Effects of a Collaborative Intervention on the Quality of Preservice Teachers' Data Based Decision Making

A Dissertation

SUBMITTED TO THE FACULTY OF

UNIVERSITY OF MINNESOTA

BY

Jennifer A. Wilson

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Susan Hupp, Adviser

October, 2013

Acknowledgements

Many people have influenced my life, career path, and this dissertation study. First, my gratitude goes to my advisor, Dr. Susan Hupp, who provided patience and mentoring throughout my graduate program. I thank my committee members Dr. Kristen McMaster and Dr. Kyla Wahlstrom for their direction and feedback during the dissertation process. I want to especially acknowledge committee member, Dr. Jennifer McComas, for encouraging me throughout this program and giving me both educational and personal advice.

I want to thank my family for their unwavering support during this process. My deepest gratitude goes to my husband Juan and my daughter Jayda for their love and understanding. Juan, you cooked a lot of dinners, wiped a lot of tears, and celebrated along with me as WE went on this endeavor. There is absolutely no way I could have done this without your support. Jayda, you always make me smile. You often asked if I had to “work again” or “go to class” and I’m excited to say we are almost done! I also want to thank my parents and my sister for their encouragement. You have supported my goals and dreams with enthusiasm for as long as I can remember.

This experience would not have been complete without my editors and friends, Amy Kunkel and Naomi Rahn. Your dedication and kind words have helped. And thank you to my “mentors” in this study – Dana Brandes, Andrea Egan, Meredith Peterson, Naomi Rahn, and Anne Zaslofsky. Thank you to the class instructors, Amy Kunkel and Quannah Parker, and the preservice teachers who participated in this study. I also appreciate Stephanie Snidarich for doing the inter-rater reliability. It has been a fun trip surviving graduate school with all of you!

Additional colleagues have also helped support me through this process and I am grateful: Christine Peper, Doug Marston, Don Healy, Ryan Gibbs, Morgan Stangl, Kara Lucas, Pari Beyzavi, Mary Lindell, and Laura Le.

Dedication

This dissertation is dedicated to my favorite campers, who inspired me to enter this field.

Abstract

Effective teaching practices, including the ability to make data-based decisions, are necessary to close achievement gaps. The purpose of this study was to investigate the effects of a collaborative, data-based decision making (DBDM) intervention on the quality of preservice teachers' DBDM. Participants were 45 preservice general educators enrolled in a teacher education course required for elementary teacher licensure. An experimental group design was used to investigate the effects of the intervention on the quality of preservice general educators' data-based decisions. Participants were randomly assigned to either the experimental condition, which was a 75-min, small-group, collaborative intervention on DBDM, or the control condition. Data were analyzed using a Mann Whitney U Test. The results showed that the intervention influenced the quality of preservice teachers' DBDM and that this collaboration influenced the confidence levels of preservice teachers with regard to DBDM.

Table of Contents

List of Tables

Chapter 1

Introduction

Statement of the Problem

As achievement gaps widen and the push for highly qualified and effective teachers intensifies, it is important that teachers have the tools to provide impactful instruction to all children. Accountability requirements from No Child Left Behind (NCLB; 2001) have increased the pressure on educators to close achievement gaps. Gaps, however, are continuing to widen between students living in poverty, students of color, and students receiving specialized services such as special education and English Language Learner services and their peers. Results from a comparison of achievement outcomes in reading and math prior to and after the enactment of NCLB are alarming (Lee & Reeves, 2012). Post-NCLB reading outcome trends show that average achievement, as well as the pace of achievements gains, has stayed the same or declined while maintaining or accelerating in math (Lee & Reeves, 2012).

Results from the Nation's Report Card (Institute of Education Sciences (IES; 2011), National Center for Education Statistics (NCES; 2011) suggest that, of students in fourth grade who performed at the "basic level" (below the 25th percentile) on the National Assessment of Educational Progress (NAEP) reading assessment, 74% were eligible for free/reduced price school lunch. Furthermore, of fourth graders who scored above the 75th percentile on the reading assessment, 71% were White, 7% were Black, 11% were Hispanic, and 8% were Asian. Of those students who scored above the 75th percentile, only 2% were English Language Learners (ELL; IES, NCES, 2011).

Similarly, among eighth graders who scored below the 25th percentile, 67% were eligible for free/reduced price school lunches, and of those who scored above the 75th percentile, 72% were White, 6% were Black, 11% were Hispanic, and 8% were Asian (IES, NCES, 2011).

Teachers play an important role in impacting these students and their achievement; educators must have the knowledge, skills, and ability to implement evidence-based instruction (Taylor, Pearson, Peterson, & Rodriguez, 2003). Fuchs, Fuchs, Hamlett, Phillips, and Bentz (1994) offered evidence that assessment data, especially when coupled with instructional recommendations, may increase student achievement. These researchers randomly assigned 40 teachers into three groups: Curriculum Based Measurement (CBM) with classwide reports and instructional recommendations, CBM with classwide reports but no instructional recommendations, and contrast (no CBM). They found that CBM strategies designed with classwide reports and instructional recommendations resulted in improved student achievement. There needs to be a linkage between the evidence-based instruction teachers are using and student assessment data (Mercier Smith, Fien, Basaraba, & Travers, 2009). However, there is a failure to link teaching strategies with student achievement (Foorman, 2007; Foorman & Schatschneider, 2003) and teachers do not have the appropriate data or tools to plan and individualize that instruction (USDOE, 2007).

Definitions

Key terminology will be used throughout this paper. It is important that a conceptual understanding of response to intervention (RtI), data-based decision making

(DBDM) and professional learning communities (PLCs) is understood so that the purpose and intervention are explicit and the results are clear.

Response to intervention. Response to intervention (RtI) is considered the "systematic use of assessment data to most efficiently allocate resources in order to enhance student learning for all students…" (Jimerson, Burns, & VanDerHeyden, 2007, p. 4). Components of RtI include DBDM, quality core instruction, tiered intervention, and a focus on enhanced learning (Burns & Gibbons, 2008). In their Institute of Education Sciences (IES) paper, Gersten et al. (2009) provide five recommendations for implementing RtI: (a) screen students for potential problems; (b) provide differentiated instruction (tier one); (c) provide intensive systematic instruction (tier two); (d) monitor student progress (tier two); and (e) provide additional intensive intervention if not responding (tier three). Accountability requirements from NCLB have increased the pressure on educators to provide these RtI components. The Individuals with Disabilities Education Act (IDEA; 2004) has also deemed the practice of examining student data in response to instruction as acceptable for making high stakes decisions about special education (Mercier Smith et al., 2009).

Data-based decision making. Data-based decision making (DBDM) is a hallmark of special education that dates back to the 1930's when Monroe argued for systematic analysis of reading errors (Fuchs, Fuchs, & Stecker, 2010). Deno (1985) stated "Measurement of student achievement is basic to evaluating the success of our educational program" (p. 219). The term DBDM, defined by the United States Department of Education Office of Planning, Evaluation and Policy Development, is "the analysis and use of student data and information concerning educational resources and processes to inform planning, resource allocation, student placement, and curriculum and instruction" (U.S. Department of Education, Office of Planning, Evaluation and Policy Development [USDOE], 2007, p. 1). Furthermore, the practice includes regular data collection and ongoing implementation of a continuous improvement process (USDOE, 2007).

Five explicit components of DBDM are offered by the USDOE (2011): (a) data location, (b) data comprehension, (c) data interpretation, (d) data use, and (e) question posing. USDOE identified and defined these skills and concepts for an interview study with over 180 teachers and 35 administrators or specialists (e.g., data coach). Personnel at the USDOE define data location as the ability to find the right data to use from a complex graph or table and to manipulate the data to support reasoning. Data comprehension is “figuring out what the data says” (p. 8). The target skills for data comprehension are moving fluently between different representations of data (e.g., understanding data points on a table and how they relate to visual inspection on a graph), understanding a histogram, interpreting a contingency table, and distinguishing between cross-sectional and longitudinal data. Data interpretation is going beyond the comprehension of the data by interpreting the meaning of the data. School personnel must be able to examine score distributions, understand the effect of outliers, understand the relationship between sample size and generalizability, and understand the concept of measurement error. Data use is the ability to apply the data to planning instruction. Educators need to use subscale and item data in addition to understanding how to differentiate instruction based on data and synthesizing multiple sources to inform practice. The final area is question posing (i.e., figuring out questions that will generate useful data). Target skills are aligning questions with purpose and data, forming queries that lead to actionable data, and appreciating values of multiple measures. This final area is important for district and school level leadership teams.

Professional learning communities. In many districts, PLCs are the main way professional development (PD) is provided (Vescio, Ross, & Adams, 2008). Eaker, DuFour, and DuFour (2002) argued that the most promising strategy for school improvement (e.g., closing the achievement gaps) is the capacity of school personnel to function as a PLC. DuFour, Eaker, and DuFour (2005) described PLCs as a collaborative group of educators who ensure student learning with a focus on results. “The concept of a PLC rests on the premise of improving student learning by improving teaching practice” (Vescio et al. 2008, p. 82). DuFour et al. (2005) emphasized that schools often suffer from the “DRIP syndrome—Data Rich/Information Poor” (p.40); it is imperative that PLC teams turn data into useful and relevant information for planning instruction and meeting students’ needs.

The core components of PLCs, as summarized by Vescio et al. (2008), are (a) shared values and norms, (b) clear and consistent focus on student learning, (c) reflective dialogue and continuing conversations, (d) deprivatized practice (i.e., making teaching public), and (e) a focus on collaboration. Murphy and Lick (2005) added to those core components by explaining that (a) the shared values and norms are about children, teaching, and learning, (b) the focus on student learning drives decisions, (c) the reflective dialogue allows for teachers to help view the practice from another's perspectives, (d) deprivatizing practice allows for teachers to practice their 'craft' openly to problem solve together, and (e) the collaboration should exist across grade-level groups. DuFour et al. (2005) suggested that the shift from focusing on teaching to focusing on learning has profound implications. Educators must be able to explain how they will respond when a student experiences difficulty learning. Their response must be systematic, timely, based on intervention, and directive (i.e., require students to devote time in areas in which they need additional help (DuFour et al., 2005).

Purpose of the Study

Components of a response to intervention (RtI) model, making data-based decisions, and improving instruction through PLCs are strategies that hold promise of assisting schools to close the achievement gaps. Educators are on the front lines making instructional decisions whether or not data are involved (U.S. Department of Education, Office of Planning, Evaluation and Policy Development [USDOE], 2011). Furthermore, the USDOE (2007) reported that teachers do not have appropriate data or tools to plan and individualize instruction. The purpose of this study is to teach preservice teachers to make data-based decisions through a collaborative Professional Learning Community (PLC)-type approach. The effects of this collaborative approach will be measured to determine the quality of preservice teacher's DBDM skills.

