Title

Using Performance Feedback to Increase Procedural Integrity of Progress Monitoring Decision Rule

Permalink

https://escholarship.org/uc/item/9n17w64z

Author

Hickey, Rebecca Elizabeth

Publication Date

2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Using Performance Feedback to Increase Procedural Integrity
of Progress Monitoring Decision Rule

A Thesis submitted in partial satisfaction
of the requirements for the degree of

Master of Arts

in

Education

by

Rebecca Elizabeth Hickey

August 2014

Thesis Committee:
      Dr. Mike Vanderwood, Chairperson
      Dr. Cathleen Geraghty
      Dr. Cixin Wang

The Thesis of Rebecca Elizabeth Hickey is approved:

Committee Chairperson

University of California, Riverside

ABSTRACT OF THE THESIS

Using Performance Feedback to Increase Procedural Integrity
of Progress Monitoring Decision Rule

by

Rebecca Elizabeth Hickey

Master of Arts, Graduate Program in Education
University of California, Riverside, August 2014
Dr. Mike Vanderwood, Chairperson

Central to Response to Intervention (RTI) is the analysis of progress monitoring (PM) data to determine the need for an instructional change (IC). This study examined the extent to which performance feedback (PF) increased the procedural integrity (PI) of a decision rule, indicated by the percentage of correct decisions to make an IC or continue instruction. Three teachers received a brief training in regards to a decision rule. Teachers were asked to analyze PM data of students in their reading intervention group weekly and indicate whether an IC was necessary. The brief training led to initial relatively high levels of PI for two of the three teachers. However, the addition of PF was effective in increasing the percentage of correct IC decisions for two of the three teachers, as evidenced by visual analysis. The results suggest a brief training about the decision rule, monitoring of teachers' initial PI, and providing brief weekly PF to teachers with low to moderate percentages may improve data analysis within an RTI system.

Table of Contents

Chapter 1: Introduction

The most recent report from the National Assessment of Educational Progress (NAEP) states that a mere 34 % of fourth and eighth grade students in the U.S. performed at or above proficient in reading in 2013 (National Center for Education Statistics, 2013). This data suggests there is a necessity to improve performance in reading for students in the U.S.; Response to Intervention (RTI) has been proposed as a model to meet this need. RTI is an increasingly known and implemented multi-tier model in education that aims to ensure core curriculum meets the needs of most students and to identify students who are at risk for later reading failure and provide targeted intervention to those identified students. This is achieved through three tiers that increase in intensity of instruction. The critical components of RTI are universal screening, evidence-based interventions, and progress monitoring (PM) of students. In Tier I, about 80% of students needs should be met through the core curriculum, defined as the curriculum that covers standards that all students in that particular grade level receive. Universal screening is completed three times a year with each student in order to evaluate Tier I and identify students for Tier II (Reschly, 2008).

Ten to 15% of students identified as at risk for later reading failure through universal screening are provided with supplemental instruction (i.e. intervention) in Tier II, within a small group. For those receiving intervention, their progress is monitored (i.e. PM). Those students who do not make adequate growth in Tier II, based on the progress monitoring data, are moved to Tier III, where a more intense intervention is provided for about 5% of students. Progress monitoring continues in Tier III, occurring more

frequently than in Tier II (Tilly, 2008). As evidenced by the description of RTI above, the

analysis of PM data is critical in RTI, and is the focus of this study.

Progress Monitoring

Progress monitoring (PM) is largely influenced by Stanley Deno's (1985) work

with curriculum-based measurement (CBM), which is a reliable and valid measurement

system designed to measure student outcomes, as well as the Fuchs' (1986) research of

formative evaluation. A primary purpose of CBM is to guide teachers in improvement of

instructional planning (i.e. instructional changes; Deno, 1985; Fuchs, Fuchs, & Hamlett,

2007). In addition, PM has been impacted by the Fuchs' (1986) work with formative

evaluation, which focuses on continual evaluation of data, commonly using CBM, and

modification of programs. Fuchs and Fuchs (2011) have more recently described the use

of PM specifically.

Within an RTI framework, PM data are analyzed to determine the effectiveness of

the intervention. The conclusions made from the data analysis about effectiveness may

lead to a student moving to a more intensive tier in the RTI system. However, the primary

purpose of PM should be to inform the instruction the students are receiving in

intervention (Busch & Reschly, 2007; Fuchs & Fuchs, 2007). In other words, changes

should first be made to the current intervention before considering a more intensive

intervention. This continuous data collection using CBM, in order to modify

interventions to maximize student progress, is associated with improved academic

performance (Stecker, Fuchs, & Fuchs, 2005). Further, a seminal meta-analysis

examining the effects of formative evaluation on student achievement found that

teachers' use of formative evaluation with CBM and data evaluation techniques yielded

higher effect sizes over teacher judgment (Fuchs & Fuchs, 1986). Overall, the research

suggests that teacher analysis of PM data leads to increased student achievement due to

instructional changes (IC; Fuchs & Fuchs, 1986; Stecker et al., 2005).

Instructional Changes

There is little discussion of the nature of ICs in studies measuring the frequency

(e.g. Jenkins & Terjeson, 2011). However, recommendations about what ICs may be

include group reduction, increased time, more explicit instruction, and simplified

activities (Gettinger & Stoiber, 2012). However, teachers may have limited power to alter

the group size, the amount of time, and the intervention program; therefore, ICs focusing

on student engagement, specifically student-teacher interactions and delivery of

instruction can be used (Gettinger & Seibert, 2002). In regards to student-teacher

interactions, ICs include increased opportunities to respond, reinforcement of desired

behaviors (e.g. point system), and using specific and academically-relevant feedback.

Using cues to signal the start of an intervention or transition between activities and

challenging students to reduce the time it takes to prepare for a lesson and activities are

ICs related to the delivery of instruction (Gettinger & Seibert, 2002). In order to

determine the need for one of these ICs, however, a decision rule must be used.

