005, 2007; Gilbertson et al., 2007; Lerman et al., 2008). Other researchers collect data during baseline, provide “typical training” to a certain criterion, then continue data collection and reporting after the training has been completed, but not during training (e.g., Bingham, 2005; Sarokoff & Sturmey, 2004; Schepis et al., 2000). Researchers may have many reasons for not monitoring application of the training to the classroom setting during training. For example, training may include only one session, so the first data point after training constitutes monitoring of application to the classroom. Acquisition of skills may result in a gradual increase in the skills, rather than a distinct change in the level of skills, so experimental control may appear to be compromised. Including these data, however, may help researchers develop a better understanding of how much training may be generally effective for different skills. These theories could
guide parametric research to determine optimally effective amounts of training and appropriate criteria to use.

The results of this study suggest that additional training time might be necessary for most participants to achieve criterion levels of performance in the classroom. Four of the five participants required performance feedback because at least one target behavior (presentation rate, praise rate, error correction accuracy) did not remain above criteria during the maintenance phase. This training is promising, because only five hours resulted in an increase in praise rate slopes for all participants, distinct level increases for error correction accuracy, and decreased variability in positive to negative comments ratios for participants with variability in baseline levels. Possibly, researchers could increase the amount of fluency training in order to improve paraprofessional treatment integrity to criterion levels, while maintaining the efficiency of training multiple individuals at the same time. If effective, this training might increase maintenance of skills and reduce the need for individualized follow-up (e.g., Lerman et al., 2004, 2008).

Treatment fidelity for fluency training was 88%. This reflects lower treatment fidelity for Ms. Jones, because she participated in the training every day, but was unable to complete training probes for health reasons. Lower treatment fidelity of fluency training may have contributed to lower acquisition of adequate praise and error correction accuracy for Ms. Jones, although additional research would be required to explore this issue further. Treatment fidelity for fluency training and interobserver agreement on treatment fidelity ratings were otherwise high.
Performance Feedback

We initiated performance feedback for four of the five paraprofessionals due to lower presentation or praise rates. This proportion of participants needing performance feedback is similar to what other researchers have found after “typical training” or didactic training. Most participants do not maintain gains and need performance feedback to maintain skills (e.g., DiGennaro et al., 2005, 2007; Gilbertson et al., 2007). We were not able to establish experimental control across participants for the difference between maintenance and performance feedback, so we cannot draw firm conclusions about the effects of performance feedback following fluency training and maintenance. However, across the three behaviors (presentation rate, praise rate, error correction accuracy) for which performance feedback might have been needed, there were 15 maintenance phases. Of those 15, only 8 required performance feedback due to low or variable skills (the others were at or above criterion during maintenance). Of the eight, performance feedback was clearly effective for three (Ms. Allen, presentation rate and error corrections; Ms. Dean, slope of praise rate); two showed potentially positive effects, but not enough time remained in the study to determine this (Ms. Lewis, presentation rate and error corrections); two had unclear effects (Ms. Dean, error corrections; Ms. Jones, praise rate); and one included only one data point, so it was not possible to determine an effect (Ms. Jones, error corrections). Only Ms. Allen and Ms. Lewis received performance feedback for lower presentation rates. The other paraprofessionals’ presentation rates were at or near criterion. To establish experimental control in this study, a minimum of three participants would have to demonstrate a distinct improvement from maintenance to performance feedback; however, only two participants per instructional variable even had
the possibility of improving dramatically. For Ms. Jones and Ms. Lewis, we could only conduct two or three sessions of performance feedback. If the study could have continued, we might have been able to help these participants increase their skills to criterion levels. Also, some researchers gradually fade out performance feedback to help maintain effects. We did not have time to do this.

In general, the interventions studied here appeared to be effective across paraprofessionals’ behaviors, rather than differentially effective for particular behaviors. This feature is important, because it means that many behaviors may be addressed within the same training and performance feedback sessions without compromising the effects on some behaviors. Fluency training and performance feedback in this study were relatively efficient, because multiple behaviors improved in a relatively short amount of time.

Limitations and Future Directions

Limitations include weaker experimental control on paraprofessional presentation rate, and lack of a clear effect on student behaviors. Future research could find participants with greater needs in presentation rate, because two participants in this study had relatively high presentation rates to start. This would provide more opportunities to replicate the effects across participants that we found for at least two participants in this study. In addition, the social validity measure suggested that our index of presentation rate may have been too high. Although no criteria exist in the literature or when talking to DI implementers, presentation rates that are too high may decrease the quality of
instruction as much as presentation rates that are too low. Future research could explore the range of presentation rates and quality that result in optimal outcomes for students.

In this study, we did not detect a clear effect on student reading accuracy or on-task behavior. Student reading accuracy for most students remained relatively high across the study. We may have observed a ceiling effect on this behavior. In addition, if students were placed in groups that were a bit easy for them, factors such as presentation and praise rate might not have an effect on student accuracy, since it was relatively high. Luke’s word reading accuracy decreased gradually during the study. This suggests that changes in student first-time corrects may occur more slowly than changes in intervention fidelity. This dependent variable would not be conducive to establishing experimental control in this time frame. Student on-task behavior also appeared to be relatively insensitive to the changes in intervention fidelity. Future research could explore the features of the independent variable and measures that promote changes in important student behaviors such as word reading accuracy and on-task behavior.

Only one paraprofessional maintained all skills after fluency training was completed, avoiding the need for performance feedback. Ms. Tate presented items correctly at a high rate during baseline (typically above the classroom criterion), which may have helped her acquire the other skills more quickly and maintain them after fluency training ended. Also, she was the only paraprofessional to end the study with all skills above the classroom criterion on three consecutive days. Future research with larger groups may determine whether mastery of certain prerequisite skills prior to the study enables some paraprofessionals to benefit differentially from this type of training. If so, researchers could improve this training to better meet the needs of individuals with
lower prerequisite skills. Also, resources might be allotted accordingly: providing group training to all, with the awareness that some may need more individualized follow-up feedback or instruction.

Fluency training included multiple components, such as rationale for specific behaviors, modeling, accuracy practice, fluency practice, training probes, and feedback on training performance. In addition, multiple skills were taught during the training, such as correct presentations, praise, praise around, signal error corrections, content error corrections, and giving “think time” when needed. We do not know which components of the training were effective or ineffective. Future research would be needed to compare components and composites to determine the most effective and efficient delivery of this type of training. Also, the results suggest that while the paraprofessionals made some improvements, most paraprofessionals did not meet the criteria of three consecutive days with presentation and praise rates above the classroom criteria before training ended. Better generalization and maintenance of skills may have occurred with more hours of this type of training (e.g., Lerman et al., 2008; Lerman et al., 2004) and training to some “mastery” criterion (although the appropriate criterion would need to be determined as well). These aspects of the training warrant more research.