Chapter 2

Literature Review

Given the purpose of this study, I will review three areas of literature: data-based decision making, preservice teacher preparation, and inservice teacher professional development (i.e., PLCs). Strengths and limitations of the research studies reviewed in this chapter will be evaluated against quality indicators of research design proposed by Gersten et al. (2005), Thompson, Diamond, McWilliam, Snyder, and Snyder (2005), and Brantlinger, Jimenez, Klinger, Pugach, and Richardson (2005) in a special edition of *Exceptional Children* (2005). (See Appendices A, B, and C.)

Gersten et al. (2005) offered an “organizer of critical issues” (p. 149) to be addressed when evaluating studies using experimental or quasi-experimental designs in special education. These critical issues include: conceptualization underlying the study, participants/sampling, implementation of the intervention and description of comparison conditions, outcome measures, and data analysis. Gersten et al. supplied specific questions to determine if these quality indicators have been met, as provided in Appendix A.

Thompson et al. (2005) proposed quality indicators of correlational research designs, addressing the following areas: measurement, practical and clinical significance, avoiding common macro-analytic mistakes, confidence intervals for reliability coefficients, statistics, and effect sizes. They posited multiple indicators for each area, as reported in Appendix B. The authors prefaced that these indicators are "insufficiently honored" and infrequently found in current practice (p.184).

Credibility measures and quality indicators for qualitative studies were presented by Brantlinger et al. (2005). The credibility measures include triangulation, disconfirming evidence, researcher reflexivity, member checks, collaborative work, external auditors, peer debriefing, audit trail, prolonged field engagement, thick, detailed description, and particularizability (i.e., degree of transferability). The quality indicators are provided separately for interview studies, observation studies, document analysis, and data analysis. The authors carefully defined each of these variables, as presented in Appendix C. As I review DBDM, preservice teacher preparation, and inservice teacher professional development (i.e., PLCs) studies in the remainder of this chapter, I will critique them against these proposed standards.

Data-based Decision Making

Data-based decision making (DBDM) includes analyzing and using student data to inform planning, resource allocation, placement, and instruction (USDOE, 2007). Studies reviewed in this section include USDOE survey and interview results, those in which student achievement was impacted by data-based decision making, and those examining how teachers perceive their students' achievement levels compared to what data shows.

USDOE survey and interview results. The U.S. Department of Education Office of Planning, Evaluation and Policy Development (USDOE) found common themes about inservice teachers and the use of DBDM in survey and interview studies conducted between 2007 and 2011. In 2007 they conducted a national survey with 6,017 K-12 general education teachers and 1,039 district technology professionals to determine the prevalence of their access to and use of data management systems (USDOE, 2007). They found that about half of all teachers in the sample had access to an electronic data system that provided them with student data. They also reported that these teachers did not necessarily have the information or tools necessary to make use of the student data available to them and that the most commonly reported use of the data was to inform parents about their child's progress. It was less common for teachers to use the data systems to plan and individualize instruction. They also found that it was common to use the system in collaboration with school colleagues; educators worked on district activities with mentors to analyze student data and response to instruction. This study met many of the quality indicators as proposed by Brantlinger et al. (2005). A strength of the study was the authors' selection of appropriate participants, who were representative of the population. A limitation of this study was that the authors did not provide the survey questions, making it impossible for the reader to determine if the questions were reasonable and limiting replication of the study by other researchers.

In 2008 the same researchers (USDOE, 2008) sampled 1,779 teachers and found that there was an increase in the use of data systems but that this sample of teachers was more likely to have access to grades and attendance data than achievement data. The main reason teachers reported accessing the system was to provide information to parents. Less than 20% of the teachers reported that they had the tools to make instructional decisions informed by the data. The most common support for learning to use student data systems was through their district (e.g., professional development) and not through formal coursework. This study contained the same strengths and limitations as the USDOE 2007 study. Namely, the participants were represented fairly and selected appropriately, but the survey questions were not provided.

In 2009 the USDOE reported data were being used for school improvement but not specifically for informing instruction. More specifically, in survey and case study research,147 teachers from 27 different schools were asked to respond to scenarios involving multiple components of data-informed decision-making. The authors concluded that teachers are “generally capable” of finding information in data representations but that when they have to do computations their ability decreases (p. 42). The authors also reported that teachers have limited understanding of how to interpret data, including how to use individual scores to make predictions and instructional prescriptions. The main barriers reported by the educators were lack of training on how to use the system to derive instructional implications, lack of time, and weakness of available data. The authors met many of the quality indicators as proposed by Brantlinger et al. (2005). A strength of the study was the purposeful sampling used to determine participants. More specifically, the authors found a representative sample of teachers from districts where educators had experience using data. Another strength was the transparency of the interview questions and scenarios. Limitations of the study were the lack of disconfirming evidence and external auditors.

In 2010 the USDOE surveyed 427 districts and made site visits to 36 schools in 12 districts. They found that districts had multiple data systems and that having these systems was driven by accountability requirements. Districts also reported that they were incorporating data into school improvement planning and that strategies were in place for building capacity. Ways to build capacity included offering professional development, hiring personnel in support positions, and developing tools for generating and acting on data. The greatest need reported by districts was for models on how to connect the data to instructional practice. The authors met many of the quality indicators as proposed by Brantlinger et al. (2005). Strengths of the study were that the authors triangulated their data (e.g., they did national surveys and site visits) and used multiple researchers to meet the criteria of collaborative work. One limitation is that the authors did not provide a comprehensive list of the survey questions; however, they provided the majority of the questions throughout the text.

In 2011 the USDOE evaluated how small groups of teachers, administrators, and specialists (e.g., data coaches) were able to master areas of DBDM including data location, comprehension, interpretation, instructional decision making, and question posing. Seventy small groups at 35 schools, including 180 teachers and 35 administrators or specialists, were interviewed. The researchers found that their participants had little difficulty in locating data; however, some struggled when they had to perform calculations. The authors found that teachers and small groups demonstrated "reasonable skill" in comparing data but difficulties with evaluating statements about data that required calculation and differences in data (p. 30). In the area of data interpretation most educators demonstrated understanding of measurement error and the majority understood the importance of sample size for generalizability. The majority of small groups demonstrated that they understood the importance of examining subscale scores and conducting item analyses but many had difficulty with formulating instructional plans for hypothetical students with inconsistent results on a variety of measures. These educators often based their decisions on experience with real students rather than on the data. Finally, the interviewees were able to demonstrate the need to look at multiple measures to inform decisions. Over one-third of the educators, however, asked for data that were irrelevant to the specific goal in the scenario. The authors met many of the quality indicators of interview studies as proposed by Brantlinger et al. (2005). A strength of this study was the authors' thick, detailed description. More specifically, they reported sufficient quotes. Furthermore, in an appendix they provided all of the interview questions and correct answers to the data scenarios. The participants were represented sensitively and fairly in the report.

The USDOE studies allow readers to gain insight into teachers' beliefs and uses of data-based decision making along with their ability to make instructional decisions based on hypothetical data. Overall, results from these studies suggest many districts have data systems in place but that some educators do not have the tools or ability to interpret the data. Furthermore, some teachers continue to struggle to make connections between the data and appropriate instructional decisions. To close achievement gaps, however, we need to know if making data-based decisions impact student achievement.

DBDM and student achievement. Knowing what it means to make data-based decisions and how to make those decisions is important because using data to make instructional decisions leads to improved student achievement (Fuchs, L.S., Fuchs, Hamlett, Phillips, & Bentz, 1994; Gischlar, Hojnoski, & Missall, 2009). Research on increases in student achievement as a result of teachers using assessment data to inform instruction is more prevalent in special education literature (e.g., Fuchs, Deno, & Mirkin, 1984; Fuchs, Fuchs, & Hamlett, 1989; Fuchs, Fuchs, Hamlett, & Stecker, 1991) than general education literature (e.g., Fuchs, L.S., et al., 1994). Many of these studies measure how teachers use Curriculum Based Measurement (CBM) to inform their instruction and increase student achievement. CBM is an assessment tool to monitor student growth in academic areas and evaluate the effects of instruction (Deno, 1985). The validity and technical adequacy of CBM has been reviewed and shown to be effective in predicting student achievement (Wayman, Wallace, Wiley, Tichá, & Espin, 2007). Seminal studies examining the relation between DBDM and student achievement are reviewed below.

Fuchs et al. (1989) randomly assigned 29 special education teachers and 53 students with learning and/or emotional disabilities to CBM or control groups. For 15 weeks teachers in the CBM group used the measurement tool to track student progress in reading. The process included goal selection, measurement, and evaluation. The authors reported that the teachers in the control group set goals for their students using typical Individual Education Program (IEP) methods, such as unsystematic observation and worksheets. After 15 weeks the authors divided the CBM group into two subgroups: CBM measurement only and CBM measurement and evaluation. The CBM measurement only group administered, scored, and graphed CBM data whereas the CBM measurement and evaluation group did these same procedures and introduced at least one instructional modification in response to the data. Students in the CBM measurement with evaluation group performed significantly better than did students in the CBM measurement only group on subtests of the *Stanford Achievement Test,* suggesting the importance of using CBM to make instructional changes when needed to impact student achievement. Researchers met all of the quality indicators for experimental research as outlined by Gersten et al. (2005). Strengths of the study included the authors' use of multiple measures, a clear description of interventions in both conditions, and sufficient information about the participants. One limitation of the study is that the authors did not capture the interventions' effects beyond an immediate posttest.

Fuchs, Deno, and Mirkin (1984) randomly assigned 39 special education teachers to a CBM or control group. For 18 weeks teacher trainers taught the CBM group data-based program modification strategies. They taught the control group "conventional special education evaluation" strategies for diagnosing and treating learning and social problems, how to manage and instruct their instructional groups, and how to work with technology and paraprofessionals (p. 449). Students whose teachers were in the CBM group achieved statistically significant better results as measured by a *Passage Reading Test* and subtests of the *Stanford Diagnostic Reading Test*. Strengths of this study included the authors detailed description of participants, multiple outcomes measures, and appropriate data analysis techniques (e.g., effect size calculations provided). The authors did not, however, provide a description of assessments results for fidelity of implementation.

Stecker and Fuchs (2000) explored how 22 special educators made data-based decisions to examine the achievement of 84 students assigned to CBM (n = 42) and non-CBM (n = 42) groups. Teachers were asked to match pairs of students with like mathematics abilities for the CBM or non-CBM group using objective or subjective methods. For students in the CBM group, the teachers monitored math progress using CBMs and made instructional adjustments. For students in the non-CBM group, the teachers made instructional adjustments without using CBM data. The authors used the *Mathematics Operations Test-Revised* to measure achievement. They found that students who were in the group in which the teacher made instructional decisions based on CBM data performed significantly better than did their peers for whom teachers made instructional adjustments without data. Strengths of this study included the authors' sufficient description of the participants, intervention, and comparison conditions. The authors also used multiple measures and linked data analysis to the research questions. One limitation of this study is noted. The authors had the teachers "use objective or subjective" means to match students for the conditions (p. 129) rather than using random assignment. The authors used statistical analysis to determine that the two groups were equivalent, but this was not done a priori and limits the conclusions that can be drawn.