Decision Rules

Two primary decision rules when analyzing PM data, slope rule and points below,

have been posited as ways to determine the necessity for an IC. Slope refers to the

calculation of the least squares regression slope of the student's data and comparing that

to the slope of the previously established goal line. Slopes that do not meet or exceed the target goal slope indicate a need for an IC. In regards to the points below rule, three consecutive scores below the goal line indicate a need for an IC. On the other hand, if any of the three most recent data points is above the goal line, current instruction should continue (Jenkins and Terjeson, 2011). Jenkins and Terjeson (2011) compared these two decision rules. Although the results suggest that using the slope method produced more frequent instructional changes, no student outcomes were compared to suggest which method produced higher student achievement, as these were hypothetical ICs. In addition, the researchers, rather than the teachers, were responsible for calculating the least squares slope and examining the number of ICs that would have occurred using each data analysis rule. Therefore, two major limitations of this study is the generalizability to teachers being responsible for the data analysis and the impact of those changes on student outcomes. In addition to these limitations, the points below method, rather than the slope method, may be a more practical for teachers.

Barriers to Data Analysis

Beyond the knowledge of decision rules, there are various documented barriers related to teacher data analysis. The U.S. Department of Education (2011) reviewed case studies of teachers' data analysis skills and found that many teachers make invalid decisions, particularly when trends are analyzed. In addition, it was found that teachers continued to use heuristics and maintain biases, despite documentation of those leading to incorrect conclusions in naturalistic observations. Further, teachers had substantial difficulty when working with individual data. Because there are documented difficulties

4

of data analysis for teachers, yet empirical support for increased student achievement through data analysis of progress monitoring and subsequent ICs, it is important to identify how school psychologists can increase teachers' correct use of decision rules with PM (U.S. Department of Education, 2011).

Professional Development and Coaching

Traditionally, professional development (PD), consisting of a single meeting, has been the means to improve teachers' instructional strategies (Kinkead, 2007). Kinkead (2007) discusses how this format of PD, specifically isolated training or informational session, without follow-up, has been documented as ineffective. In an effort to strengthen PD, coaching was suggested as a possible follow-up procedure. Coaching refers to model teachers working with colleagues to support PD and increase effective instruction. However, the results of coaching studies have produced mixed results (e.g. Carlisle & Berebitsky, 2010; McIntyre, Kyle, Chen, Munoz, & Beldon, 2010). Due to the inconsistent results within the coaching literature, a different follow-up method is necessary. An alternative follow-up procedure, which has been extensively studied to improve instruction, specifically treatment integrity (TI), and to a lesser extent procedural integrity (PI), is PF.

Performance Feedback

Performance feedback (PF) is the monitoring of a behavior that is the focus of concern and providing feedback to the individual, in regards to that specific behavior (Noell et al. 2005). Solomon, Klein and Politylo (2012) synthesized the single-case literature on PF targeting teacher TI. The results indicated that TI was low at the onset,

generally declined in phases prior to PF implementation, and increased following PF. PF was effective in preschool through high school and grade level did not alone significantly moderate the effectiveness of PF. Additionally, PF was more effective in general education and when targeting academic skills. Further, teachers have positively rated the acceptability of and satisfaction with the PF procedures (Pellecchia et al., 2011). The following sections discuss specific studies that provide empirical support for the use of PF to improve TI and PI.

Treatment integrity. Performance feedback (PF) has primarily been used to increase TI. TI refers to the extent to which teachers implement school-based interventions consistently and accurately (Gresham, 1989; Noell et al., 2000). The following sections summarize several studies that have documented the effectiveness of PF as a technique to increase TI following training.

Academic interventions. Noell and colleagues (2000) compared two follow-up methods after training teachers on peer-tutoring targeting reading comprehension, including a daily follow-up meeting in the absence of integrity or student data, and PF. The PF meeting included graphic displays of student performance and implementation data, identifying missed or incorrectly completed steps and how to improve. Teachers implemented at a mean integrity level of 41% during baseline. Two of the five teachers improved during the follow-up phase; however, four of the five teachers increased implementation during the PF phase. Additionally, every student's performance increased during the PF phase (Noell et al., 2000).

Behavioral interventions. Noell and colleagues (1997) studied the effectiveness of PF on treatment integrity for a reinforcement-based intervention implemented by general education teachers. Implementation of the daily PF by a consultant significantly improved TI and improved student academic performance for two of the three participating students. In a similar study, Noell and colleagues (2005) compared three behavioral consultation follow-up strategies for increasing TI, including a follow-up using an abbreviated Plan Evaluation Interview, a commitment emphasis condition where teachers were asked to commit to the intervention at the final meeting prior to implementation, and a PF condition. The analyses indicate that the PF condition had substantially higher levels of TI and student behavior improved significantly above the other conditions (Noell et al., 2005).

In an additional study with general education teachers and students with academic performance deficits, Witt, Noell, LaFleur, and Mortenson (1997) examined the effect of PF on the TI of a reinforcement-based intervention. PF resulted in significant increases in TI, as well as increases from pre-treatment (53%) to post-training baseline (71%) to maintenance (81%) in student academic performance. In a similar study, PF was effective in increasing the TI of a pre-referral, reinforcement-based intervention to increase academic performance. Three of the four dyads had improvements in TI (Mortenson and Witt, 1998). In examining the effects of PF on the implementation of individual Behavioral Support Plans (BSP) using a concurrent multiple baseline across participants design, the results from Codding, Feinberg, Dunn, and Pace (2005) suggest that PF is effective in increasing TI of individual BSPs in a special education setting. Also within

the special education population, DiGennaro, Martens, and Kleinmann (2007) studied the effectiveness of goal setting and PF regarding TI to increase student behavior. Each teacher increased in TI and all students exhibited reductions in target behavior (DiGennaro et al., 2007).

Data Collection. PF has also been shown to be effective in increasing data collection of teams for a target student in an autism classroom, extending the PF efficacy to team behavioral change. Further, data collection of those teams generalized to additional students in the classroom (Pellecchia et al., 2011).