We provided training on skills which we did not assess during this study (e.g., “praise around,” correcting signal errors). Also, we did not provide comprehensive training on all aspects of implementing the Corrective Reading: Decoding program; therefore some aspects of treatment fidelity on Direct Instruction are missing from this study. For example, we did not assess nor train on issues of lesson completion,
administration of mastery tests, story reading, or completing workbooks. Future research could include these aspects of treatment integrity for Direct Instruction reading programs.

During training, we focused on increasing fluency in the target skills. The rates of these behaviors may have increased, but the apparent quality did not always increase. We included quality in our measures as much as possible (e.g., rate of accurate presentations; error correction accuracy), but some features of quality teaching skills are more difficult to capture. For example, during training, we discussed features of effective praise: it is specific to the behavior, appropriate to the student, varied in content and delivery, enthusiastic and sincere. Paraprofessionals practiced giving praise at high rates using varied praise words, while focusing on different behaviors. In the classroom, however, we evaluated only the overall rate of praise, not whether it was specific, appropriate, varied, enthusiastic and sincere. Anecdotally we observed that some of the increases in rate of praise reflected addition of lower quality praise statements. Ms. Allen’s praise rate increased dramatically after fluency training, but she typically said, “Good” in the same tone of voice as giving an instruction and moved quickly to the next item. Ms. Dean did not praise as much as Ms. Allen, but the praise appeared to be higher quality: she frequently appeared enthusiastic, provided specific praise, and varied how and what she praised. With the emphasis on fluency in training, some paraprofessionals may have thought that speed was the most important feature of their instructional skills.

Since we defined instructional skills topographically rather than functionally, we did not directly assess the effects of each paraprofessional response on student behavior. For example, accurate presentations were defined according to the topography of the paraprofessionals’ behavior (i.e., an appropriate cue, pause, signal), rather than from the
effect of the presentation on students’ group unison responses. We could have defined accurate presentations as ones which resulted in adequate group unison responses, and inaccurate presentations as ones which resulted in “signal errors” from the students (times when students did not answer together at the time of the signal). We defined negative comments as those which appeared negative (e.g., telling students what not to do, rather than saying what to do), rather than comments which followed inappropriate student behavior. That definition did not account for seemingly neutral comments or actions (e.g., looking at a student, tapping them on the shoulder, saying the student’s name) that may have acted as reinforcers for off-task behavior. Possibly, these types of comments and behavior from the paraprofessional were “better” reinforcers for off-task behavior than low quality praise was a reinforcer for on-task behavior. By counting praise statements, we were attempting to quantify the paraprofessional’s use of her attention to reinforce student behaviors. However, our definition of praise may not be a close enough approximation of attention. Future research may be warranted to determine measures that more closely approximate the paraprofessionals’ use of attention and its effects on student behavior.

During fluency training, the trainer modeled the target skills and typical errors in the target skills before the paraprofessionals practiced the skills. This modeling was limited in scope and content. Studies have shown that the range of examples and non-examples provided in modeling affects generalization of skills (e.g., Moore & Fisher, 2007; Sprague & Horner, 1984). Moore and Fisher found that video models that included a wider range of examples of the target behavior resulted in better acquisition of functional analysis skills than videos with a more limited set of examples of the target
behavior. The fluency training in the current study may have been more effective if the
trainer had provided a wider range of examples and non-examples of instructional
behaviors. Video modeling may be a more efficient means of presenting the range of
skills in typical settings, rather than the limited example of the trainer modeling in a
training setting without students.

One strategy for increasing generalization and maintenance of skills is to ensure
that the new skills will be reinforced by situations in the target environment (Stokes &
Baer, 1977; White et al., 1988). Ideally, more effective teaching skills would result in
immediate improvements in student outcomes, which would in turn reinforce the use of
the new teaching skills. In the current study, we taught paraprofessionals some skills that
might actually have resulted in temporary decrements in student behavior. For example,
we taught paraprofessionals to correct errors by modeling the missed word, asking
students to say it, spell it, say it again, then go back to the beginning of the word list (at
least one or two words back), and asking students to read the words again. During
baseline, paraprofessionals typically did not include all of these steps during error
corrections. When paraprofessionals included complete error correction procedures
during and after training, students sometimes complained about having to do all of the
steps. For example, some students said, “No, not again!” when the paraprofessional asked
them to go back to the beginning of the list after error corrections. Other students put
their heads down after numerous error corrections had been conducted in a lesson. In
some groups, this situation was more likely than in others because of the
paraprofessionals’ baseline behaviors. For example, Ms. Lewis often modeled harder
words during baseline when it was not scripted. When she stopped doing this during
fluency training and maintenance, students’ errors increased, and more error corrections were necessary. Students appeared to react to this with off-task behaviors, especially during maintenance (see Figure 5, Luke). Future research could explore what procedures might be implemented to help participants maintain intervention fidelity when faced with student behavior that degrades or does not noticeably change in the short term due to new skills learned in the training.

Anecdotally, paraprofessionals gave different reasons for decreases in their own intervention fidelity after fluency training, but these were not evaluated formally. For example, one paraprofessional said she did not use more varied or enthusiastic praise because she felt that it distracted students. Another paraprofessional stopped using a cue in some situations because she felt it might be easier for students to respond accurately. Two paraprofessionals were going to end structured reading instruction several days before the last day of tier II reading groups. Separately, they said that they were going to start a new story in the curriculum and did not want students to be left in the middle of the story when groups ended. These rules appeared to compete with continued implementation fidelity. While we do not know if these rules controlled their behavior, it may be productive to explore the role that rule-governed behavior may play in the generalization and maintenance of training. Also, researchers and trainers may attempt to anticipate these types of barriers to maintenance of implementation fidelity and address them in the training more effectively.

We assessed the fidelity of fluency training and performance feedback with checklists of steps completed. We did not explicitly evaluate the quality of each step (e.g., fluency of delivery, etc.), which limits the validity of this measure (Gersten et al.,
2005); however, no norms exist for defining the quality of the steps of this training or performance feedback. In addition, the trainer served as the primary or secondary observer on many occasions. Sometimes the author of this study self-assessed intervention fidelity of the training or feedback as she did the training or provided the feedback, rather than from watching a video later. The agreement on ratings between the author and the other observer was high on these measures (99%), so it is unlikely to be a source of measurement error.