While the above studies examined the use of DBDM with special education teachers, in this review only one study was identified that examined general educators' use of DBDM. Stecker, Fuchs, and Fuchs' (2005) found the same results in their review of the literature on the use of CBM to improve student achievement.

Fuchs et al. (1994) evaluated the effectiveness of CBM decision-making strategies in the general education setting. They randomly assigned 40 general education teachers to three groups: CBM with classwide reports and instructional recommendations, CBM with classwide reports but no instructional recommendations, and contrast (no CBM). They measured academic achievement using the *Math Operations Test-Revised* and CBM slope data. They found that CBM strategies designed with classwide reports and instructional recommendations resulted in improved student achievement compared to the contrast conditions. Fuchs et al. (1994) met all of the essential quality indicators for an experimental study (e.g., provided sufficient information about participants, described the intervention clearly, described and assessed fidelity of implementation, used multiple measures). Furthermore, they met 75% of the desirable indicators for experimental studies (e.g., provided interrater reliability data, assessed fidelity and quality of implementation, presented the results clearly). The authors did not provide data on attrition rates, capture the intervention's effect beyond a posttest, or include audio or videotaped excerpts.

In a study that showed different effects than Fuchs et al. (1994), Graney and Shinn (2005) worked with 44 second-grade general education teachers and their 184 students who were in low reading groups. The authors randomly assigned teachers and students to one of three groups: CBM with group feedback, CBM with individual student feedback, and control. After five weeks of reading instruction, teachers in the CBM with group feedback condition were shown a graph that included all of their students' CBM growth. Teachers in the individual feedback group were shown just one of their students' CBM graphs. Teachers in the control group did not receive any progress monitoring feedback. The researchers found that reviewing progress monitoring data did not have an effect on students' reading progress. Interestingly, they did state, however, that all three groups' reading progress improved reliably, "suggesting some reactive benefits of progress monitoring" (p. 184). The authors of this study, unlike Fuchs et al. (1994), did not provide instructional implications for the teachers after they reviewed the progress monitoring data. The study authors met the majority of the essential and desirable quality indicators for experimental and quasi-experimental research. They did not, however, use multiple measures or measure student achievement beyond a posttest.

Overall, results of these studies suggest that special education teachers make data-based decisions (e.g., review progress monitoring data) that impact student achievement. The study by Fuchs et al. (1994), that included general education teachers, showed similar results. The study by Graney and Shinn (2005), however, did not show that reviewing student progress monitoring had a significant effect on academic achievement, however, all students improved their reading. The final section summarizes literature on teachers' perceptions of their students compared to how they actually perform on direct measures of academic achievement, thus arguing the importance of looking at student data to make instructional decisions.

Teacher perceptions compared to student data. Teachers' judgment of students' achievement levels is often used to make instructional decisions but these perceptions do not always match how students perform, particularly on more direct measures of academic achievement (Begeny, Eckert, Montarello, & Storie, 2008; Feinberg & Shapiro, 2003). In a review of literature on teacher-based judgments of academic achievement, Hoge and Coladarci (1989) examined 16 studies in which mostly general educators judged their students academically and then compared that to actual student achievement on standardized tests. These authors found that there were "generally high levels of agreement between the judgmental measures and the standardized achievement test scores" (p. 308). Begeny et al. (2008) and Hamilton and Shinn (2003) argued, however, that teacher judgments need to correlate with student progress on measures that are more direct, such as reading fluency or comprehension rather than "global judgments of reading" (p. 230). Direct measures, such as CBM, are linked to the content of instruction more so than are norm-referenced and standardized tests (Feinberg & Shapiro, 2003). The remainder of this section will summarize literature in which teachers make judgments on student achievement and its relation with academic achievement on direct measures.

Hamilton and Shinn (2003) explored teachers' judgment of reading comprehension levels of "word callers" (i.e., students who read fluently, yet do not comprehend what they read; p. 229). The authors recruited 29 third grade teachers who had identified 35 "word callers" and 33 similarly fluent peers. They compared the teachers' judgment of both groups of students on three measures: CBM in reading, CBM with maze passages, and a comprehension oral question answering test. Results disconfirmed teachers' perceptions. The identified "word callers" did have lower comprehension levels, but they also had lower CBM fluency scores than predicted by their teachers and compared to their "similarly" fluent peers. Furthermore, the teachers significantly overestimated the CBM reading, maze, and comprehension levels of all students. The authors met many of the quality indicators as proposed by Thompson et al. (2005). Strengths of this study include the use of multiple measures and a clear description of the participants. A noted limitation is external validity; the study should be replicated with other grade level students from different geographic areas.

Feinberg and Shapiro (2003) examined the accuracy of teacher predictions of oral reading fluency. Thirty teachers and 30 students in third, fourth, and fifth grades participated. The authors found a moderate correlation between teacher judgment and actual reading performance. However, when effect sizes were used to compare teacher judgment on the actual performance they found a substantial difference between the judgment and scores at the third grade level and a large difference at the fourth grade level. The authors argued that teachers may be less accurate at predicting specific performance levels although they can predict relative differences in students. The authors met many of the quality indicators as proposed by Thompson et al. (2005). Strengths of this study include reliability and validity of the measurement tools, providing effect size statistics, and avoiding macro-analytic mistakes. A limitation, however, is that they did not include confidence intervals. Additionally, they did not explain if the assumptions of statistical methods were met.

Eckert, Dunn, Codding, Begeny, and Kleinmann (2006) sought to extend the results of Feinberg and Shapiro (2003) by examining the correlation between two general education teachers’ perceptions of 33 second grade students' achievement in reading and math and students' actual scores on CBMs in these content areas. The teachers were interviewed and completed assessment charts to judge the levels of their children in math and reading. Similar to Feinberg and Shapiro (2003) and Hamilton and Shinn (2003), these researchers found that correlations between teachers' judgments and actual student scores were low for math and moderate to high for reading. More specifically, the teachers overestimated their students' performance across math skills (e.g., addition and subtraction) and on reading material that was below or at grade level. Strengths of this study were similar to those of Feinberg and Shapiro (2003); they met many of the quality indicators. A noted limitation of this study was the small sample size of two teachers making judgments on 33 students. The results should be viewed with caution as they may not be generalizable.

Begeny et al. (2008) examined teachers' judgments across a continuum of assessment methods including oral reading fluency, teacher rating scales, interview data, and class ranking charts. The authors had 10 teacher participants and 87 students in grades one, two, and three. Similar to the previous results, these authors found that teachers' judgments were moderately correlated with students' oral reading fluency scores. The teachers were able to rank order their students' reading ability compared to peers. Again, the teachers often overestimated their students' reading abilities. The educators also had difficulty identifying the students who read material at the instructional or frustration level (compared to the independent level). These findings suggest that teachers may have difficulty identifying instructional levels for their students. The authors met quality indicators in measurement, practical and clinical significance, and avoided making common macro-analytic mistakes. They did not provide confidence intervals. Overall, the results from the literature on teacher perception showcase their ability to make fairly accurate estimations of their students' ability on standardized tests but that they often overestimate students' skills on more direct measures that are a closer link to classroom instruction.

Hamilton and colleagues, in an IES Practice Guide (2009), stated that the "existing research on using data to make instructional decisions does not yet provide conclusive evidence of what works to improve student achievement" (p. 6). They argued that this type of research is difficult to carry out, is linked to education technology, and entails multiple elements (e.g., instruction in addition to data-based decisions). This review of the literature suggests the following: (a) teachers lack sufficient tools and ability to make instructional decisions, (b) special education teachers who make data-based decisions positively impact the achievement of special education students, but research on general education teachers does not warrant the same conclusion because there have not been enough studies to document this effect, and (c) teachers' perceptions of their students' academic abilities do not clearly match their achievement on direct assessments. The remainder of this chapter will focus on literature on how to effectively work with preservice and inservice teachers so that they can make data-based decisions to improve student outcomes.

Preservice Teacher Preparation

Teacher preparation has been an under-studied area in teacher education (Wilson, Floden, & Ferrini-Mundy, 2002). Wilson et al. were asked by the U.S. Department of Education to review research on teacher preparation and found no "high quality" research on the effects of subject matter preparation, pedagogical preparation, and student learning. They also reported that the effects of student teaching consisted of small interpretive studies only. Additionally, Maheady, Jabot, Rey, and Michielli-Pendl (2007) suggested that more studies on how preservice teachers' skills develop and how they impact student learning are necessary. Additionally, for RtI to be successful, it is imperative that professionals have appropriate training in the components of DBDM (Hawkins, Kroeger, Musti-Rao, Barnett, & Ward, 2008). In a literature review by Hawkins et al. (2008), however, they found no studies related to RtI and preservice training.

Data-based decision making. The literature on DBDM and preservice teachers is not voluminous; only three qualitative studies from peer-reviewed journals were found that provided information on this topic. In a study by Hawkins et al. (2008) university faculty members structured a field experience for preservice special education teachers and school psychologists to collaborate in a school with kindergarten general education teachers. The researchers created a guiding checklist for the trainees, supervised the students, wrote case notes, had students complete a portfolio and teacher work sample, and measured kindergarten student outcomes. The trainees conducted universal screening, implemented tiered instruction, and made data-based decisions (e.g., visual analysis of data from graphs, team meetings). Results from the guided checklist to the case notes were descriptive in nature. Results measured from kindergarten student progress showed that 91% of students were proficient in letter sound fluency by the end of the school year compared to 52% at the beginning of the study. Hawkins et al. concluded that the core building blocks of RtI need to be taught to “professionals in training,” including the development of field experiences for these trainees, training in

DBDM, collaboration on the results of assessment data, and effective instruction (p. 757). The authors met many of the quality indicators and credibility measures proposed by Brantlinger et al. (2005). One strength of this study was triangulation; the authors used multiple and varied data sources (e.g., work samples, student data). A second strength was the collaborative work by the researchers, graduate students, and preservice teachers. A limitation was the lack of sufficient quotations or field notes to support the conclusions made.

Buck, Trauth-Nare, and Kaftan (2010) conducted action research to study how 30 preservice science teachers understood formative assessment. The course provided four days of explicit instruction on assessment and used a field-based case study. The researchers collected data on pre- and post-course questionnaires, meeting transcripts, course documents, focus group interviews, and field notes. Based on those data sources, the authors concluded that the preservice teachers “demonstrated a substantial increase in understanding of formative assessment” and could provide the purposes of formative assessment and the link between assessment and instructional planning (p. 417). The authors met many of the quality indicators and credibility measures proposed by Brantlinger et al. (2005). Strengths of this study included prolonged field engagement, use of thick detailed description, triangulation, and connections made with related research. A limitation was that the authors did not have an external auditor confirm if the inferences made were grounded in findings.

Graham (2005) worked with 38 preservice secondary English teachers and used both observation and interviews to examine preservice teachers’ knowledge on DBDM.