Procedural integrity. As previously mentioned, PF has also been documented as effective in increasing PI. Within an RTI system, Burns, Peters, and Noell (2008) refer to PI as the extent to which components of the RTI process adhere to the proposed process. In addition to being among the first researchers to apply PF to PI, these authors have generated one of few articles studying the effectiveness of PF for the PI of a process within an RTI framework. The purpose of the study was to examine the effects of PF on the PI of the problem solving team (PST) meeting process (Burns et al., 2008).

The first school's PST included the teacher who referred the student, school psychologist, physical education teacher, two special education teachers, two general education teachers, and others who were necessary, such as a social worker. The referring teacher, school psychologist, social worker, speech and language pathologist, occupational therapist, physical education teacher, and special education teacher, as well as the principal inconsistently, were involved in the PST of the second school. All teachers from the grade of those students being discussed during a given week, assistant

principal, school psychologist, school counselor, speech and language pathologist, special education teacher, and a special education teacher certified to support students with emotional and behavioral disorders comprised the PST of the third school. A 20-item checklist was used during the observations of the PSTs (Burns et al., 2008). This checklist was previously developed based on identified activities for PSTs from the literature (Burns, Wiley, & Viglietta, 2008). Although the authors state the checklist may not reflect the optimal set of PST activities, the included activities could be used to test the effectiveness of PF to increase PI of PST meetings (Burns et al., 2008).

A multiple-baseline design across three schools within an urban school district was used.

The dependent variable was a percentage calculated by dividing the number of items observed by the total number of items on the checklist. The visual analysis, specifically the immediacy of the effects, and the percentage of non-overlapping data points (PND), revealed an effect for all three schools. The PND for school 1, 2, and 3 were 90.0%, 90%, and 100%, respectively. These results suggest that PF may be effective in the PI for the PST process, particularly within an RTI system. Additionally, this provides rationale to examine the effectiveness of PF when applied to other key processes of an RTI system (Burns et al., 2008).

Purpose

Past research supports the use of PF in increasing TI of academic (e.g. Noell et al., 2000) and behavioral interventions (e.g. Noell et al., 1997), and data collection (Pellechia et al., 2011). Further, Burns and colleagues (2008) provide initial evidence that

9

PF may be effective in increasing PI of components within an RTI system. Furthering

the use of PF to increase PI of critical features of RTI, this study will be amongst the first

to use PF to target PI of decision rules for determining the need for ICs. Therefore, the

purpose of this study is to examine the effectiveness of PF in increasing PI of a decision

rule to make ICs when analyzing PM data. More specifically, the following four research

questions will be studied: (1) To what extent do teachers use the decision rule correctly,

indicated by the percentage of correct IC decisions (i.e. PI), with only a brief training? (2)

To what extent is there a functional relationship between the use of PF and the accuracy

of the teachers' ICs based on a decision rule (i.e. PI)? (3) To what extent do teachers'

correct ICs, when the data indicates a need for an instructional change, differ from

baseline to treatment? Although similar to research question 1, this question is limited to

only when the data suggests an IC is necessary (4) To what extent do teachers rate PF to

be socially valid?

Chapter 2: Methods

Participants

   School. Teachers were recruited from an elementary school in southern California. The school consisted of approximately 680 students in kindergarten through sixth grade, with 10% who received special education, and 7% classified as English Language Learners. Approximately 41% of the students received free or reduced lunch. The majority of the students were Caucasian (66%), 31% were Hispanic, and Asian, American Indian, Filipino, and African American students each made up 1% of the school's population. In addition, the elementary schools within the district implemented RTI.

   Teachers. The participants were teachers who were responsible for implementing a small group reading intervention for students in Tier III. After obtaining IRB approval, each teacher was approached by the principal to ask if they would like to participate in the study, for which they were provided with a brief description of the responsibilities of participation. The principal clearly stated that the study was voluntary. From this recruitment process, three teachers agreed to participate in the study.

   Teacher A was a female kindergarten teacher with 19 years of teaching experience. She taught two intervention groups that varied in the number of children, comprised of 7 first and 4 second grade students in Tier III. Voyager Passport was implemented for both groups. Teacher B was a female kindergarten teacher with 21 years of teaching experience. Her intervention group contained eight third grade students in Tier III, and Voyager Passport was used. Teacher C was a female kindergarten teacher

with 20 years of teaching experience. She taught two intervention groups, a second and fifth grade Tier III group. Voyager Passport and Language! were implemented with the second and fifth grade groups, respectively.

Measures

      Progress monitoring. As was used by Noell and colleagues (2005), student performance was measured with CBM. Specifically, grade-appropriate Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency (DORF) was used for the progress monitoring data for grades second through sixth, as well as first grade students who read at least 25 words correct during the one minute test. DORF words read correctly (WRC) has been shown to have adequate psychometrics. DORF WRC has alternate-form reliability and test-retest reliability across grades above .90, and inter-rater reliability of 0.99. Predictive criterion validity across the grades included in the study are 0.60 and above (Good et al., 2013). In addition, DIBELS Nonsense Word Fluency (NWF) correct letter sounds (CLS) was used for students in 1st grade who were unable to reach 25 correct words per minute on DORF WRC. NWF CLS has a 1st grade two week alternate-form reliability of 0.88, a test-retest reliability of 0.90, and inter-rater reliability of 0.99. The predictive criterion validity of NWF CLS at winter and spring screening are 0.51 and 0.56, respectively, indicating moderately strong validity (Good et al., 2013).

      Procedural integrity of decision rules. The dependent variable, PI of the decision rule, was the percentage of correct IC decisions. The percentage was calculated by dividing the number of correct IC decisions by the number of opportunities to make a

correct IC decision (i.e. number of student graphs). A correct decision could be achieved

in the following two ways: data suggested the need for an IC, using the decision rule, and

the teacher indicated an IC, or the data suggested no IC was necessary and the teacher did

not make an IC. On the other hand, an incorrect decision was determined when the data

showed the need for an IC and the teacher failed to make an IC, or the data did not

support an IC, yet the teacher made an IC. The teachers' decision to make an IC or not

make an IC was indicated on the students' PM graphs, with the nature of the IC specified

on the back. Three of the weeks (approximately 27% of the weeks), an independent rater

determined the percentage of correct ICs for all participants. Interrater reliability was

calculated by dividing the number of agreed upon PI percentages by the total number of

opportunities to calculate PI and multiplying by 100%. This process resulted in 100%

interrater agreement of the percentage of correct instructional changes.