We conducted this study during the last quarter of the school year, which may have affected some of the results. For example, paraprofessionals probably had been teaching in ways similar to that seen in baseline for several months to several years (i.e., paraprofessionals had been teaching *Corrective Reading: Decoding* for 6 months to 5 years at the beginning of the study). The instructional behavior evident in baseline was well-practiced. The training may have been more effective at the beginning of the school year, when expectations for groups may be more flexible, or with all new paraprofessionals, who had not practiced lower implementation fidelity for months to years before the training. Paraprofessionals and students may have adjusted to the changes required during training more easily at the beginning of the year. Also, four of five students ended the study with low and decreasing levels of on-task behavior, regardless of the intervention being conducted with the paraprofessional (i.e., maintenance or performance feedback). Possibly, students were anticipating the end of groups and were less on-task as a result.
Implications and Conclusion

Studies have shown that paraprofessionals feel that they need more training in managing student behavior and teaching strategies (e.g., Downing, Ryndak, & Clark, 2000; Riggs & Mueller, 2001; Wallace, 2003) and providing training is a high priority for schools (e.g., Giangreco, Edelman, & Broer, 2003). The current study suggests that fluency training for paraprofessionals in instructional and behavior management strategies may be worth investigating further. This training may be effective especially for increasing praise statements, error correction accuracy and decreasing variability in positive to negative ratios. Adding more training time might enhance treatment effects for paraprofessionals and students. Researchers could also study the addition of other components to the training to try to improve on-task behavior for students.

Maintenance of improvements might be enhanced by adding components to the training or performance feedback. For example, studies have suggested that teachers and paraprofessionals can maintain treatment effects with the use of self-monitoring procedures (Allinder, Bolling, Oats, & Gagnon, 2000; Bingham, 2005; Sutherland & Wehby, 2001; Van Vonderen & De Bresser, 2005). Future research on training in Direct Instruction could include a strategy for maintenance such as self-monitoring. This feature could also enhance the efficiency of the interventions, since self-monitoring requires less supervisor oversight.

Finding efficient and effective means for promoting intervention fidelity of research-based instruction is critical to ensuring that RTI is an effective system for intervening with students at risk for academic failure and identifying students with
learning disabilities. The results of this study suggest that fluency training followed by performance feedback for paraprofessionals is a potential means to increase the efficacy and efficiency of professional development. Five hours of training resulted in increased praise, error correction accuracy and decreased variability of positive to negative comments ratio. However, effects on students were more subtle or unclear. Future research could determine if more training or changes in measurement methods could enhance the effects or ability to detect effects. Our measures appeared to detect features of instruction that were necessary, but not sufficient to obtain acceptable ratings from DI experts. Administrators who would be responsible for continuing the use of this training in the future rated the feasibility highly. In addition, participants rated the training as acceptable and useful to them. This study suggests a practical approach to improving professional development for paraprofessionals working with students at risk for reading failure within an RTI system.
REFERENCES


Hearing on literacy: Why kids can't read: Hearings before the Committee on Education and the Workforce, 105th Cong. 1st Sess., page number (1997) (testimony of Reid G. Lyon).


APPENDICES
Appendix A

Teacher Training in Reading Instruction Review Details
This literature review extends the NRP review on teacher education in reading instruction to 2007.

Search Methods

The literature search was conducted using ERIC, PsycInfo, and Google Scholar with different combinations of the following keywords: inservice, teacher training, education, reading, achievement, instruction, and professional development. Abstracts of articles were reviewed, and articles were chosen for further review if it appeared that the article was a study of teacher training and student achievement in reading from 1999 to the present, inclusive. To be included in the following review, the article had to detail an intervention conducted with inservice teachers with at least one goal of improving student (K-12) reading achievement. Inservice teachers were considered teachers who were currently teaching and not participating in an initial teacher certification program. “Teachers” could also include other interventionists (e.g., special education teachers, reading specialists, paraprofessionals, speech pathologists, etc.) who were not the researchers or their immediate trainees. Student achievement in reading had to be assessed in the study. So, articles that included preservice teachers, or interventions with students run by the researchers, or only teacher outcomes were not included. This search resulted in eight articles.

Each article was coded for the number of teachers and students involved; the type of student population and grade levels; whether or not a comparison group was included and if it was a quasi-experimental, random or other type of assignment to groups; the type of teacher intervention; teacher outcomes; the type of student intervention; student
outcomes; and the reading components included in the interventions (i.e., phonemic awareness, phonics, fluency, vocabulary, comprehension; taught with a systematic, explicit approach).
Appendix B

Paraprofessional Demographic Questionnaire
Paraprofessional Demographic Questionnaire

Please answer the following questions. The information will be anonymous and summarized as group data.

1. Age: ________

2. Educational background (please check the highest level completed):
   a. _____ 8th Grade
   b. _____ Some college (no degree)
   c. _____ Associate’s degree
   d. _____ Bachelor’s degree
   
   What was your major?: __________________________
   
   e. _____ Some graduate school (no degree)
   f. _____ Master’s degree
   
   What was your degree in?: __________________________

3. Do you have a teaching credential?  Y       N
   a. If yes, what state is the credential from? __________________________
   b. In what teaching area is the teaching credential (e.g., elementary education, special education, etc.)?

   __________________________

4. Number of years teaching or paraprofessional experience:

________________________
5. Number of years teaching with a Direct Instruction program:

____________________

6. Number of years teaching with Corrective Reading: Decoding:

____________________

7. Initial training in Direct Instruction (please check all that apply):
   a. _____ I participated in the training that the district provided at the beginning of the school year for the large group of paraprofessionals.
      Year: _________________
      i. I went to the following sessions (please check all that apply):
         1. _____ General training on Direct Instruction (morning session)
         2. _____ Corrective Reading
         3. _____ Reading for all Learners
         4. _____ Reading Mastery
         5. _____ Early Reading Intervention
   b. _____ I received my initial training on an individualized or small group basis with a literacy coordinator in the district (do not mark this if you attended the training at the beginning of the year).
   c. _____ I received training in Direct Instruction from an Association for Direct Instruction conference, such as the one in Eugene, Oregon, or a regional ADI conference.
d. _____ I received training in Direct Instruction through a teacher preparation program at a College or University.

i. College or University name:


8. Follow-up coaching in Direct Instruction (please check all that apply):

a. _____ A literacy coordinator or other district employee has observed me teaching at least one of my groups and has provided feedback based on this observation.

i. If you checked this item, please include the following information:

1. Number of times this has been done this school year:

   __________

2. Average amount of time the coach observed each time:

   ________

3. Average amount of time the coach provided feedback on the session each time:

   ____________________________

4. Did the coach stop you during any session and show you how to perform a particular skill?  Y  N

5. Did the coach show you how to perform a particular skill after any session?  Y  N
b. _____ I have participated in small group training with other paraprofessionals conducted by a literacy coordinator that was not related to any particular observation of my skills.

   i. If you checked this item, please include the following:

      1. Number of times this has been done this school year:

         ____________

      2. Average amount of time these sessions took:

         ______________

      3. Average number of paraprofessionals who participated:

         ________

c. _____ Other, please describe:

_________________________________________
_________________________________________
_________________________________________
Appendix C

Direct Instruction Reading Instructional Rating Forms
Instructions: Please watch the video of each session provided in the order presented on the DVD.