The author stated that before students took courses and worked with mentor teachers they were “clueless” about establishing goals for students and how to assess them (p. 612). After being enrolled in courses and working with mentor teachers for two years, Graham indicated that preservice teachers gained a clearer definition of what assessment is, understood that formative assessment guided instruction, and benefited from working with mentor teachers. She also reported that there were unresolved concerns after the two year experience. More specifically, preservice teachers were concerned with designing goals, grading and fairness, grading and motivation, validity of assessments, and the time required to make data-based decisions. Graham concluded that preservice teachers are “doomed” to replicate more traditional, unexamined assessment practices, if they are not taught data-based decision making or don’t observe experienced teachers (p. 619). The authors met many of the quality indicators for observation and interview studies outlined by Brantlinger et al. (2005). Strengths of this study included member checks to confirm the accuracy of notes and thick, detailed description including sufficient quotes to provide evidence of the researchers’ interpretations. A limitation was that the author stated that observation was used as a mechanism to collect data but then did not explain what their observation entailed.

In summary, the main avenues through which preservice teachers received instruction on DBDM were collaboration with mentor teachers (Graham, 2005; Hawkins et al., 2005), explicit instruction (Buck et al., 2010; Graham, 2005), and use of case studies (Buck et al., 2010). The three available articles were all qualitative in nature. Brantlinger et al. (2005) argued that reviewers of qualitative research should not “make authoritative pronouncements about what works" but should instead use evidence gathered to inform policy and practice (p. 202). These studies, viewed with caution, provide some evidence to show that working with preservice teachers to make data-based decisions may be impactful.

Preservice teacher preparation programs. In addition to knowing the literature on working with preservice teachers to make data-based decisions, it is important to know the literature, in general, on how to teach future educators, as they are the participants in this study. Brownell, Ross, Colon, and McCallum (2005) conducted comparative research and documented the characteristics of effective general education teacher programs based on previous literature in these areas (e.g., National Commission on Excellence in Elementary Teacher Preparation for Reading Instruction [NCEETPRI], 2003). They found seven common features of effective programs: (a) coherent program vision, (b) conscious blending of theory, disciplinary knowledge, subject-specific pedagogical knowledge and practice, (c) carefully crafted field experiences, (d) standards for quality teaching, (e) active pedagogy, (f) focus on meeting the needs of a diverse student population, and (g) collaboration as a vehicle for building professional community.

One of the reviews from the Brownell et al. (2005) article was the National Commission on Excellence in Elementary Teacher Preparation for Reading Instruction (NCEETPRI, 2003) study. The International Reading Association asked the Commission to study teacher preparation in reading and provide leadership for change. The researchers identified eight institutions with excellent credentials. Using a quasi-experimental design, they followed 101 beginning teachers and compared teaching practices and student achievement results from those who had attended institutions with influential reading preparation programs to those who did not. They found eight critical features of institutions with excellent reading preparation programs: content, apprenticeship, vision, resources and mission, personalized teaching, autonomy, community, and assessment. More specifically, apprenticeships allowed preservice teachers the opportunity to interact with excellent models and mentors. Community included active learning with faculty, students, and mentor teachers. The authors shared their findings in an executive summary so the specific details regarding the method and data analysis were not provided, thus making it hard to measure how they met the quality indicators. An overall strength of the study, however, was the amount of detail the authors provided in explaining the features of the reading programs.

In a comprehensive review, Wideen, Mayer-Smith, and Moon (1998) critically analyzed the research on learning to teach. The authors found that long term interventions resulted in positive effects on teacher candidates' beliefs. The long term interventions included a small number of participants, group work, and a close relationship with instructors. For example, in a three-year longitudinal study Gunstone, Slattery, Baird, and Northfield (1993) provided a one-year teaching education program, that included active learning, with 13 preservice teachers and measured their cognitive and academic development. In comparison, when educators provided only knowledge to preservice teachers, rather than active learning, there was minimal to no impact on teacher candidates' beliefs. A concerning limitation of the research in this area, including the studies in Wideen et al.'s review, is that researchers generally did not directly measure the impact of teacher education programs on student achievement.

In summary, more studies are warranted on teacher preparation (Wilson et al., 2002), and specifically, on how teachers develop their skills and the impact these skills and preservice education has on student learning (Maheady et al., 2007). With that caveat, the authors of the previous articles (Brownell et al., 2005; NCEETPRI, 2003; Wideen et al., 1998) provided recommendations for teacher preparation programs. All teams emphasized the use of active teaching methods such as modeling, collaboration, and groups. Active-learning methods include discussion techniques, working in collaborative groups, and reviewing case studies (Nilson, 2003). Nilson argued that active-learning methods are integral for developing problem solving skills and requiring students to assume responsibility for their own learning. Preservice teachers may benefit from engaging in these types of activities. Similar methods are being used with inservice teachers during professional development (PD) and the next section will explain how these teachers receive PD through Professional Learning Communities (PLCs).

Inservice Teacher Professional Development - Professional Learning Communities (PLCs)

Given that the main opportunity for learning about DBDM is through inservice training (USDOE, 2008), it is appropriate to review literature available on the PD that these teachers receive. No literature specifically on PD and DBDM was found so PD for inservice teachers, in general, was reviewed. The use of Professional Learning Communities (PLCs) is now apparent in many districts as the main way to provide PD

(Vescio et al., 2008). The premise of PLCs lies in the core belief that improving teaching practices will lead to improved student learning (Vescio et al.). Vescio et al. argued that researchers have only recently begun to empirically examine how teaching practices change in response to student learning because of PLCs.

In a review of 11 empirical studies on the impact of PLCs on teaching practice and student learning, the authors explained that collaborative strategies include teachers who encourage, reflect, take risks, share lessons, use protocols for decision making, rely on note taking to inform colleagues about work, observe each other teaching, investigate teaching problems, collectively generate new ideas, and engage in literature circles. Similarly, the authors documented in their review that most studies included teachers who focused on student learning. For example, in one study the teacher teams focused on improving learning for low and under achieving students while one focused on improving student literacy. Another important element in this review was teacher authority; it is critical that educators are able to make decisions about their PLCs and overall school governance. The researchers emphasized the importance of continuous teacher learning and highlighted the use of educators reading scholarly articles. Vescio et al. also studied the impact PLCs had on student achievement. According to the authors there is “modest evidence” that PLCs impact teaching but a “resounding and encouraging yes” that student learning increases when teachers participate in these PLCs (p. 87). The remainder of this section will look more closely at the eight studies that attempted to make the connection between PLCs and student achievement. Three of the studies were not included because they did not link PLCs with student achievement. An additional two articles published after Vescio et al., and reflective of the impact of PLCs, will also be reviewed.

Supovitz and Christman (2003) wrote a policy brief that Vescio et al. (2008) included in their review of the literature on PLCs. They used a qualitative approach to review instructional practices from the Cincinnati and Philadelphia public schools. Reform initiatives in both districts used small learning communities to impact instruction. The authors summarized that the reform movement (a) had significant and positive influences on the school environment, (b) influenced instructional practice, and (c) influenced student performance (i.e., PLCs that engaged in “structured, sustained, and supported instructional discussion and that investigated the relationships between instructional practices and student work produced significant gains in student learning”) (p. 5). The authors did an adequate job explaining the reform movements but the results on student achievement cannot be linked explicitly to PLCs.

Berry, Johnson, and Montgomery (2005) implemented qualitative research studying the impact of teacher leadership, specifically with PLCs, on transforming student achievement. The authors conducted observations of 25 teachers (nine of whom were nationally board certified) at a rural school with 560 students in grades 3-5. With the help of a consultant, the teachers were part of PLCs where they created collaborative teams, used reflective skills, discussed instructional strategies around assessment, and were involved in data analysis and decision making. Within four years the school had gone from one in which only about half of the students were performing at or above grade level to one with more than 80% meeting grade level standards. Only a small number of the quality indicators proposed by Brantlinger et al. (2005) were met. A strength of this study is the collaborative work involving multiple researchers. This research should be viewed with caution, however, because there is no way to tease out what impacted student learning (e.g., having nationally board certified teachers or PLCs). Additionally, the authors did not triangulate their data or provide enough thick, detailed description to provide evidence of their conclusions.

Phillips (2003) used qualitative data to describe how one middle school, with over 600 students, implemented a learning community that impacted student learning. The author used interviews, classroom observations, focus groups, and select documents (e.g., student work). Five key themes emerged from the study; the school personnel had high quality professional development, used research-based literature, shared leadership, used collaborative processes, and considered the school context (e.g., understanding of their student population). Additionally, the school showed "dramatic academic growth, especially in reading and mathematics achievement” according to the Texas Education Agency (p. 256). Phillips met few of the quality indicators proposed by Brantlinger et al. (2005) for qualitative studies. She did provide numerous vignettes that included sufficient quotes, but did not provide enough information about researcher reflexivity, collaborative work, or provide specific examples of the interview questions.

Strahan (2003) used qualitative methods to explain how three schools have ‘beaten the odds’ (p. 127). The author constructed case studies by collecting demographic and achievement data, conducting interviews, and doing observations at three elementary schools servicing about 1,400 children, over 70% of whom received free or reduced price lunch. Strahan documented that student scores on achievement tests rose from less than 50% proficient to more than 75% proficient. He concluded that the educators at these schools used ‘data-directed dialogue’ to inform their instruction (p. 143). More specifically, the educators met in grade level teams to identify student needs, develop strategies to improve student learning, and enact more effective strategies. The author met the majority of the quality indicators from Brantlinger et al. (2005). He triangulated his data, provided explicit examples to support his claims, and represented the participants fairly.

Hollins McIntyre, DeBose, Hollins, and Towner (2004) also used qualitative approaches to gather and analyze data on how 12 teachers (10 African American, 2 European American) promoted a self-sustaining PLC while impacting student outcomes of predominately under-achieving African American children in kindergarten through fourth grade. The authors conducted interviews, transcribed meetings, recorded field notes, and used quantitative data to measure the impact of the PLC. The teachers met biweekly in collaborative groups where they described challenges, identified strategies to meet the challenges, implemented the strategies, evaluated the implementation, and determined future practices. The authors suggested that their qualitative findings showed that conversation amongst teachers became more positive, teachers made links between themselves and the children, and they shared strategies and developed new approaches collaboratively. The quantitative data suggested that second graders from the target school made an overall 28% gain on the *Stanford Achievement Tests* compared to the district overall gain of 12%. Similarly, third graders made a 31% overall gain compared to 9% at the district level. The authors met many of the quality indicators proposed by Brantlinger et al. (2005). The researchers were reflexive (i.e., self-disclosed their assumptions and biases) in their paper, which allowed the readers to understand their perspective. Furthermore, they triangulated their data and provided thick descriptions of the PLCs. One limitation was that the authors did not provide their interview questions.