Consultation procedural integrity of performance feedback. Noell (2008)

broadly defines consultation procedural integrity (CPI) as the extent to which

consultation procedures are implemented as planned. The consultation procedure

implemented within this study was PF. A checklist of the steps in delivering the PF,

similar to Pellecchia and colleagues (2011), was used to measure the CPI of PF (Table 2).

The CPI for the PF procedures was 99%. Only one activity, scheduling the next meeting,

was not completed for only one week. An independent rater observed 40% of the PF

sessions and completed the PF checklist. Interrater reliability was calculated by

correlating the PI percentage of the consultant by the percentage of the independent

observer and multiplying by 100%. This process resulted in 100% interrater agreement for the CPI of the PF.

Teacher acceptability survey. Similar to Noell and colleagues (2005), a survey was developed to measure the social validity of the PF process (Table 2). The survey used a Likert scale ranging from 1 (Strongly Disagree) to 7 (Strongly Agree). Each of the participants was asked to complete the survey following the data collection.

Procedures

Training. Consent from the district was obtained prior to recruiting participants. The teacher consent was obtained on the day of training, which occurred after the teachers were informed of the study. The teachers received an initial 30 minute training that occurred in the school they worked in. The purpose of this training was to provide specific steps for the teachers to use when analyzing PM data, in order to determine the need for an IC. The following activities were reviewed: using the 3-day rule to decide the need of an IC; determining which of the previously mentioned ICs would be implemented; and graphing an IC.

Progress monitoring. Students were given one probe from the DIBELS Next measures, either NWF or DORF. The goals to create the goal lines on the PM graphs were based on the Spring DIBELS cut offs, based on the students' grade level. Staff from the school collected and graphed the PM data weekly, which were delivered to the teachers to analyze the data. Following the PF meeting, the students' graphs were collected from the teachers and returned to staff to complete the following week's PM.

Performance feedback. Each of the participating teachers received PF weekly at week 3, 6 and 9 for Teacher A, B, and C, respectively. The order was determined by a random selection, as the participants and principal asked to have an indication of when the meetings would occur. Although the majority of the PF literature occurred daily, weekly feedback leads to increases in TI and may be more practical than daily feedback (Mortenson & Witt, 1998). In most studies, PF was reviewed the following day or week, rather than the same day (Codding, Feinberg, Dunn, & Pace, 2005). Therefore, PF meetings were held the week after the intervention teachers reviewed the data and documented ICs. In addition, components such as goal setting and graphic displays have been found to increases the effectiveness of PF (Jones, Wickstrom & Friman, 1997; Martens, Hiralall & Bradley, 1997). Consequently, both student goals and graphs were used during the PF meeting.

Modeled after Noell and colleagues (2005), PF consisted of meeting briefly with the teacher, reviewing student graphs, and discussing ICs. Following the PF checklist, the consultant first provided praise for reviewing the data. Next, the students' graphs were displayed and reviewed by the consultant. For students that the steps were incorrectly completed, indicated by a lack of an IC that was needed or an unnecessary IC, the teacher was asked to explain the steps to analyze each of those students' PM data. The consultant then reviewed any steps of the decision rule that were completed incorrectly. The final activity was to schedule the next meeting on the following week.

Experimental Design

      This study employed a concurrent multiple baseline design across participants. The introduction of the PF for the participants were weeks 3, 6, and 9, respectively. The order in which the PF would be introduced to the teachers was randomly selected prior to beginning the study. This study was designed to meet What Works Clearinghouse (WWC) standards with reservations. The WWC standards with reservations state the researcher must systematically manipulate the independent variable, there must be interrater data collected for each outcome, three replications of an effect (e.g. three participants), and a minimum of three data points per phase (i.e. baseline and intervention phases; WWC, 2010). Although this study was designed to meet WWC standards with reservations, due to the school's decision to end intervention one week early, the third participant's intervention phase included only two, rather than three, data points.

Analyses

      Research question 1. To evaluate the first research question of teachers' use of the data analysis rule with only a brief training, the mean percentage of correct instructional changes were calculated for the baseline phase.

      Research question 2. To examine the functional relationship between PF and teachers' PI of decision rules, both visual analysis and effect size estimation were used.

      Visual analysis. Changes in the percentage of correct instructional changes will be examined using the traditional method of visual analysis. Visual analysis includes an examination of within-phase patterns and between-phase patterns. The within-phase patterns analyzed are level, trend, and variability (Horner et el., 2005). Level refers to the

average within a phase, which can be calculated using the mean or median (Kennedy, 2005). As suggested by Horner and colleagues (2005), the mean will be used to determine level in the present study. When inspecting the trend of the data, both slope and magnitude are considered. In regards to slope, the data can be identified as positive, flat, or negative. The second element, magnitude, refers to the extent of the slope, which can be considered high, medium, and low. The final within-phase pattern, variability, refers to the degree individual data points vary from the trend, and is considered high, medium, or low. The extent to which these three within-phase patterns differ in the baseline and intervention phases suggests the functional relationship between the dependent and independent variables (Kennedy, 2005).

The between-phase components that are examined in single-case research are immediacy, overlap, and consistency (Horner et al., 2005). The immediacy of the effect refers to how rapidly a change occurs from baseline to intervention. The change is considered rapid or slow; more rapid effects indicate a functional relationship. The second element is the overlap of the data points between phases; in general, less overlap is desired (Kennedy et al., 2005). The final between-phases pattern, consistency, examines the similarity across all phases (Horner et al., 2005). The less consistent the phases are from one another, the greater evidence of a functional relationship is suggested (Kennedy et al., 2005).