After watching a session, please answer the following questions by selecting the choice that most closely represents your evaluation of the instructor’s skills in presenting the Direct Instruction formats for that session alone. After you have rated that session, watch another session and rate that session on a separate form.

Reviewer ID: ___________________                             Video session number: ___________________

Date of this review: ______________

Please circle your rating for each item.

1. The **pace** with which the instructor presented items accurately was:
   a. Excellent; I would consider this to be close to ideal.
   b. Good; I would find this acceptable in a DI classroom.
   c. OK; I would recommend more practice and coaching for this person.
   d. Poor; I would find this unacceptable in a DI classroom.

2. The **amount of praise** the instructor provided to students was:
   a. Excellent; I would consider this to be close to ideal.
   b. Good; I would find this acceptable in a DI classroom.
   c. OK; I would recommend more practice and coaching for this person.
   d. Poor; I would find this unacceptable in a DI classroom.

3. The **quality of praise** the instructor provided to students was:
   a. Excellent; I would consider this to be close to ideal.
   b. Good; I would find this acceptable in a DI classroom.
   c. OK; I would recommend more practice and coaching for this person.
   d. Poor; I would find this unacceptable in a DI classroom.

4. The **ratio of praise to corrective** or negative statements was:
   a. Excellent; I would consider this to be close to ideal.
   b. Good; I would find this acceptable in a DI classroom.
   c. OK; I would recommend more practice and coaching for this person.
   d. Poor; I would find this unacceptable in a DI classroom.

5. The **quality of error corrections** provided to students was:
   a. Excellent; I would consider this to be close to ideal.
   b. Good; I would find this acceptable in a DI classroom.
   c. OK; I would recommend more practice and coaching for this person.
d. Poor; I would find this unacceptable in a DI classroom.

6. The **overall quality of instruction** presented in this session was:
   a. Excellent; I would consider this to be close to ideal.
   b. Good; I would find this acceptable in a DI classroom.
   c. OK; I would recommend more practice and coaching for this person.
   d. Poor; I would find this unacceptable in a DI classroom.

7. Do you have other comments on the quality of the instruction in this video? If so, please explain.

________________________________________________________________________________
________________________________________________________________________________

*Comparison Rating Form for DI Videos*

Reviewer ID: __________

Date of this review: __________

Please pretend that you would be hiring a person for a job at a school teaching small-groups using *Corrective Reading: Decoding*. Based on these two videos, pretending they are two different people, which one represents a person you would be more likely to hire based on her presentation of word attack alone (choose one answer)?

   a. Video ______

   b. Video ______

   c. Both are equivalent to me.
Appendix D

Questionnaire for District Reading Coordinators
Instructions: Please circle the answer that most closely represents your opinion about the feasibility of the training (not including classroom observations) as you observed or as described to you by the participants. The primary concern is whether or not this type of training would be practical and worth the time and money for the district to conduct in the future.

1. Considering the results of the training, the amount of time taken to provide the fluency training was reasonable.
   

2. Considering the results of the training, the amount of money required to provide the fluency training was reasonable.


3. The training was simple enough to be easily implemented by district staff.


4. The paraprofessionals enjoyed this enough that we would feel comfortable providing them this training in the future.
5. We would be able to provide more than 5 hours of this type of training to paraprofessionals in the future.

6. Overall, we are likely to use this training in the future.

7. Are there other barriers to the use of this training in the future? If so, please describe them (continue on the back of this page, if needed).
Appendix E

Questionnaire for Paraprofessionals Regarding Training
Instructions: Please circle the answer that most closely represents your level of agreement with each statement about the observations and training you received. This questionnaire is anonymous and will have no information identifying you.

1. I felt that the training I received helped me improve my classroom teaching.

2. I enjoyed participating in the training.

3. I feel that the training I received helps my students read better.

4. I feel that the training I received helps my students behave better during class.

5. I feel that the training helped me present items faster and more effectively in class.

6. I feel that the training helped me praise more during class.
7. I feel that the training helped me provide more positive than negative statements to my students.

8. I feel that the training helped me provide more accurate error corrections to students in class.

9. I thought that practicing the teaching components at a fast rate during training helped me apply them to my teaching in the classroom.

10. I would recommend this training to other people.

11. This training had little effect on my teaching in the classroom.

12. Overall, I found this training to be worthwhile and helpful to me.

13. Comments:
Appendix F

Outline of Fluency Training Sessions
## Table 14

**Fluency Training Outline**

<table>
<thead>
<tr>
<th></th>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Day 4</th>
<th>Day 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Presentation Rate</strong></td>
<td>Rationale, model</td>
<td>Review rationale</td>
<td>Look at students</td>
<td>SS – Acc</td>
<td>Intro. varied think time</td>
</tr>
<tr>
<td></td>
<td>SS – accuracy</td>
<td></td>
<td></td>
<td>SS – Flu</td>
<td>SS w/E – Acc</td>
</tr>
<tr>
<td></td>
<td>SS – fluency</td>
<td></td>
<td></td>
<td></td>
<td>SS – Flu</td>
</tr>
<tr>
<td><strong>Praise</strong></td>
<td>Rationale, features</td>
<td>Review rationale</td>
<td>Say praise words – Flu</td>
<td>Introduce, model “Praise Around”</td>
<td>Review Praise Around</td>
</tr>
<tr>
<td></td>
<td>Alphabet activity</td>
<td></td>
<td>SS – Acc</td>
<td>SS – Acc</td>
<td>SS – Acc</td>
</tr>
<tr>
<td></td>
<td>Say praise words – fluency</td>
<td></td>
<td>SS w/P – Flu</td>
<td>SS – Flu</td>
<td>SS – Flu</td>
</tr>
<tr>
<td><strong>Error Corrections</strong></td>
<td></td>
<td>Signal Error Corrections</td>
<td>Rationale, model</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SS w/E (40%) – Acc</td>
<td>SS w/E (40%) – Acc</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SS w/E (40%) – Flu</td>
<td>SS w/E (40%) – Flu</td>
<td></td>
</tr>
<tr>
<td><strong>Fluency Practice</strong></td>
<td>3S – Flu (presentations)</td>
<td>SS w/P, SE – Flu</td>
<td>SS w/P, E (40%) – Flu</td>
<td>SS w/P, E (40%) – Flu</td>
<td>MS w/SE, P, E (40%) – Flu</td>
</tr>
<tr>
<td><strong>Training Probe</strong></td>
<td>DS w/P, E (10%)</td>
<td>DS w/P, E (10%)</td>
<td>DS w/P, E (10%)</td>
<td>DS w/P, E (10%)</td>
<td>DS w/P, ± (10%)</td>
</tr>
</tbody>
</table>