A more robust study was conducted by Bolam, McMahon, Stoll, Thomas, and Wallace (2005) who used a mixed methods approach (qualitative and correlational) to answer research questions about the components of PLCs and the impact they have on student achievement. The authors surveyed over 390 schools and followed up with 16 case studies. They used the responses from the PLC questionnaire and correlated them with student achievement on national assessments. They also calculated value-added scores. The authors then examined PLCs using a case study approach that included interviews, document analysis, and observations at schools that varied on questionnaire responses (i.e., studied three low PLC implementation schools, three medium implementation schools, and three high implementation schools). Bolam et al. concluded that there was a positive link between PLCs and student outcomes, especially when value-added performance was included. "It appears that the greater the extent of reported staff involvement in professional and pupil learning, the higher was the level of pupil performance and progress in both primary and secondary schools" (Bolam et al., p. 132). The authors met the majority of the credibility measures and quality indicators of qualitative research proposed by Brantlinger et al. (2005). They triangulated their data and gave very detailed descriptions of their results. However, they only met a handful of the quality indicators for correlational research proposed by Thompson et al. (2005). A strength was in how they reported the reliability and validity of the measurement tools. Limitations were the common macro-analytic mistakes they made (e.g., they did not prove they met the assumptions of statistical methods) and they did not report confidence intervals.

Louis and Marks (1998) also used mixed methods to study the impact of PLCs on authentic academic achievement. The authors defined authentic achievement as one in which children construct their own knowledge, exhibit disciplined inquiry, and elaborate written communication. Louis and Marks used a correlational design to determine if PLCs, impacted authentic achievement, while controlling for other variables such as race and academic achievement scores. The authors analyzed 910 teacher surveys, conducted 25 interviews and observations, scrutinized 144 instructional practices, and analyzed over 5,000 pieces of student work. They found that in schools with strong professional communities the students performed significantly higher on authentic achievement measures, after controlling for student variables. To further support these findings the authors offered qualitative insights based on case studies. They documented that the strong PLCs had common themes: shared values and norms, focus on student learning, collaboration, deprivatized practice, and reflective dialogue. The authors met many of the quality indicators proposed by Thompson et al. (2005) and Brantlinger et al. (2005). A strength of the correlational research was that the authors provided reliability coefficients of the measures, whereas one limitation was that they did not provide validity data of the measures. Strengths of the qualitative part of the research included the authors' use of multiple sources of data and prolonged engagement in the field.

Supovitz (2002) used mixed methods to understand more about PLCs. The author addressed three research questions: (a) did teaming influence the culture of the school?; (b) did teaming change instructional practices?; and (c) did teaming improve student learning? Supovitz used survey data from about 3,000 administrators and teachers, hundreds of interviews, numerous observations, as well as hierarchical linear modeling, to answer the research questions. He concluded that teaming did influence the culture of the school; for example, in schools where there were more teams, the teachers felt more involved in decision making and had higher levels of collaboration. He found that there were no differences between instructional practices for team and non-team based schools. Finally, he said there were "no discernible patterns of higher student performance in team-based schools when they were compared to nonteam-based schools" (p. 1612). He did state, however, that when the teams focused on group instructional practices the students had statistically significantly higher scores on achievement measures. Supovitz met about half of the correlational quality indicators. One strength was the interpretation of effect sizes. One limitation was that he did not provide confidence intervals. He also met many of the credibility measures of qualitative studies. He triangulated his data and provided the survey questions, but did not provide information on member checks.

Two remaining articles published after Vescio et al.'s (2008) review examined how PLCs impacted student achievement. Both research teams employed qualitative research. Kanold, Toncheff, and Douglas (2008) explored how two high schools focused on the use of collaborative teams to teach, plan, and assess. The districts hired consultants to help their educators deprivatize practice (i.e., remove teacher isolation and encourage meaningful teacher collaborative work time), a main tenet of PLCs. The consultants also worked with site leaders to build collaborative teams. Finally, the educators were required to create standards, goals, and assessments to understand what students were learning. The authors stated "district achievement rates reached unprecedented levels" (p. 27). Kanold et al. met some of the quality indicators supplied by Brantlinger et al. (2005). They triangulated their data and had prolonged field engagement. They were not, however, explicit about their methodology.

Maxwell, Huggins, and Scheurich (2010) examined how an underperforming diverse rural high school teaching staff helped facilitate a "turnaround" in state accountability scores (e.g., 11% of African Americans were passing in math in 2002-2003 compared to 71% in 2007-2008). The authors conducted prolonged interviews with 13 educators and found that reform initiatives, specifically using PLCs, created the turnaround needed in the low achieving school. The authors further explained that the PLC included educators doing a needs assessment, conducting research in the community, implementing recommendations (e.g., changing the schedule), and providing ongoing professional development support. The authors were very explicit in their methodology and met all of the quality indicators of qualitative research.

Overall, 10 articles were reviewed on the impact of PLCs on student achievement; three author teams employed mixed methodologies (i.e., qualitative and correlational research designs) and seven used qualitative approaches. The seven qualitative studies should be viewed with caution as this type of research cannot make causal statements; rather, it informs policy and practice (Brantlinger et al., 2005). The authors who used correlational research designs met only some of the quality indicators, which according to Thompson et al. (2005) is not uncommon in educational research. In summary, PLCs include the following components: (a) collaborative practices, (b) reflective dialogue, (c) focus on student learning based on data, (d) changes in instructional strategies or instruction, and (e) deprivatized practice. There is some evidence also that PLCs may impact student learning and achievement (Bolam et al., 2005; Louis & Marks, 1998).

Summary

In this comprehensive review I have summarized literature on DBDM, preservice teachers, and PLCs. Overall, there are six key findings. First, the USDOE studies suggest that districts have data systems in place but that teachers struggle to make linkages between the data and how to interpret it to make instructional decisions. Second, special education teachers are more adept at making data-based decisions that impact student achievement compared to their general education teacher peers. Third, teachers have accurate perceptions of children's ability on standardized measures of achievement, but their perceptions on more direct assessments are not adequate. Fourth, preservice teachers have received instruction on DBDM via collaboration with mentor teachers, explicit instruction, and the use of case studies. Fifth, there is a need for more studies on teacher preparation; what studies exist highlight excellent teacher preparation program components (e.g., blend of theory, pedagogy, and practice, collaboration to build professional communities. Sixth, PLCs are a common way to work with inservice teachers. When educators work within a PLC framework they may be able to impact student achievement.

Research Questions

This literature review has suggested that teachers need additional work with data to make decisions that inform instruction and impact student learning. Moreover, preservice teachers need explicit instruction on DBDM so that they can be informed practitioners when they enter the field. Finally, the main avenue for working with teachers is through PLCs where they work collaboratively with others to make decisions about student learning.

The purpose of this study is to investigate the effects of a DBDM intervention, using a collaborative PLC-type approach, on the quality of preservice teachers' decision-making. Specifically, the study will address three questions:

1. Does collaborating with a mentor and peers influence the quality of preservice teachers' data-based decision making?
2. Does collaborating with a mentor and peers influence the confidence levels of preservice teachers' in regards to data-based decision making?
3. Does collaborating with a mentor and peers influence the value preservice teachers place on data-based decision making to make instructional decisions?

Chapter 3

Method

Participants and Setting

Study participants were 45 undergraduate students enrolled in a teacher education course (Educational Psychology 5616: Behavior Analysis and Classroom Management) required for general education teacher licensure. The course focuses on principles of behavior analysis and procedures used in the assessment and management of classroom behavior. Students study techniques for collecting data and determining appropriate interventions in addition to reviewing aggregate data. Forty-one of the 45 participants (91%) were female with an average age of 21.5 years old (range 20 to 30). The majority of participants were Caucasian (87%). The remaining participants' ethnicities were Asian (4%), Hispanic/Latino (2%), Asian and Caucasian (2%), and Hispanic/Latino and Caucasian (4%). All study activities were incorporated into the teacher education course and occurred during regularly scheduled class times in the regular classroom setting. Additional classrooms were needed for small groups, but all were held in traditional campus classrooms.

Measures

Data-based decision making pre- and posttest. The pre- and posttest measure was initially developed by the United States Department of Education (USDOE; 2011) as part of their interview study. The interview questions were rewritten in test format so that participants could answer questions in a written format that could then be scored quantitatively (see Appendix D). Participants were given a hypothetical case study and directed to look at student data. They were then asked questions about data location, data comprehension, data interpretation, data use, and question posing. There were five main questions in the pretest, each of which had two to three intended answers according to the USDOE (2011). Participants' answers were only considered to be correct if they supplied one or more of the intended answers (as shown in Appendix E). Additionally, the participants received an extra point if they accurately extended the answer beyond one of the correct answers. For example, one question asked how the teacher would group a particular student or which approach to use with the student. One of the intended answers is to recommend an individualized instructional plan for the child based on the assessment results. If the participant stated that the child should get one to one instruction in reading she would receive one point for the correct intended answer. If the participant went on to write "the child should receive individualized instruction where he is taught high frequency words, decoding strategies, and fluency" she would receive the extra point for extending the answer. Extended answers needed to be accurate and relevant to the original intended answer. In total, participants could receive a maximum score of 26 (i.e., 5 questions with 13 intended answers plus 13 opportunities to extend an answer). A comparable posttest was developed and used (see Appendices F and G).

Social validity and confidence tool. Participants also completed a social validity and confidence tool that measured the level to which they agreed or disagreed (on a 1-5 point Likert-type scale) with a declarative statement about data-based decision making and their confidence levels with making appropriate decisions in the five areas of data-based decision making (e.g., data use, data comprehension) (see Appendix H).

Demographic information. Participants were also asked to complete a demographic survey that included their gender, ethnicity, and age (see Appendix I).

Procedure

The investigator secured informed consent (see Appendix J) for a total of 48 participants. Two participants were excluded because they did not attend class on the day of the pretest. All study activities were incorporated into the teacher education course and occurred during regularly scheduled class times.

Mentors and training. Five doctoral students from the University of Minnesota Educational Psychology department served as the mentors for the intervention. The mentors were in either their third or fourth year of the program and all were Caucasian females ranging in age from 25-40 years old. Four of the five mentors were on a USDOE Leadership Training Grant fellowship helping implement RtI in the Minneapolis Public Schools. The fifth mentor was a School Psychology doctoral student who also had experience implementing RtI in the Minneapolis Public Schools. All five mentors participated in a 90-minute training provided by the investigator. The training included an overview of the research and questions of the study followed by an overview of the activities occurring throughout the study (e.g., pretest on study day 1, intervention on study day 2). The trainees were then given overall directions about the intervention (e.g., audio record the duration of your group work conversation). The majority of the training was then spent going over the semi-scripted USDOE (2011) case study approach with specific scenarios, questions, and follow-up questions. The mentors asked the investigator questions throughout the training to better understand how to implement the intervention. The training ended with any final questions or concerns.