Effect size estimation. Percentage of All Non-Overlapping Data (PAND) and the Phi Procedure were used to evaluate the effect of PF on the percentage of correct instructional changes. PAND is calculated by dividing the total number of intervention

data points that overlapped baseline by the total number of data points collected across the baseline and intervention phases. This resulting number is then subtracted from 1.0 and multiplied by 100 to compute the percentage. For example, if a study contains 5 baseline data points, and 15 intervention data points, and 4 intervention data points overlapped the baseline data points, 4 would be divided by 20, or .20. In this case, after subtracting from 1.0 and multiplying by 100, the PAND is 80%. The PAND will be used to determine the effectiveness of PF. A benefit of using PAND is the ability to convert this to the common effect size Phi Parker, Hagan-Burke & Vannest, 2007; Riley-Tillman & Burns, 2009).

Phi is used to examine the practical effect of the intervention. Phi is calculated by using a 2x2 table (Table 3), and using the following equation:

$$[\, a \, / \, (\, a + c \,) \,] - [\, b \, / \, (\, b + d \,) \,]$$

(Parker et al., 2007; Riley-Tillman & Burns, 2009). Based on Phi, the effects are determined to be small (.10), medium (.30), or large (.50; Cohen, 1988). There is concern around calculating an effect size with a study using single case design due to autocorrelation, which inflates the effect size. Phi overcomes the problem of autocorrelation . Although, there remains a concern for insensitivity of magnitude of large effects and baseline trend is not considered (Riley-Tillman & Burns, 2009). Due to these concerns, the visual analysis was used as the primary analysis.

Research question 3. The means of the correct number of times a teacher implemented an IC when the data analysis rule suggested the need for an IC were

calculated for both phases across each participant. A t-test was conducted to test whether the means were significantly different.

      Research question 4. The means for each question and overall on the Teacher Acceptability Survey were calculated to determine the social validity of PF.

Chapter 3: Results

Research Question 1

The first research question regarded the extent to which teachers use a decision rule correctly to analyze PM and determine the need for an IC (i.e. PI), after receiving only a brief training. The average PI across participants during the baseline phase was 73.4%, ranging from 22.1% to 85.8%. This suggests teachers' skills in using a decision rule to determine the need for an IC after a brief training varies greatly, with one teacher having adequate PI.

Research Question 2

Visual analysis. Results of the effectiveness of PF for PI of the decision rule are presented in Figure 1. In single case design, WWC (2010) requires that three data points be collected within each phase for all participants. Clearly, only two data points were collected for Teacher C in the PF phase; therefore, the conclusions from this participant are limited. Further, WWC states there must be three replications to conclude a treatment is evidence-based. Without the inclusion of Teacher C, this study cannot meet this standard. In addition, a slight positive trend was observed for Teacher A during baseline, yet a stable baseline is needed. As will be discussed, other components of visual analysis provide support that PF had some effect on PI for Teacher A. Despite these limitations, analysis of the data from this study continues to be informative, particularly for future research.

A positive effect on level was seen for each teacher, according to the mean increasing substantially. Teacher A, B, and C increased from 22.1% to 93.9%, 85.8% to

100%, and 82.2% to 100%, respectively. A positive effect of trend was seen for Teacher A, in which the slope during baseline was relatively flat, while clearly positive during the PF phase. The variability decreased across each participant from baseline to the PF phase. Teacher A ranged from 8.3% to 33.3% during baseline phase and 75% to 100% during the PF phase, narrowing the variability. Teacher B decreased from 78% to 100% to no variability, with 100% PI during the PF phase. Similarly, Teacher C ranged from 50% to 100% during baseline, and had no variability during the PF phase. When considering the variability across participants, there was a substantial decrease. During the baseline phase, the PI ranged from 8.3% to 100%; during the PF phase, the PI ranged from 75% to 100%. As is evident, the range tightened significantly towards higher percentages for each participant and across all participants.

A rapid change (i.e. immediacy), as well as no overlap, was observed with Teacher A. Due to both Teacher B and C making 100% in at least one week during the baseline phase, there was high overlap and difficulty observing immediacy for both participants. In regards to consistency, little consistency was observed across the two phases for Teacher A and B. Taken together, these results suggest that the use of PF was effective in increasing PI for Teacher B, and to an extent Teacher A. However, as previously mentioned, there was a slight positive trend in baseline phase for Teacher A. Further, too few data points were collected for Teacher C to conclude effectiveness.

Effect size estimations. PAND for the percentage of correct ICs was 75%, which resulted in a Phi coefficient of 0.59. Based on Cohen's (1988) guidelines, there was a large effect.

Research Question 3

The PI for the baseline and treatment phases, when considering only when the need for an IC was suggested based on the decision rule, were 92.89 and 94.48, respectively. Although the mean during the treatment phase was greater than the baseline phase, the t-test indicates the means were not significantly different ($p = 0.87$).

Research Question 4

An attempt was made to collect the Teacher Acceptability Survey from each teacher. The documents were given to each teacher at the conclusion of the data collection, and teachers were asked to return it to the school psychologist's box. After collecting one survey, each of the remaining teachers was given an additional copy. A third attempt to collect the survey was made by emailing each of the teachers asking them to complete the survey electronically. However, only one survey was completed and collected from Teacher B. Teacher B rated "6" or "7" on the Likert scale ranging from "1" to "7" for each of the five statements. The mean rating across the questions was 6.8.

Chapter 4: Discussion

Research examining the effect of PF suggests it is effective for increasing TI of academic (e.g. Noell et al., 2000) and behavioral interventions (e.g. Noell et al., 1997), and data collection (Pellechia et al., 2011) . Further, Burns and colleagues' (2008) study examining the effectiveness of PF for increasing PI of the PST provide initial evidence that PF may be effective in increasing PI of various components within an RTI system. To further study the use of PF to increase PI of critical features of RTI, this study is amongst the first to use PF to target PI of decision rules for determining the need for ICs. The purpose of this study was to examine the effectiveness of PF in increasing PI of a decision rule to make Ics when analyzing PM data.