**Note.**

- SS = Simple script
- P = Praise required
- Acc = Practice for accuracy
- MS = Moderate script
- E = Student errors included
- Flu = Practice for fluency
- DS = Decoding script
- SE = Signal error
Fluency Training Details

Materials needed each day:

- Handouts for paraprofessionals (guided notes to encourage active responding)
- Curriculum for practice (Scripts for teachers and “students”)
- Countdown timer for each group
- Pens/pencils
- Data sheets and graphs for paraprofessionals to record their own practice
- Data sheets for trainer to record during training probe
- Graphs for trainer to graph performance of each skill during training probe

Training Day 1 (60 minutes)

<table>
<thead>
<tr>
<th>Minutes</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Introductions of trainer and participants</td>
</tr>
<tr>
<td>3</td>
<td>Purpose of training sessions and different types of training (e.g., rationale, model, accuracy, fluency practice, teaching pace practice, training probe), and schedule for the day.</td>
</tr>
<tr>
<td>30</td>
<td><strong>Presentation Rate</strong></td>
</tr>
</tbody>
</table>

Trainer explained rationale for group responses and signals.

Trainer modeled cue, pause, signal and point out critical features (e.g., accurately presenting script, not answering with students or mouthing answers, keeping pause consistent, signal audible and visible, etc.).
Participants practiced cue, pause, signal with simple script (“What word?”), no student errors, while trainer gave feedback (positive and corrective when needed).

Participants practiced cue, pause, signal for fluency with simple script (aim = 20 presentations per minute minimum). Repeated for one-minute timings until aim was met, or three one-minute practices at most. Graphed results.

**Praise Rate/ Basic Behavior Management**

Trainer explained rationale for using praise (a way to focus teacher attention on appropriate behaviors).

Trainer presented features of effective praise (specific, varied, sincere, age appropriate, enthusiastic).

Alphabet activity (participants wrote down one praise statement per letter of the alphabet).

Participants conducted fluency practice on praise statements alone (number of varied praise statements per minute).

**Training Probe**

Participants were given a script from the *Corrective Reading: Decoding* program from the lesson they most recently taught in class. The trainer timed the target paraprofessional for 2 minutes while they taught “students” who made approximately 10% errors on words with no off-task behavior. The trainer recorded data on the
paraprofessional’s presentation rate, praise rate, and percent accurate error correction steps during this probe.

The trainer calculated the values for each skill from this 2-minute observation, put the data on graphs and discussed the results with the paraprofessional.

**Training Day 2 (57 minutes)**

<table>
<thead>
<tr>
<th>Minutes</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Trainer reviewed the schedule for the day.</td>
</tr>
<tr>
<td>15</td>
<td><strong>Praise Rate</strong></td>
</tr>
<tr>
<td></td>
<td>Trainer briefly reviewed features of effective praise.</td>
</tr>
<tr>
<td></td>
<td>Participants conducted fluency practice on praise statements alone (number of varied praise statements per minute).</td>
</tr>
<tr>
<td></td>
<td>Participants practiced presenting simple script with praise while trainer gave feedback.</td>
</tr>
</tbody>
</table>

**Fluency practice**

Participants practice the simple script as quickly as possible (aim = 20 or more presentations per minute), with as many praise statements as possible (aim = 6 or more praise statements per minute), with no student errors.

<table>
<thead>
<tr>
<th>Minutes</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td><strong>Signal Error Corrections</strong></td>
</tr>
</tbody>
</table>
|         | Trainer identified difference between signal errors (students answer late, early or do not answer) and content errors (students say the wrong answer), presented rationale for correcting signal errors and critical
features of correcting signal errors (correct ALL signal errors; stop immediately; tell students what you want them to do; repeat item until all answering together on signal).

Trainer modeled correcting signal errors, pointing out critical features.

Participants practiced correcting signal errors (marked in student script) with no student errors on content, while trainer gave feedback.

15 Presentation Rate

Trainer briefly reviewed rationale and critical features of cue, pause, signal. Trainer introduced the rationale for looking at students when they are responding.

Participants practiced cue, pause, signal, emphasizing looking at students when they answer, no student errors, while trainer gave feedback.

Participants practiced cue, pause, signal while looking at students and correcting signal errors for fluency (aim = 20 presentations per minute minimum). Repeat for one-minute timings until aim was met, or three one-minute practices at most.

5 Training Probe

Participants were given a script from the Corrective Reading: Decoding program from the lesson they most recently taught in class (and different from other training probes). The trainer timed the target paraprofessional for 2 minutes while they taught “students” who made approximately 10% errors on words with no off-task behavior. The trainer recorded data on the paraprofessional’s
presentation rate, praise rate, and percent accurate error correction steps during this probe.

The trainer calculated the values for each skill from this 2-minute observation, put the data on graphs and discussed the results with the paraprofessional.

**Training Day 3 (52 minutes)**

<table>
<thead>
<tr>
<th>Minutes</th>
<th>Topic:</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Trainer reviewed the schedule for the day.</td>
</tr>
<tr>
<td>15</td>
<td><strong>Praise Rate and Signal Errors Review</strong></td>
</tr>
<tr>
<td></td>
<td>Participants conducted fluency practice on praise statements alone (number of varied praise statements per minute).</td>
</tr>
<tr>
<td></td>
<td>Participants conducted accuracy practice on correcting signal errors.</td>
</tr>
<tr>
<td>30</td>
<td><strong>Content Error Corrections</strong></td>
</tr>
<tr>
<td></td>
<td>Trainer reviewed the two types of errors (signal and content)</td>
</tr>
<tr>
<td></td>
<td>Trainer presented rationale for providing explicit error corrections.</td>
</tr>
<tr>
<td></td>
<td>Trainer modeled error correction procedure (“model, read, spell, read”), pointing out critical features.</td>
</tr>
<tr>
<td></td>
<td>Participants practiced error correction procedure using simple script with a “student” error on approximately 40% of words (marked in the “student” scripts), while the trainer provided feedback.</td>
</tr>
<tr>
<td></td>
<td>Participants practiced error correction procedure for fluency, by presenting the simple script as quickly as possible with 40% “student” errors while maintaining a target rate of 20 presentations per minute, 6</td>
</tr>
</tbody>
</table>
praise statements per minute and 95% accuracy on content error corrections. Participants graphed results for all skills.