All participants. The main avenue for instructing teachers on DBDM is through the use of collaborative work in Professional Learning Communities (PLCs; USDOE, 2008). To begin the study, however, participants were given a brief overview about Data-based Decision Making (DBDM) and the purposes of assessment. A 15-minute mini-lesson was provided by the investigator prior to participants taking the pretest. During this time, a PowerPoint presentation (see Appendix K) was used to explain important concepts. For example, participants were taught that the main purposes of data collection are screening, progress monitoring, instructional planning, resource allocation, eligibility for special education, program evaluation, and accountability decisions (Salvia, Ysseldyke, & Bolt, 2007). The participants viewed hypothetical student data and were told general information about when and how to use the given data (e.g., screening data can inform a practitioner that some students may need remedial work). They then took a pretest that took approximately 20 minutes. After the 15-minute instruction and pretest all study participants then received a 30-minute lecture on how to make data-based decisions. They viewed screening data and were told the steps to make appropriate data-based decisions (see Appendix L). Finally, they were randomly assigned to experimental or control conditions for day two of the study.

Experimental group. On the second day of the study, the participants were randomly assigned to five collaborative work groups to work for 30 to 45 minutes. A mentor (graduate students previously trained in RtI, as described above) was randomly assigned to each work group and asked to follow a guided script to engage participants in making data-based decisions. Following the group work the investigator led a 30-minute debriefing session where participants were encouraged to share both strengths and challenges of the collaborative work groups and ask questions of the mentors and the investigator. On the third day of the study, students in the experimental condition took the posttest and then received regular instruction (i.e., course content that had been presented to the control group on day two of the study) from one of the two instructors of the course.

Control group. On the second day of the study, participants in the control condition took the posttest and then received regular instruction from one of the two instructors of the course. On the third day of the study, students in the control group, who had already taken the posttest, received the intervention as explained above in "Experimental group."

Intervention condition. The main component of the intervention was a mock PLC, in which groups worked collaboratively to examine data. Five work groups were led by the doctoral student mentors with experience leading data meetings and making instructional decisions based on data. In their collaborative work groups, the mentors led a discussion to facilitate the preservice teachers to make accurate data-based decisions. An adapted case study approach that provides hypothetical screening data from the USDOE (2011) (see Appendix M) was examined by the work groups. The mentors engaged in semi-scripted dialogue so that all participants received the same general instruction. For example, the first scenario asked participants to determine the number of third grade students who scored above average on the given assessment. The mentor asked the participants the question and then waited for a response. If the participants were not able to answer the question accurately the mentor used a structured response and questions to elicit the accurate answer. The guiding questions as well as the expected answers were given to all mentors along with a time guide to make sure that they were able to complete all scenarios and discussion points in the allotted time. Overall, the mentors worked with the participants on six hypothetical scenarios and questions. Towards the end of the intervention the investigator saw that groups were finishing earlier than was expected. To address this situation mentors were asked to work with their groups to sort all students into reading groups. They were specifically asked to complete the prompt: “Group all students into a group – write out who would be in each group and what they would focus on; the instruction include…” On the prior six scenarios the correct responses had been scripted, but for this question the participants were encouraged to work in pairs or small groups within their group to make instructional reading groups based on the provided data. The participants then shared out their group information and engaged in conversation regarding the decisions they had made. A participation sheet was provided to each mentor so she could document student participation (see Appendix N).

The final part of the intervention was a 30-minute debriefing session in which participants were asked strengths and challenges of the collaborative work. They were also encouraged to ask the panel of mentors questions about making data-based decisions. The investigator then provided concluding comments on DBDM and the study (see Appendix O).

Control condition. Participants in the control condition took the pre-and posttest as previously explained. On day two, when the experimental group was participating in the intervention, they received instruction from one of the two course instructors on regular course content. The participants were then provided with the same intervention as described above on day three, after they had taken the posttest.

Data Analysis

A Mann Whitney U test was used to examine if there was a pre-existing difference on the distributions of the pretest scores between the experimental and control conditions. To answer the research question, the Mann Whitney U test was again used to analyze the data. This nonparametric test allowed me to see if there are differences on the distributions of the gain scores (posttest - pretest) between the experimental and control conditions on the dependent variable. In a similar way, the Mann Whitney U test was used to examine if there was a pre-existing difference in the view on validity and confidence levels of the preservice teachers in each condition. The Mann Whitney U test was also used to see if there was a difference in the gain scores between the experimental and control conditions on their views on the validity of DBDM and their confidence with making data-based decisions.

Treatment Fidelity and Interrater Agreement

Procedural fidelity of implementation of the 30-minute whole group instruction on data-based decision making was assessed (see Appendix P). Procedural fidelity of implementation of the intervention for all five work groups from the experimental condition was also assessed (see Appendix Q). All five groups audio recorded their group discussion and the investigator listened to each recording and assessed the fidelity of implementation to the protocol.

To check the inter-rater reliability of scoring the tests, 25% of the assessments were randomly selected and scored by the investigator and a second rater who was blind to the condition that each participant was assigned to.

Chapter 4

Results

The purpose of this study was to investigate the effects of a Data-Based Decision Making (DBDM) intervention, using a collaborative PLC-type approach, on the quality of preservice teachers' decision-making. Specifically, the study addressed three research questions:

1. Does collaborating with a mentor and peers influence the quality of preservice teachers' data-based decision making?
2. Does collaborating with a mentor and peers influence the confidence levels of preservice teachers' in regards to data-based decision making?
3. Does collaborating with a mentor and peers influence the value preservice teachers place on data-based decision making to make instructional decisions?

Quality of Preservice Teachers' Data Based Decision Making

To answer the first research question regarding the quality of data-based decision making (DBDM), a Mann Whitney U test was used to determine if there was a pre-existing difference on the distribution of the pretest scores between the experimental and control group on the Data-Based Decision Making test. The test consisted of five questions, each of which had two to three intended answers according to the USDOE (2011) and addressed the five areas of DBDM (e.g., data location, question posing). Table 1 provides a summary of descriptive statistics resulting from the DBDM pre- and posttest. The mean pretest score for the experimental group was 6.68 and the mean for the control group was 7.43 (out of 26 possible), with no statistically significant difference (*p* = 0.4021) between the group's pretest scores (see Table 2). A Mann Whitney U test was also used for the second analysis to determine if there were differences on the distributions of the gain scores (posttest - pretest) between the experimental and control conditions on the DBDM test. The mean posttest gain score for the experimental group was 2.09 whereas the gain score for the control group was -0.04. Results from the Mann-Whitney U test indicate that the difference between the experimental and control groups was statistically significant, with the experimental group outperforming the control group on the DBDM posttest (*p* = 0.0272).

Table 1

Descriptive Statistics of the Quality Measure (Data-Based Decision Making Test)

	n	Pretest *M*	Pretest Median	Posttest *M*	Posttest Median	Gain Score *M*
Control	23	7.43	7	7.39	7	-0.04
Experimental	22	6.68	7	8.77	9	2.09

Table 2

Mann Whitney U Test Results for the Quality Measure (Data-Based Decision Making Test)

Difference	Rank	*p*-Value
Pretest Scores Experimental - Control	479	0.4021
Gain Scores Experimental - Control	590.5	0.0272*

The participants were asked specific questions from five different areas of data-based decisions: data location, data comprehension, data interpretation, data use, and question posing (USDOE, 2011). The results for these specific areas are provided in Table 3. The experimental group mean for each tested area was arithmetically higher than that of the control group on the posttest. For example, in the area of Data Use the control group scored a mean of 3.04 on the posttest and the experimental group scored 3.77 (out of 12 possible). Additionally, the gain score in each area was arithmetically higher for the experimental group. For example, in the area of Question Posing the experimental group gained 1.04 on their posttest compared to the control group which gained 0.09 (out of 6 possible). Only one of the five areas, Question Posing, showed a statistically significant difference in the gain scores between the experimental and control group ($p = 0.0078$).

Table 3

Descriptive Statistics of the Five Areas of Data-Based Decision Making on the Quality Measure

		Data Location (2 possible)			Data Comprehension (2 possible)			Data Interpretation (4 possible)			Data Use (12 possible)			Question Posing (6 possible)		
	n	Pretest *M*	Posttest *M*	Gain	Pretest *M*	Posttest *M*	Gain	Pretest *M*	Posttest *M*	Gain	Pretest *M*	Posttest *M*	Gain	Pretest *M*	Posttest *M*	Gain
Control Group	23	0.83	0.91	0.08	0.74	0.83	0.09	0	0.04	0.04	3.39	3.04	-0.35	2.48	2.57	0.09
Experimental Group	22	0.77	1.0	0.23	0.59	0.95	0.36	0	0.09	0.09	3.36	3.77	0.41	1.91	2.95	1.04
Gain (*p*-Value)		0.2881			0.2942			0.5457			0.3232			0.0078*		

Confidence Levels of Preservice Teachers

To answer the second research question regarding the confidence levels of the preservice teachers, participants completed a tool that measured the level to which they agreed or disagreed (on a 1-5 point Likert-type scale) with statements regarding their confidence levels with making appropriate decisions in the five areas of DBDM (e.g., data use, data comprehension). Table 4 provides a summary of descriptive statistics from the results of this assessment. A Mann Whitney U test was used to determine if there was a pre-existing difference on the distribution of the pretest scores between the experimental and control group on the confidence tool. The mean score for the experimental group was 18.88 and the mean for the control group was 18.06 (out of 20 possible) with no statistically significant difference (p = 0.2037) between the groups' pretest scores (see Table 5). A second Mann Whitney U test was used to determine if there were differences on the distributions of the gain scores (posttest - pretest) between the experimental and control conditions on the confidence tool. The mean posttest gain score for the experimental group was 2.35 whereas the gain score for the control group was 0.51, and the difference was statistically significant (p = 0.0054). These results indicated that the experimental group was more confident, based on this tool, than their control group peers after the intervention.

Table 4

Descriptive Statistics of the Social Validity and Confidence Tool

		Confidence			Validity		
	n	Pretest *M*	Posttest *M*	Gain Score *M*	Pretest *M*	Posttest *M*	Gain Score *M*
Control Group	23	18.06	18.57	0.51	4.35	4.35	0
Experimental Group	22	18.88	21.23	2.35	4.25	4.4	0.15

Table 5

Mann Whitney U Test Results for the Confidence Tool

Difference	Rank	*p*-Value
Confidence Pretest Scores Experimental – Control	562	0.2037
Confidence Gain Scores Experimental – Control	618	0.0054*

Preservice Teachers Value of Data Based Decision Making to Make Instructional Decisions

To answer the third research question regarding preservice teachers' views of the value of DBDM for instructional decisions, a survey that had the statement "Data based decision making is valuable for making instructional decisions" was given to the participants. Participants rated how much they agreed or disagreed with the statement on a 1-5 Likert-type scale. Table 6 provides a summary of descriptive statistics from this measure. A Mann Whitney U test was used to determine if there was a pre-existing difference on the distribution of the pretest scores between the experimental and control group on this statement. The mean score for the experimental group was 4.25 and the mean for the control group was 4.35 (out of 5 possible) with no statistically significant difference (p = 0.5347) between the groups. A second Mann Whitney U test was used to determine if there were differences on the distributions of the gain scores (posttest - pretest) between the experimental and control conditions on the validity statement. The mean posttest gain score for the experimental group was 0.15 whereas the gain score for the control group was 0, with no statistically significant difference between the groups (p = 0.126).