The four specific research questions of this study were as follows☹1) To what extent do teachers use the decision rule correctly, indicated by the percentage of correct IC decisions (i.e. PI), with only a brief training? (2) To what extent is there a functional relationship between the use of PF and the accuracy of the teachers' Ics based on a decision rule (i.e. PI)? (3) To what extent do teachers' correct Ics, when the data indicates a need for an instructional change, differ from baseline to treatment? This is similar to question 1; however, this question is limited to only when the data suggests an IC is necessary (4) To what extent do teachers rate PF to be socially valid?

Brief Training and Procedural Integrity

The first purpose of this study was to examine the PI of a decision rule following a brief training. Past research of traditional professional development, which refers to an isolated meeting without follow up for the purpose of training individuals, suggests a

single training is ineffective in behavioral change of teachers' instructional practices (Kinkead, 2007). Further, the U.S. Department of Education (2011) found that teachers had substantial difficulty with analyzing individual data. The results from this study partially replicate this finding, as there was a large amount of variation between teachers' PI following training, with some having high PI without PF. Specifically, only teacher A was low at the onset after the initial training, which replicates findings that one meeting is insufficient for changes to instruction; on the other hand, both Teacher B and C had relatively high PI following the initial training and remained high within the baseline phase. In practical terms, this suggests that some teachers can successfully analyze data with only a brief training. However, other teachers may need additional support from the school psychologist.

Performance Feedback and Procedural Integrity

The primary purpose of this study was to examine the effectiveness of PF on PI. Based on the visual analysis, a functional relationship was demonstrated between PF and increases in the percentage of correct IC decisions for Teacher A and B. In addition, a Phi coefficient of 0.59 indicates PF had a large effect on the percentage of correct IC decisions. However, the data indicates that PF may be more effective for teachers that have lower initial PI, specifically Teacher A.

In the previously described meta-analysis of PF, conducted by Solomon and colleagues (2012), the results suggest TI was low at onset and decreased prior to implementation of PF. This was not observed for the PI in the case of any of the three participants. Further, this study partially replicates previous research with PF to increase

TI of academic (Noell et al., 2000) and behavioral interventions (Codding et al., 2005; DiGennaro et al., 2007; Noell et al., 1997; Noell et al., 2005; Mortenson & Witt, 1998; Witt et al., 1997), and data collection (Pellechia et al., 2011). In the first study with PI as the outcome and the application of PF, Burns and colleagues (2008) examined the effectiveness of PF to increase PI of the PST process, within an RTI framework. The results indicate that PF can increase PI of the PST process, an important process in an RTI system. The results from the present study also examined a critical component to RTI, data analysis, particularly of PM data. The effect of PF observed for Teacher A and B provide initial evidence. However, due to methodological concerns, replication will be necessary within a more methodologically rigorous study.

The effectiveness of PF on PI of the decision rule was specifically examined in relation to times when the data suggested a need for an IC. The mean during the treatment phase was greater than the baseline phase. However, the means were not significantly different. Upon further analysis of the data, incorrect decisions were observed to occur more frequently when teachers indicated an IC, while the data suggested there was no need. This supports the finding from the U.S. Department of Education (2011) that despite incorrect conclusions being drawn, teachers continue to use heuristics and maintain biases. For example, during PF meetings, teachers would defend a decision based on external factors, such as a student being from a difficult family.

Performance Feedback and Social Validity

In regards to the fifth research question, although only one completed survey was collected, the ratings indicate that the PF related to data analysis is socially valid. This is

similar to past research findings on the social validity of PF (Noell et al., 2005; Pellechia et al., 2011).

Implications

The variation found amongst the teachers in regards to PI of the decision rule after a brief training suggests a brief training on analyzing individual PM data may be sufficient for some teachers to correctly use a decision rule to determine the need for an IC. Further, by monitoring teacher PI of IC's and providing brief PF weekly, teachers with initially low and moderately high PI can increase the percentage of correct IC decisions to an acceptable level. This has important implications for an RTI system. Data analysis of PM is critical within an RTI framework to determine the need for changes of an intervention (i.e. IC; Busch & Reschly, 2007; Fuchs & Fuchs, 2007). The findings from this study imply that a more effective data analysis for Tier II and Tier III within a multi-tiered system can be achieved through a brief training of a decision rule, monitoring of IC decisions and brief weekly PF for teachers who maintain low PI of IC decisions. In addition, the procedures used for PF, as indicated by past research and the present study, are socially valid.

Limitations

There are limitations in this study that restrict the conclusions to be drawn about the effectiveness of PF and the generalizability. WWC (2010) requires there be three replications within a single case study to conclude effectiveness of a treatment. However, because only one data point was collected during the PF phase for Teacher C, only two replications could be analyzed. Although PF may have been effective for Teacher C, had

the intervention continued as anticipated, two intervention data point are insufficient to make conclusions about the effectiveness of PF. Additionally, the baseline for Teacher A was not stable, with a slight positive trend, when PF was introduced. This is a concern because the effect from baseline to treatment may not be due to the PF. A final limitation of this study was that the social validity measure was collected from only one teacher. Although past research indicates that PF is socially valid, as evidenced by positive ratings by teachers receiving PF, a social validity measure from all participants would have provided more support for the social validity.

Future Research

Studies employing a single case design should ensure all the standards are met according to WWC (2010), particularly the collection of three data points in each phase for all participants and to produce three demonstrations of effectiveness. In addition, future research examining PF should produce studies using other methodologies (i.e. randomized controlled trial). A distal goal in PF research with data analysis, if effectiveness is established, should be exploring the joint effect of PF targeting both TI of interventions and PI of a decision rule simultaneously. Finally, in future research, individuals should explore the effects of PF on the PI of various key RTI components and processes.

Chapter 5: References

Burns, M. K., Peters, R., Noell, G. H. (2008). Using performance feedback to enhance implementation fidelity of the problem-solving team process. *Journal of School Psychology, 46*, 237-550.

Burns, M.K., Wiley, H. I., & Viglietta, E. (2008). Best practices in facilitating problem-solving teams. In A. Thomas, & J. Grimes (Eds.), Best practices in school psychology (pp. 1633−1644)., 5th ed. Bethesda, MD: National Association of School Psychologists.