5 Training Probe

Participants were given a script from the Corrective Reading: Decoding program from the lesson they most recently taught in class (and different from other training probes). The trainer timed the target paraprofessional for 2 minutes while they taught “students” who made approximately 10% errors on words with no off-task behavior. The trainer recorded data on the paraprofessional’s presentation rate, praise rate, and percent accurate error correction steps during this probe.

The trainer calculated the values for each skill from this 2-minute observation, put the data on graphs and discussed the results with the paraprofessional.

Training Day 4 (57 minutes)

<table>
<thead>
<tr>
<th>Minutes</th>
<th>Topic:</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Trainer reviewed the schedule for the day.</td>
</tr>
<tr>
<td>10</td>
<td>Review of content error corrections</td>
</tr>
<tr>
<td>40</td>
<td>Positive/Negative Comments Ratio: Praise Around</td>
</tr>
</tbody>
</table>

Trainer presented rationale for “praise around” procedure.

Trainer presented steps and critical features of praise around.

Participants identified inappropriate behaviors and corresponding incompatible behaviors.
Trainer modeled praise around with participants acting as students.

Participants practiced using praise around with simple script with others acting as students and trainer providing feedback (positive and corrective when needed).

Participants practiced praise around for fluency with other skills at previously stated criteria. Participants graphed results for presentation rate, praise rate, and content error corrections.

5 Training Probe

Participants were given a script from the Corrective Reading: Decoding program from the lesson they most recently taught in class (and different from other training probes). The trainer timed the target paraprofessional for 2 minutes while they taught “students” who made approximately 10% errors on words with no off-task behavior. The trainer recorded data on the paraprofessional’s presentation rate, praise rate, and percent accurate error correction steps during this probe.

The trainer calculated the values for each skill from this 2-minute observation, put the data on graphs and discussed the results with the paraprofessional.

Training Day 5 (62 minutes)

<table>
<thead>
<tr>
<th>Minutes</th>
<th>Topic:</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Trainer reviewed the schedule for the day.</td>
</tr>
<tr>
<td>15</td>
<td>Review Praise Around</td>
</tr>
</tbody>
</table>
Trainer briefly reviewed praise around. Participants practiced praise around for accuracy while trainer gave feedback.

**Presentation Rate with Varied Think Time**

Trainer gave participants rationale for varying think time appropriately (e.g., giving a little more time on words that have been difficult in the past; go quickly through words that have not been hard for students).

Trainer modeled providing varied think time during presentations.

Participants practiced providing varied think time for accuracy while correcting signal and content errors, and praising students.

**Putting it all together**

Trainer modeled moderate script for participants (“What sound? What word?”) using all skills previously practiced.

Participants practiced presenting the moderate script for accuracy including all skills previously practiced (i.e., presentations, praise, monitoring students, praise around, signal and content errors, and varied think time).

Participants practiced presenting moderate script for fluency including skills previously practiced, except for providing varied think time.

**Training Probe**

Participants were given a script from the *Corrective Reading: Decoding* program from the lesson they most recently taught in class (and different from other training probes). The trainer timed the target
paraprofessional for 2 minutes while they taught “students” who made approximately 10% errors on words with no off-task behavior. The trainer recorded data on the paraprofessional’s presentation rate, praise rate, and percent accurate error correction steps during this probe.

The trainer calculated the values for each skill from this 2-minute observation, put the data on graphs and discussed the results with the paraprofessional.
Appendix G

Verbal/Graphic Performance Feedback Protocol
Performance Feedback Protocol

Materials needed:
  Tx Fidelity forms
  Graphs
    Pos/Neg Ratio
    Presentation Rate
    Praise Rate
    Error Corrections

1. Review graph _______________
   a. Point out latest data point (tell what the value is for that point).
   b. Point out the overall level and/or trend for the current phase.
   c. Point out the criterion level for this skill.
   d. Positive for what has been done with fidelity
   e. Corrective for what needs to be done differently
   f. Recommendations for how to do it differently
   g. Any questions?

2. Review graph _______________
   a. Point out latest data point (tell what the value is for that point).
   b. Point out the overall level and/or trend for the current phase.
   c. Point out the criterion level for this skill.
   d. Positive for what has been done with fidelity
   e. Corrective for what needs to be done differently
   f. Recommendations for how to do it differently
   g. Any questions?

3. Review graph _______________
   a. Point out latest data point (tell what the value is for that point).
   b. Point out the overall level and/or trend for the current phase.
   c. Point out the criterion level for this skill.
   d. Positive for what has been done with fidelity
   e. Corrective for what needs to be done differently
   f. Recommendations for how to do it differently
   g. Any questions?

4. Review graph _______________
   a. Point out latest data point (tell what the value is for that point).
   b. Point out the overall level and/or trend for the current phase.
   c. Point out the criterion level for this skill.
   d. Positive for what has been done with fidelity
   e. Corrective for what needs to be done differently
   f. Recommendations for how to do it differently
   g. Any questions?

5. Written summary: Recommendation from each graph.
Appendix H

Written Performance Feedback Form
Observation Recommendations

Date: ______________
Name: ___________________________

Presentation Rate:

Praise Rate:

Error Corrections:

Positive to Negative ratio:

Other:
Appendix I

Fluency Training Treatment Fidelity Forms
Fluency Training Treatment Fidelity, Day 1

_____ Introductions, Rationale for Study (5 min.; Actual: ______ min.)

Focus, Cue, Pause, Signal (30 min.; Actual: _____ min.)

_____ Rationale for group responses
_____ Description of the steps
_____ Modeling of the steps
_____ Accuracy Practice
_____ Fluency Practice
_____ Graph results

Basic Behavior Management/ Praise (25 min.; Actual: _____ min.)

_____ Keep them engaged
_____ Use your attention
_____ Features of Effective Praise (Specific, Varied, Sincere, etc.)
_____ Alphabet of Praise Words
_____ Fluency practice with praise words (?If there’s time)

Training Probe

_____ 2 minute timing with Decoding Script
_____ 10% errors
_____ Graph results
Fluency Training Treatment Fidelity, Day 2

___ Review presentation (cue, pause, signal) (5 min.; Actual: ______ min.)

Praise words (15 min.; Actual: _____ min.)

   ___ Review effective praise
   ___ Fluency practice with praise words
   ___ Fluency practice with presentation rate and praise, List 2

Signal Error Corrections (20 min.; Actual: ____ min.)

   ___ Two types of errors (content and signal errors)
   ___ Review steps for correcting signal errors
   ___ Model correcting signal errors
   ___ Accuracy practice with signal errors

Monitoring Students (15 min.; Actual: _____ min.)

   ___ Rationale for monitoring students
   ___ Practice script with presentation rate, praise and monitoring students, List 3

Training Probe

   ___ 2 minute timing with Decoding Script
   ___ 10% errors
   ___ Graph results
Fluency Training Treatment Fidelity, Day 3

_____ Review Fluency practice with praise words and signal error corrections

(15 min.; Actual: ______ min.)