Table 6

Mann Whitney U Test Results for the Social Validity Tool

Difference	Rank	*p*-Value
Validity Pretest Scores Experimental -Control	481.5	0.5347
Validity Gain Scores Experimental – Control	551.5	0.126

Treatment Fidelity and Interrater Agreement

Procedural fidelity of implementation of the 30-minute whole group instruction on data-based decision making was assessed by one of the course instructors who observed the instruction. To determine the fidelity the investigator divided the total number of correctly implemented components by the total number of components (18 components). Treatment fidelity for whole-group instruction was 100%. Procedural fidelity of intervention implementation for all five work groups from the experimental condition was also assessed. All five groups audio recorded their group discussion, and the investigator listened to each recording to assess the fidelity of implementation to the protocol by dividing the total number of correctly implemented components by the total number of components (28 components). Treatment fidelity was assessed using the checklist/protocol for 100% of the groups. Fidelity ranged from 93% to 100% of intervention components met with a mean of 96%.

To check the inter-rater reliability of scoring the Data-Based Decision Making test, 27% of the assessments were randomly selected and scored by the investigator and a second trained graduate researcher blind to the research conditions. Agreement ranged from 73% to 100% with a mean of 92%. Only two scores were below 85%.

Chapter 5

Discussion

Summary of Findings

The purpose of this study was to teach preservice teachers to make data-based decisions through a collaborative Professional Learning Community (PLC)-type approach. This discussion section will include a summary of the overall results, a comparison of the current findings to previous literature, implications for practice and future research, and limitations. The findings of this study support the hypothesis that collaborating with a mentor and peers influenced the quality of preservice teachers' data-based decision making (DBDM). Given the random assignments of participants, lack of statistically significant differences on pretest scores, and presence of statistically significant differences on posttest scores, the results of this study support that collaboration with mentors and peers improves the quality of preservice teachers' data-based decision making. The results of this study also support the hypothesis that this collaboration influenced the confidence levels of preservice teachers in making data-based decisions. Contrarily, the results from this study do not indicate that the collaboration influenced the value preservice teachers' had on DBDM to make instructional decisions.

This intervention was conducted by trained mentors with experience leading inservice teachers to use data to make instructional decisions in the public school setting. Furthermore, the mentors had semi-scripted questions and answers that allowed the dialogue to match the expectations of the Data-Based Decision Making test while still using different types of data (e.g., during the intervention preservice teachers viewed CBM and MAP assessments results compared to the pre- and posttest where they viewed vocabulary, comprehension, and sight reading scores) and question prompts so that the results were not complicated by training effects. The mentors routinely asked the participants questions such as “Is there additional information you’d like to know about [student’s] reading ability?, What type of instruction might be best for [student]?, and How would looking at the different strands of the MAP or areas of the F&P help determine what [student’s] needs may be?” The participants had engaging conversations with their peers to determine best-practice answers, with guidance from the mentors.

Relation to Current Literature

Six major findings were summarized in the review of the literature regarding DBDM, preservice teachers, and Professional Learning Communities (PLCs). These major findings will be reviewed in light of the current study which suggests that a collaborative intervention can help increase the quality of preservice teachers’ DBDM. Some of the major findings will be reviewed in Implications for Future Research. The first key finding was from the United States Department of Education (USDOE) survey and interview studies. The USDOE suggested that districts have data systems in place, but teachers struggle to make linkages between the data and how to interpret data to aid in instructional decision-making. While the current study contains a different sample of participants, the results suggest that preservice teachers are better able to make connections between data and instructional decisions after participating in a collaborative intervention. More specifically, preservice teachers from the control group were not able to make decisions of as high quality as the intervention group on the posttest measure of quality. DuFour et al. (2005) highlighted that schools suffer from the "DRIP syndrome—Data Rich/Information Poor" (p.40); the inservice teachers from the USDOE studies struggled with using data, whereas the participants in the experimental group of the current study had a lot of data and were able to make quality instructional decisions with the information. This study extends the qualitative nature of the USDOE studies by including an experimental design with quantitative results, and suggests that a collaborative intervention may increase the quality of preservice teachers' data-based decisions.

A second key finding of the literature review was that teachers have accurate perceptions of children's ability on standardized measures of achievement, but their perceptions on more direct assessments are not adequate (Begeny, et al., 2008; Feinberg & Shapiro, 2003). While this was not directly assessed in the current study, it was evidenced anecdotally in the recorded intervention conversations. For example, one of the case study students reviewed during the intervention performs low on the Curriculum Based Measurement (CBM) reading probe, but high on standardized assessments. The participants had conversations about how his test score may have been an aberrant result. The mentors guided the participants to better understand how to interpret both standardized measures of achievement and direct assessments.

A third key finding was that preservice teachers have received instruction on DBDM via collaboration with mentor teachers, explicit instruction, and the use of case studies as described in qualitative studies (Buck et al., 2010; Graham, 2005; Hawkins et al., 2005). In the current study the preservice teachers received similar instruction (e.g., mentors, case studies), and a randomized, experimental design strengthens the results and interpretations. The types of summative assessment data reviewed by the preservice teachers in the current study were also different than the formative assessments reviewed in both the Buck et al. (2010) and Graham (2005) studies.

A fourth key finding of the literature review was that there is a need for more studies on teacher preparation (Wilson et al., 2002). Existing studies highlight excellent teacher preparation program components (e.g., blend of theory, pedagogy, and practice, collaboration to build professional communities) (Brownell et al., 2005), but quantitative, experimental analyses of effective teacher preparation components and interventions is needed. The current study adds to the literature on teacher preparation and provides evidence of an intervention that teaches preservice educators how to make higher-quality data-based decisions. The current study goes beyond general information about excellent teacher preparation program components and gives specific details about an effective way to collaborate with future educators.

A fifth key finding was that PLCs are a common way to work with inservice teachers and when educators work within a PLC framework they may be able to impact student achievement (Vescio et al., 2008). The current study extends the work of PLCs into preservice teacher preparation. The current study also emphasizes the collaborative nature of PLCs and how working together with a mentor and peers enables preservice teachers to make higher-quality data-based decisions. The richness of the conversations between the mentors and peers were caught on the audio of the intervention. For example, the mentors asked preservice teachers what additional information they'd like to know about a particular student. Two participants responded as follows:

> I think I'd be apprehensive putting a student with a lot of green scores into a tier 3. I think I'd start with a tier 2…I think looking at strands of the MAP would help…I would want to know the different scores [strands] on the MAP, or maybe like if there is a way to see what her comprehension is compared to her oral reading fluency. Maybe she can read super good but is not comprehending… (Audio from April 23, 2013).

Providing the intervention through a collaborative, PLC-type framework may have enhanced the conversations of the preservice teachers and their mentors, thus enhancing preservice teachers' understanding of using data to make decisions.

The results of this study directly extend or tangentially relate to many of the key findings from the review of literature on DBDM, preservice teacher preparation, and PLCs. There continue to be implications for future research to extend these findings. There are also important implications for practice, as explained in the next section.

Implications for Practice

This study has several implications for practice at both the college and K-12 levels. The information generated by this study can impact both future and current teachers, as well as faculty at the university level who are teaching these aspiring educators.

Implications for the college level. There are at least three key implications of the current study for practitioners at the college level. First, the results indicate that the intervention was effective in improving the quality of data-based decisions in preservice teachers. Staff who teach preservice teachers should continue to prepare future educators with the information they need to make high quality data-based decisions. More specifically, college instructors should ensure that they are teaching preservice teachers how to analyze data and use it to inform instructional practice to ultimately impact student achievement.

The second key implication for practice highlights *how* university faculty can teach preservice teachers about DBDM. Relationships between key stakeholders in education are imperative, and continuing the relationships between university mentors and school districts, particularly in the practicum placement and training of preservice teachers, provides educators at all levels the chance to strengthen their programs or schools. More specifically, school districts can utilize the expertise of university staff while the university instructors can promote positive relations with local schools. Furthermore, partnerships between universities and local school districts can provide a natural bridge from research to practice through college consultation and regular evaluations of program benefits and challenges.

The third implication of the current study on practitioners at the college level is the need to incorporate DBMD into both the curricula and practica provided to preservice teachers. To help bridge the research to practice gap, it is important that preservice educators have prolonged experiences with DBDM in the field. Incorporating parts of this intervention into preservice teachers' practica may allow them to see the link between what they learn in their university coursework with what occurs in the field. The preservice teachers could bring data from their school sites and work with university mentors and their peers to analyze the data and make instructional decisions.

Implications for the K-12 school setting. The results of this study could also extend into work with inservice teachers. The results of the USDOE studies indicate that inservice teachers struggle to make connections between data and instruction. University mentors or district staff with expertise on collaboration and DBDM should work with their colleagues to use available data to inform instructional decisions. The inservice teachers could use data from their classrooms and gather as PLCs to analyze results and change their instruction (DuFour, DuFour, Eaker, & Many, 2010). It is imperative that PLCs not only "admire" the data, but also talk about instructional strategies that may have impactful results. The guiding questions from this intervention could be used in those data meetings to help focus the attention of educators on available data to ensure their students are placed in appropriate flexible groups with instruction targeted toward their specific needs.

Implications for Research

Though the results of this study are promising, several suggestions to enhance future research could strengthen the evidence base for interventions on DBDM for preservice teachers. First, the intervention was about 75 minutes in length on one afternoon. Had the intervention been longer or extended across different sessions, the results of the experimental group may have been more pronounced. Conducting the intervention at different points during the preservice teachers' program might reveal continued growth in the quality of their data-based decisions. The effect of an ongoing mentorship rather than a one-time dosage should also be studied.

A second way to enhance the research would be to look more directly at the individuals involved. For example, a researcher could look more explicitly at who the mentors are (e.g., graduate students vs. practitioners) to determine which type of interventionist may have the most impact. Researchers could look more explicitly at characteristics of the participants as well. The current study explored preservice elementary-level teachers' ability to make data-based decisions. It is important that all K-12 educators (including special education teachers, English Language Learner instructors, Gifted and Talented teachers) make data-based decisions in their instruction. Replicating the current study with different groups of participants would add to the research base. In the same essence, the intervention could be conducted with current teachers who may be struggling with making data-based decisions in order to increase the quality of DBDM in the field.

A third way to extend the results of this study would be to conduct follow-up research on the current participants to see if they are making data-based decisions once they are in the field. Furthermore, it would be important to assess if their instructional decisions were increasing student achievement. If a longitudinal view revealed improved quality of data-based decisions and subsequent increases in student achievement, the rationale for teaching preservice teachers how to make data-based decisions would be enhanced.