Busch, T. W., & Reschly, A. L. (2007). Progress monitoring in reading: Using curriculum-based measurement in a response-to-intervention model. *Assessment for Effective Intervention, 32*(4), 223-230.

Carlisle, J.F. & Berebitsky, D. (2010). Literacy coaching as a component of professional development. *Reading and Writing, 24*, 773-800.

Codding, R. S., Feinburg, A. B., Dunn, E. K., & Pace, G. M. (2005). Effects of immediate performance feedback on implementation of behavior support plans. *Journal of Applied Behavior Analysis, 38*(2), 205-219.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). New Jersey: Lawrence Erlbaum.

DiGennaro, F. D., Martens, B. K., & Kleinmann, A. E. (2007). A comparison of performance feedback procedures on teachers' treatment implementation integrity and students' inappropriate behavior in special education classrooms. *Journal of Applied Behavior Analysis, 40*(3), 447-461.

Deno, L. S., Lembke, E., & Reschly, A. (n.d.). *Progress Monitoring: Study Group Module*. Department of Special Education, University of Minnesota, Minneapolis.

Deno, L. S. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, *52*, 219–232.

Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children, 53*(3), 199-208.

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

Fuchs, L. S., & Fuchs, D. (2011). *Using CBM for progress monitoring in reading* National Center on Student Progress Monitoring. 1000 Thomas Jefferson Street NW, Washington, DC 20007.

Fuchs L. S., Fuchs D., & Hamlett, C. L. (2007). Using curriculum-based measurement to inform reading instruction. *Reading and Writing 20*, 553-567.

Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research & Practice, 18*(3), 157-171.

Gersten, R., Compton, D., Connor, C.M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W.D. (2008). Assisting students struggling with reading: Response to Intervention and multi-tier intervention for reading in the primary grades. A practice guide. (NCEE 2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved October 15, 2009 from http://ies.ed.gov/ncee/wwc/publications/practiceguides/.

Gettinger, M., & Seibert, J. K. (2002). Best practices in increasing academic learning time. *Best Practices in School Psychology.* Bethesda, MD.

Gettinger, M., & Stoiber, K. C. (2012). Curriculum-based early literacy assessment and differentiated instruction with high-risk preschoolers. *Reading Psychology, 33*(1-2), 11-46.

Good, R.H., Kaminski, R.A., Dewey, E.N., Wallin, J., Powell-Smith, K.A., & Latimer, R.J. (2013). DIBELS Next Technical Manual. Eugene, OR: Dynamic Measurement Group. Available: http://dibels.org/

Gresham, F.M. (1989). Assessment of Treatment Integrity in School Consultation and Prereferral Intervention. *School Psychology Review, 18(1),* 37-50.

Jenkins, J., & Terjeson, K. J. (2011). Monitoring Reading Growth: Goal Setting, Measurement Frequency, and Methods of Evaluation. *Learning Disabilities Research & Practice (Blackwell Publishing Limited)*, *26*(1), 28-35.

Jones, K.M., Wickstrom, K.F., Friman, P.C. (1997). The effects of observational feedback on   treatment integrity in school-based behavioral consultation. *School Psychology Quarterly, 12(4)*, 316-326.

Kinkead, S. (2007). *Improving instruction through coaching.* Olympia, WA: The Center For Strengthening the Teacher Profession, Superintendent of Public Instruction, Washington.

McIntyre, E., Kyle, D., Chen, C., Munoz, M., & Beldon, S. (2010). Teacher learning and ell reading achievement in sheltered instruction classrooms: Linking professional development to student development. *Literacy Research and Instruction,* 49(4), 334-351.

National Center for Education Statistics (2013). The Nation's Report Card: A First Look: 2013 Mathematics and Reading (NCES 2014-451). Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

Noell, G.H. (2008). Research examining the relationships among consultation process, treatment integrity, and outcomes. In W.P. Erchul & S.M Sheridan (Eds.), *Handbook of research in school consultation: Empirical foundations for the field* (pp. 323-342). Mahwah, NK: Erlbaum.

Noell, G. H., Duhon, G. J., Gatti, S. L., & Connell, J. E. (2002). Consultation, follow-up and implementation of behavior management interventions in general education. *School Psychology Review*, 31, 217-234.

Noell, G. H., & Witt, J. C. (1999). When does consultation lead to intervention implementation?: Critical issues for research and practice. *The Journal of Special Education, 33*, 29-35.

Noell, G. H., Witt, J. C., Gilbertson, D. N., Ranier, D. D., & Freeland, J. T. (1997). Increasing teacher intervention implementation in general education settings through consultation and performance feedback. *School Psychology Quarterly, 12*(1), 77-88.

Noell, G.H., Witt, J.C., LaFleur, L.H., Mortenson, B.P.., Ranier, D.D., & LaVelle, J. (2000). Increasing intervention implementation in general education following consultation: A comparison of two follow-up strategies. *Journal of Applied Behavior Analysis, 33(3),* 271-284.

Noell, G. H., Witt, J. C., Slider, N. J., Connell, J. E., Gatti, S. L., Williams, K. L., et al. (2005). Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. *School Psychology Review, 34*(1), 87-106.

Parker, R. I., Hagan-Burke, S., & Vannest, K. (2007). Percent of all non-overlapping data (PAND): An alternative to PND. *Journal of Special Education, 40*, 194–204.

Pellecchia, M., Connell, J. E., Eisenhart, D., Kane, M., Schoener, C., & Turkel, K. (2011). We're all in this together now: Group performance feedback to increase classroom team data collection. *Journal of School Psychology, 49*, 411-421.

Peterson, L., Homer, A., & Wonderlich, S. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis, 15*, 477-492.

Reschly, D. J. (2008). School Psychology Paradigm Shift and Beyond. In A. Thomas & J. Grimes (EDS) Best practices in school psychology V (pp. 3-14). Bethesda, MD: NASP.

Riley-Tillman, T.C. & Burns, M. K. (2009). Evaluating Educational Interventions: Single-Case Design for Measuring Response to Intervention. New York, NY: The Guilford Press.