Content Error Corrections (30 min.; Actual: ____ min.)

_____ Review two types of errors

_____ Introduce the steps of the “read, spell, read” error correction

_____ Model the steps of the “read, spell, read” error correction

_____ Accuracy practice with error corrections, List 4

_____ Fluency practice with error corrections, List 4

Training Probe

_____ 2 minute timing with Decoding Script

_____ 10% errors

_____ Graph results
Fluency Training Treatment Fidelity, Day 4

_____ Review content error corrections- practice (15 min.; Actual: ______ min.)

Praise around (30 min.; Actual: _____ min.)

_____ List the steps of praise around

_____ Practice identifying incompatible behaviors

_____ Model praise around

_____ Accuracy practice with praise around

_____ Fluency Practice with praise around

Training Probe

_____ 2 minute timing with Decoding Script

_____ 10% errors

_____ Graph results
Fluency Training Treatment Fidelity, Day 5

_____ Review praise around-practice (10 min.; Actual: ______ min.)

Varied Think Time (10 min.; Actual: _____ min.)

_____ Rationale for varied think time
_____ Review basic signal illustration
_____ Model varied think time
_____ Accuracy practice with varied think time

Putting it all together (30 min.; Actual: ____ min.)

_____ Accuracy practice with “what sound, what word”
_____ Accuracy practice correcting content, signal errors, praising and monitoring with “what sound, what word”
_____ Fluency practice with all skills on “what sound, what word” script.

Training Probe

_____ 2 minute timing with Decoding Script
_____ 10% errors
_____ Graph results
Appendix J

Verbal/Graphic Performance Feedback Treatment Fidelity Form
Date: __________ Teacher: _______________ Person giving feedback: ______

IOA: Y  N  Total time for feedback: __________

1. Review graph _______________
   a. Y  N  n/a Point out latest data point (tell what the value is for that point).
   b. Y  N  n/a Point out the overall level and/or trend for the current phase.
   c. Y  N  n/a Point out the criterion level for this skill.
   d. Y  N  n/a Positive for what has been done with fidelity
   e. Y  N  n/a Corrective for what needs to be done differently
   f. Y  N  n/a Recommendations for how to do it differently
   g. Y  N  n/a Any questions?

2. Review graph _______________
   a. Y  N  n/a Point out latest data point (tell what the value is for that point).
   b. Y  N  n/a Point out the overall level and/or trend for the current phase.
   c. Y  N  n/a Point out the criterion level for this skill.
   d. Y  N  n/a Positive for what has been done with fidelity
   e. Y  N  n/a Corrective for what needs to be done differently
   f. Y  N  n/a Recommendations for how to do it differently
   g. Y  N  n/a Any questions?

3. Review graph _______________
   a. Y  N  n/a Point out latest data point (tell what the value is for that point).
   b. Y  N  n/a Point out the overall level and/or trend for the current phase.
   c. Y  N  n/a Point out the criterion level for this skill.
   d. Y  N  n/a Positive for what has been done with fidelity
   e. Y  N  n/a Corrective for what needs to be done differently
   f. Y  N  n/a Recommendations for how to do it differently
   g. Y  N  n/a Any questions?

4. Review graph _______________
   a. Y  N  n/a Point out latest data point (tell what the value is for that point).
   b. Y  N  n/a Point out the overall level and/or trend for the current phase.
   c. Y  N  n/a Point out the criterion level for this skill.
   d. Y  N  n/a Positive for what has been done with fidelity
   e. Y  N  n/a Corrective for what needs to be done differently
   f. Y  N  n/a Recommendations for how to do it differently
   g. Y  N  n/a Any questions?

5. Written summary: Recommendation from each graph.
   a. Y  N  n/a Graph 1
   b. Y  N  n/a Graph 2
   c. Y  N  n/a Graph 3
   d. Y  N  n/a Graph 4
Appendix K

Interobserver Agreement: Computer Data Collection
We used a computer program for data collection on paraprofessionals’ presentation rate, praise rate, and negative comments. Close to the end of data collection for interobserver agreement, the author noticed that the two data streams for the two observers were sometimes off by one observation on many intervals for higher rate behaviors (e.g., presentation rate and some paraprofessionals’ praise rate). We started the two computers at the same time as usual, and found that the counters were frequently off by about 1 second. This discrepancy may have caused interobserver agreement for presentation rate, praise rate, and negative comments to be spuriously low in some cases.

To determine how much this discrepancy affected the interobserver agreement calculations, the author opened the data streams for two IOA sessions, changed discrepant observations by 1 second (only when two intervals had the exact same data streams, except for one observation in the second interval for the observer who started later), and recalculated the IOA (retaining the original IOA value in the data reported for the study). Each discrepant observation appeared to reduce the IOA by about 5% for presentation rate. With two to three discrepant observations in the sessions that the author explored, the IOA for presentation rate was reduced by about 10-15%.

Since our IOA was relatively high on these measures and this problem was detected toward the end of data collection, we did not redo IOA on previous observations. For about the last quarter of IOA observations, we calibrated our computers’ counters at the beginning of each IOA session by counting the seconds aloud as they appeared on the screen. If the count was off, we restarted the session before observations were recorded. Since this would only affect IOA observations, it is not a concern for primary data collection.
Appendix L

Interobserver Agreement: Kappa
We calculated interobserver agreement using a kappa correction for chance, since other measures of interobserver agreement may result in spuriously high agreement across behavioral rates presented in this study (Samaha, Vollmer, & Bosch, 2008). Table 15 shows interobserver agreement calculations adjusted for chance with kappa. For the dependent measures of presentation rate, praise rate, and negative comments, we calculated kappa using a computer program, Instant IOA, v. 0.5. For the dependent measures of error correction accuracy, student on-task behavior, and word reading accuracy, kappa was calculated according to the following equation: $K = (p_o - p_e) / (1 - p_e)$, where $p_o$ is the observed agreement between raters, and $p_e$ is the level of agreement expected by chance. Agreement of .60 or higher is typically deemed acceptable, although this is simply convention and has not been verified empirically for these measures (Viera & Garrett, 2005).