Solomon, B. G., Klein, S. A., Politylo, B. C. (2012). The effect of performance feedback on teachers' treatment integrity: A meta-analysis of the single-case literature. *School Psychology Review, 41*(2), 160-175.

Stecker, P.M., Fuchs, L.S., & Fuchs, D. (2005). Using CBM to improve student achievement: A review of research. *Psychology in the Schools, 42*, 795-818.

U.S. Department of Education, Office of Planning, Evaluation and Policy Development. (2011). Teachers' Ability to Use Data to Inform Instruction: Challenges and Supports, Washington, D.C.

What Works Clearinghouse. (2010). *Procedures and Standards Handbook (Version 2.1)*. Washington, DC: Institute of Education Sciences.

Witt, J.C., Noell, G.H., LaFleur, L.H., & Mortenson, B.P. (1997). Teacher use of interventions in general education settings: Measurement and analysis of the independent variable. *Journal of Applied Behavior Analysis, 30*(4), 693-696.

Chapter 6: Tables and Figures

Table 1

Performance Feedback Procedural Integrity

Introduction to meeting	
Provided praise for reviewing the data	
Presented graphic display of all students	
Reviewed whether the teacher indicated an instructional change for each student	
Asked the teacher to explain how he or she determined an instructional change was necessary or not for those completed incorrectly	
Reviewed decision rule step or graphing instructional changes, if completed incorrectly	
Scheduled next meeting	
Stayed within time limit (10 minutes)	

Table 2

Teacher Acceptability Survey

Statement	Strongly Disagree		Slightly Disagree		Slightly Agree		Strongly Agree
The consultant worked with me to analyze data in a supportive manner	1	2	3	4	5	6	7
The consultant provided problem solving strategies to overcome identified barriers to data analysis	1	2	3	4	5	6	7
I found the feedback sessions helpful	1	2	3	4	5	6	7
Weekly data analysis regarding students' progress is important	1	2	3	4	5	6	7
I was satisfied with the feedback sessions	1	2	3	4	5	6	7

Table 3

Phi Effect Size Calculation Table

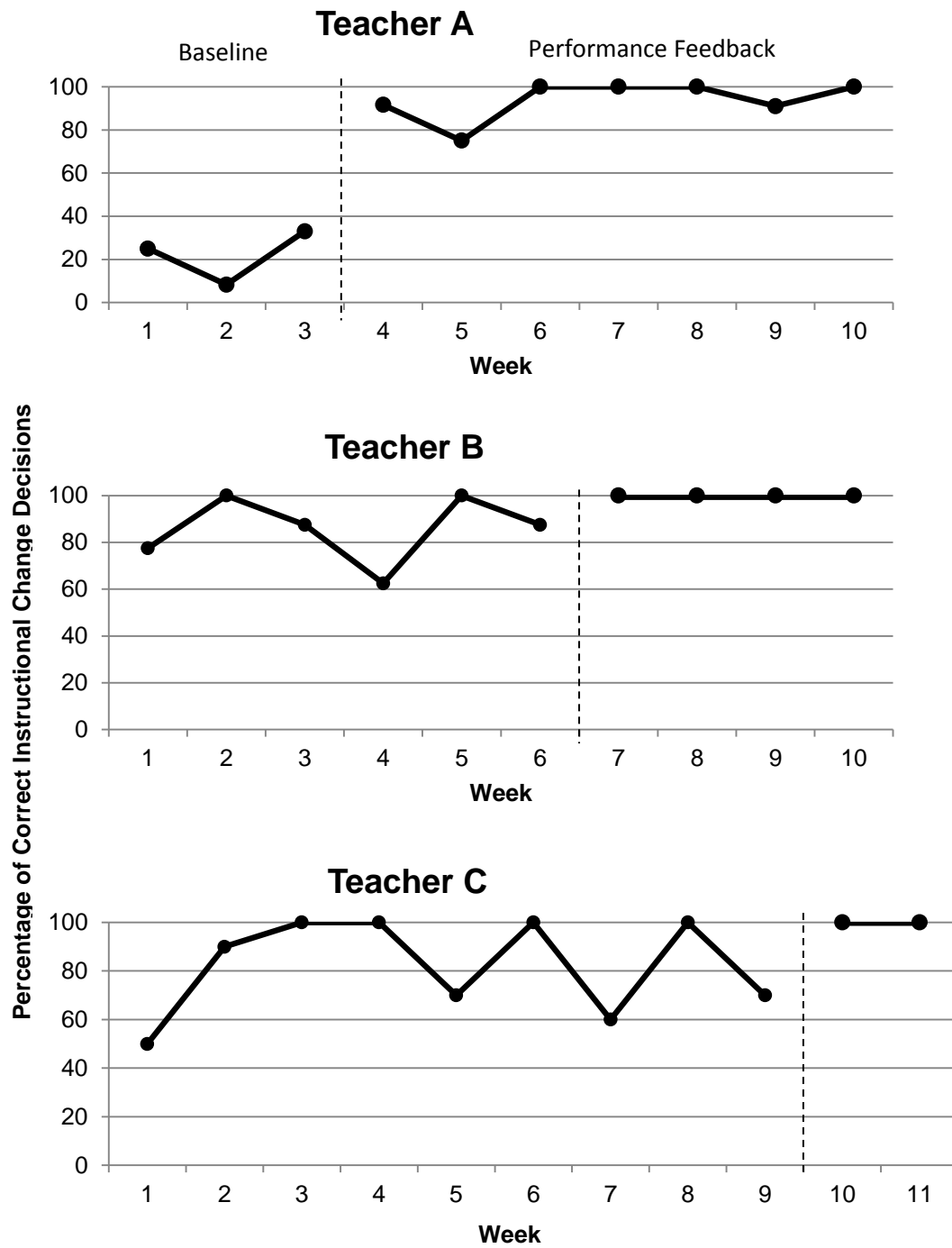|	
Cell A: % of baseline (baseline/total data points)	Cell B: % of overlapping data/2
Cell C: % of overlapping data/2	Cell D: % of treatment (treatment/total data points)

Figure 1. Teachers' Percentage of Correct Instructional Change Decisions by Week