Table 15

*Interobserver Agreement on Dependent Measures: Kappa.*

<table>
<thead>
<tr>
<th>Dependent Measure</th>
<th>Percent of sessions</th>
<th>Mean Kappa</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presentation Rate</td>
<td>32</td>
<td>.91</td>
<td>.68 – 1.0</td>
</tr>
<tr>
<td>Praise Rate</td>
<td>32</td>
<td>.91</td>
<td>.49 – 1.0</td>
</tr>
<tr>
<td>Error Corrections</td>
<td>32</td>
<td>.85</td>
<td>.00 – 1.0</td>
</tr>
<tr>
<td>Negative Comments</td>
<td>33</td>
<td>.92</td>
<td>.78 – 1.0</td>
</tr>
<tr>
<td>Student On-Task</td>
<td>34</td>
<td>.50</td>
<td>.00 – 1.0</td>
</tr>
<tr>
<td>Student Reading Accuracy</td>
<td>47</td>
<td>.61</td>
<td>.00 – 1.0</td>
</tr>
</tbody>
</table>
The agreement based on kappa was high for presentation rate, praise rate, error corrections and negative comments, according to typical interpretations (Viera & Garrett, 2005). The agreement based on kappa for student on-task and word reading accuracy was relatively low. Further inspection of the data confirms that for more rare instances of behavior (e.g., low rates of errors or on-task behavior), even one or two disagreements can reduce Kappa to near zero levels. On the other hand, for behaviors that are likely to be observed at nearly equivalent rates (e.g., correct and incorrect steps for paraprofessional error corrections), one or two disagreements do not cause kappa to drop as precipitously. Kappa appears to be a better estimate for behaviors that have more equal rates.

A kappa statistic cannot be calculated for sessions in which two behaviors are mutually exclusive, and one behavior was not observed by either observer (e.g., on-task). This occurred on 8 of the 43 (18%) sessions in which interobserver agreement was evaluated for student on-task behavior. If using exact agreement, \[ \frac{\text{agreements}}{\text{agreements} + \text{disagreements}} \times 100\% \] the observers would have 100% agreement on nonoccurrence of the behavior. When the corresponding Kappa statistic cannot be calculated, greater weight is given to the Kappa values from other sessions. In this case, those values were lower, which may mean this Kappa value (.50) is an underestimate of agreement for student on-task behavior.
Appendix M

Data Analysis: Percent of Nonoverlapping Data
Authors have used percentage of non-overlapping data (PND) as a metric to summarize effects across single-subject studies (Scruggs et al., 1987). For a behavior that should increase with intervention, PND is calculated by counting the number of data points in intervention which are above the highest data point in baseline, and dividing that number by the total number of data points in the intervention phase, then multiplying by 100%. An advantage of using PND to summarize an effect for a change between phases is that it gives a single number to summarize the effect, which could theoretically be compared to other PNDs. Other authors have described the limitations of using PNDs (e.g., Salzberg, Strain, & Baer, 1987). In this study, PNDs did not seem appropriate to summarize the data because they seemed particularly likely to greatly underestimate the effect in some cases (e.g., the decrease in variability from baseline to fluency training for positive to negative comments ratio was obvious with visual analysis, but would result in 0% PND), and overestimate effects in other cases (e.g., Ms. Allen’s presentation rate had an increasing slope in baseline, with a higher level, but decreasing slope in fluency training, which would result in 100% PND). PND might have been more appropriate for analyzing the level changes with error correction accuracy. However, visual analysis seemed to more readily accommodate the various patterns evident in the data in this study.
Breda Victoria O’Keeffe

Department of Special Education and Rehabilitation
Utah State University
2865 Old Main Hill Road.
Logan, Utah 84322
(435) 797-3608
b.okeeffe@aggiemail.usu.edu

EDUCATION

Currently enrolled in doctoral program in Disability Disciplines, Special Education,
Ph.D. anticipated July 2009
Dissertation Title: The Effects of Fluency Training on Implementation Fidelity of
a Reading Intervention Conducted by Paraprofessionals
Utah State University, Logan, Utah
Research Vice President Fellowship, $15,000 (2005-2006)

M.S., Special Education, 2001
National University, San Diego, California

B.A., Literature, Magna Cum Laude, 1996
Harvard-Radcliffe College, Cambridge, Massachusetts
Agassiz Award for superior academic achievement

PROFESSIONAL EXPERIENCE

2005-present  Research Fellow
Utah State University, Logan, Utah
Department of Special Education and Rehabilitation

2003-2005  Private Instructor
Verona, Wisconsin
Taught elementary students with disabilities reading, math,
spelling, and social studies in a 1:1 setting.

2002-2003  Special Education Teacher
Core Knowledge Charter School, Verona, Wisconsin
Taught students with mild to moderate disabilities in Kindergarten
through 8th grade.
2001-2002  
**Vocational Education Coordinator**  
*The Institute for Effective Education, San Diego, California*  
Provided professional development, curriculum development, program evaluation, and student employment placement assistance for vocational education department in special education school.

1999-2001  
**Instructor; Teaching Assistant**  
*The Institute for Effective Education, San Diego, California*  
Taught middle and high school students with cognitive, emotional and behavioral disabilities in academic, vocational, and social skills. Helped train and supervise teaching assistants.

1994-1996  
**Research Assistant**  
*Werner Sollors, PhD, Harvard, Cambridge, Massachusetts*  
Researched, proofread, and assisted in preparing literary manuscripts for publication. Helped prepare materials for academic lectures. Part-time student employment.

**PROFESSIONAL CERTIFICATION**

State of Wisconsin  
Special Education Teaching License, Cognitive Disabilities  
2004-2009

State of California  
Preliminary Level I Education Specialist Instruction Credential, Moderate/Severe Disabilities  
2001-2006

**PUBLICATIONS**


**CONFERENCE PRESENTATIONS**


**GRANTS**


**TEACHING EXPERIENCE**

Teaching Reading and Language Arts to Students with Disabilities, On campus and distance education (SPED 5310), (2005, Fall; 2006, Spring, teaching assistant; 2006, Fall, instructor).

Practicum Supervision for Reading and Language Arts to Students with Disabilities (SPED 5410), (2007, Fall).

Student Teaching Supervision in Special Education (2009, Spring)

Education of Exceptional Individuals (SPED 4000) (2008, Fall, teaching assistant; 2009, Spring, cotaught).

**SERVICE**

Guest Reviewer: Education and Treatment of Children (4 manuscripts)

Student Representative, Doctoral Program Committee, Utah State University (April 2008- May 2009)

Student Representative, Faculty Search Committee, Utah State University (April 2008- May 2009)

Consultation with Cache Valley Learning Center on treatment fidelity for Direct Instruction reading and math programs (2008, Fall)

Student Representative, Doctoral Student Recruiting Committee (2007, Fall; 2008, Spring)

Conducted training in *Language for Learning* with St. George School District (2007, Fall)

Consultation with Cache School District in evaluating implementation of Reading Success (2006, Fall)

Member, Master’s Thesis Award Committee (2006, Spring)

**PROFESSIONAL MEMBERSHIPS**

Association for Behavior Analysis

Association for Direct Instruction

Council for Exceptional Children