Additional implications for future research can be viewed in response to findings from the literature review on DBDM. First, the USDOE reported that teachers struggled to interpret data to make instructional decisions. A replication of this study and research on how inservice teachers are doing now would be beneficial. Second, general education teachers need to make data-based decisions that impact student achievement. More studies and interventions with this population would enhance literature in this area and help in the development of guidelines for best practice. Third, teacher perceptions about data should be reviewed after they have had experience with making data-based decisions. Adding this component to a replication of the current study may indicate that the intervention helps combat inaccurate perceptions. Fourth, more quantitative studies, in general, are needed on teacher preparation programs, and how DBDM instruction can enhance preservice teacher learning. Fifth, PLCs should be well-defined and quantitatively studied in relation to DBDM and its impact on student achievement. Moving forward with research in all of these areas will continue to enhance an understanding of effective instruction in DBDM, while also allowing educators a means to increase student achievement.

Limitations

Limitations of the current study should be taken into account when interpreting the results and planning for future research. While the intervention was implemented at an acceptable level, each group had a different experience with their mentor and the duration of the intervention was markedly different for some groups compared to others. Also, though the nature of the study required participants to be preservice teachers, convenience sampling was used (i.e., participants recruited from one course rather than randomly drawn from the preservice teacher program) and the participants may not be representative of the population. Furthermore, the sample size was small and the majority of the participants were Caucasian women ages 20 to 30 years old. Additionally, the participants had been in an environment where teaching faculty adhered to a data-based view of teaching. In a different institution the participants may not be engaged in that time of instruction.

The dependent variable used in this study represented an additional limitation and could have benefited from some modification. The Data-Based Decision Making test was based on the USDOE (2011) qualitative study and was adapted for this study. The questions were provided as well as intended answers although the link between the questions and their answers were not always explicit. The questions could have been rewritten to better gauge the breadth and depth of the participants responses. For example, few participants fully met criteria for the intended answers in the area of Data Interpretation (e.g., few wrote something about the concept of measurement error and variability after being asked the question "What group would you put [student] into? Why?"). Another limitation was that only one dependent measure was used to capture the effect of the intervention. In the future it would be advantageous to consider additional dependent variables or measurement tools.

A final limitation was that only one question was asked to answer the research question "Does collaborating with a mentor and peers influence the value preservice teachers place on data-based decision making to make instructional decisions?" The preservice teachers rated themselves relatively high on the pretest (4.3 out of 5) so there was not a lot of opportunity to show growth on this question. In the future it would be helpful to have additional questions and/or a different tool to help measure how much value preservice teachers put on DBDM to inform their instructional decisions.

Conclusion

Achievement gaps in reading and math are startling (Lee & Reeves, 2012) and teachers need the tools to provide effective instruction that increases student learning (Taylor et al., 2003). Providing teachers with the data to make decisions about their instruction has shown promise of increasing student achievement (Fuchs et al., 1994). Currently there is a failure to link teaching strategies with student achievement (Foorman, 2007; Foorman & Schatschneider, 2003). Teachers also do not have the appropriate data or tools to plan and individualize instruction (USDOE, 2007). The intervention implemented in this study provided preservice teachers with skills to increase the quality of data-based decisions. While there is not a direct link to student achievement, it is encouraging that these future educators have acquired the skills to work collaboratively to gather data and make data-based decisions that could ultimately be used to enhance instruction and increase student achievement.

References

Begeny, J. C., Eckert, T. L., Montarello, S. A., & Storie, M. S. (2008). Teachers' perceptions of students' reading abilities: An examination of the relationship between teachers' judgments and students' performance across a continuum of rating methods. *School Psychology Quarterly, 23*(1), 43-55. doi: 10.1037/1045-3830.23.1.43

Berry, B., Johnson, D., & Montgomery, D. (2005). The power of teacher leadership. *Educational Leadership, 62*(5), 56. Retrieved from http://login.ezproxy.lib.umn.edu/login?url=http://search.proquest.com/docview/62082469?accountid=14586

Bolam, R., McMahon, A., Stoll, L., Thomas, S., & Wallace, M. (2005). Creating and sustaining effective professional learning communities. General Teaching Council for England. Research Report RR637. Retrieved from: http://www.educationscotland.gov.uk/images/Creating%20and%20Sustaining%2OPLCs_tcm4-631034.pdf

Brantlinger, E., Jimenez, R., Klinger, J., Pugach, M., Richardson, V. (2005). Qualitative studies in special education. *Exceptional Children, 71*(2), 195-207. Retrieved from http://login.ezproxy.lib.umn.edu/?url=http://search.proquest.com/docview/62129561?accountid=14586

Brownell, M. T., Ross, D. D., Colon, E. P., & McCallum, C. L. (2005). Critical features of special education teacher preparation: "A comparison with general teacher education". *Journal of Special Education, 38*(4), 242-252. doi: 10.1177/00224669050380040601

Buck, G.A., Trauth-Nare, A., & Kaftan, J. (2010). Making formative assessment discernible to preservice teachers of science. *Journal of Research in Science Teaching, 47*(4), 402-421. doi: 10.1002/tea.20344

Burns, M.K., & Gibbons, K.A. (2008). *Implementing response to intervention in elementary and secondary schools.* New York, NY: Routledge.

Deno, S.L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*(3), 219-232.

DuFour, R., Eaker, R., & DuFour, R. (2005). *On common ground: The power of professional learning communities*. Bloomington, IN: Solution Tree.

DuFour, R., DuFour, R., Eaker, R., and Many, T. (2010). *Learning by doing: A handbook for professional learning communities at work.* (2^{nd} ed.). Bloomington, IN: Solution Tree.

Eaker, R., DuFour, R., & DuFour, R. (2002). *Getting started: Reculturing schools to become professional learning communities.* Bloomington, IN: National Educational Service.

Eckert, T. L., Dunn, E. K., Codding, R. S., Begeny, J. C., & Kleinmann, A. E. (2006). Assessment of mathematics and reading performance: An examination of the correspondence between direct assessment of student performance and teacher report. *Psychology in the Schools, 43*(3), 247-265. doi: 10.1002/pits.20147

Feinberg, A. B., & Shapiro, E. S. (2003). *Accuracy of teacher judgments in predicting oral reading fluency* Retrieved from http://login.ezproxy.lib.umn.edu/login ?url=http://search.proquest.com/docview/62235642?accountid=14586

Foorman, B. R. (2007). Primary prevention in classroom reading instruction. *Teaching Exceptional Children, 39*(5), 24-30. Retrieved from http://web.ebscohost.com. ezp1.lib.umn.edu/ehost/pdfviewer/pdfviewer?vid=2&hid=13&sid=2aa2f76e-4e1f-4729-a563-486c2907143c%40sessionmgr12

Foorman, B.R., & Schatschneider, C. (2003). Measurement of teaching practices during reading/language arts instruction and its relationship to student achievement. In S. Vaughn and K.L. Briggs (Eds.), *Reading in the classroom: Systems for observation of teaching and learning* (pp. 1-30). Baltimore, MD: Brookes Publishing Co.

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1984). The effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal, 21*(2), 449-460. doi: 10.3102/00028312021002449

Fuchs, L. S., Fuchs, D., Hamlett, C., Phillips, N.B., Bentz, J. (1994). Classwide curriculum-based measurement: Helping general educators meet the challenge of student diversity. *Exceptional Children, 60*(6), 518-537. Retrieved from http://login.ezproxy.lib.umn.edu/?url=http://search.proquest.com/docview/62811912?accountid=14586

Fuchs, D., Fuchs, L.S., & Stecker, P.M. (2010). The "blurring" of special education in a new continuum of general education placement and services. *Exceptional Children, 76*(3), 301-323. Retrieved from http://login.ezproxy.lib.umn.edu.ezp2.lib.umn.edu/login?url=http://search.proquest.com.ezp2.lib.umn.edu/docview/742865988?accountid=14586

Fuchs, L. S., Hamlett, D. F. C. L., & Stecker, P. M. (1991). Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Educational Research Journal, 28*(3), 617-641. doi: 10.3102/00028312028003617

Gersten, R., Compton, D., Connor, C.M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W.D. (2009). *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 from http://ies.ed.gov/ncee/wwc/publications/practiceguides/.

Gersten, R., Fuchs, L.S., Compton, D., Coyne, M., Greenwood, C., Innocenti, M.S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children, 71*(2), 149-164. Retrieved from http://login.ezproxy.lib.umn.edu/?url=http://search.proquest.com/docview/62126279?accountid=14586

Gishchlar, K.L., Hojnoski, R.L., & Missall, K.N. (2009). Improving child outcomes with data-based decision making: Interpreting and using data. *Young Exceptional Children, 13*(1), 1-18. doi: 10.1177/1096250609346249

Graham, P. (2005). Classroom-based assessment: Changing knowledge and practice through preservice teacher education. *Teaching and Teacher Education, 21*, 607-621. doi: 10.1016/j.tate.2005.05.001

Graney, S. B., & Shinn, M. R. (2005). Effects of reading curriculum-based measurement (R-CBM) teacher feedback in general education classrooms. *School Psychology Review, 34*(2), 184-201. Retrieved from http://login.ezproxy.lib.umn.edu/login?url=http://search.proquest.com/docview/61924658?accountid=14586

Gunstone, R.F., Slattery, M., Baird, J.R., & Northfield, J.R. (1993). A case study exploration of development in preservice science teachers. *Science Education, 77*(1), 47-73. doi: 10.1002/sce.3730770104

Hamilton, L., Halverson, R., Jackson, S., Mandinach, E., Supovitz, J., & Wayman, J. (2009). *Using student achievement data to support instructional decision making* (NCEE 2009-4067). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/.

Hamilton, C., & Shinn, M. R. (2003). Characteristics of word callers: An investigation of the accuracy of teachers' judgments of reading comprehension and oral reading skills. *School Psychology Review, 32*(2), 228-240. Retrieved from http://login.ezproxy.lib.umn.edu/login?url=http://search.proquest.com/docview/618898615?accountid=14586

Hawkins, R.O., Kroeger, S.D., Musti-Rao, S., Barnett, D.W., & Ward, J.E. (2008). Preservice training in response to intervention: Learning by doing an interdisciplinary field experience. *Psychology in the Schools, 45*(8), 745-762. doi: 10.1002/pits.20339

Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research, 59*(3), 297-313. Retrieved from http://login.ezproxy.lib.umn.edu/login?url=http://search.proquest.com/docview/63060798?accountid=14586

Hollins, E., McIntyre, L., DeBose, C., Hollins, K., & Towner, A. (2004). Promoting a self-sustaining learning community: Investigating an internal model for teacher development. *International Journal of Qualitative Studies in Education, 17*(2), 247-264. Retrieved from http://login.ezproxy.lib.umn.edu/login?url=http://search.proquest.com/docview/62121150?accountid=14586

Individuals with Disabilities Education Improvement Act of 2004, Pub. L. No. 108-446 (2004). Retrieved from http://idea.ed.gov/.

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

Kanold, T., Toncheff, M., & Douglas, C. (2008). Two high school districts recite the ABCs of professional learning communities. *Journal of Staff Development, 29*(3), 22-24. Retrieved from http://login.ezproxy.lib.umn.edu/login?url=http://search.proquest.com/docview/61968692?accountid=14586

Lee, J., & Reeves, T. (2012). Revising the impact of NCLB high-stakes accountability, capacity, and resources: State NAEP 1999-2009 reading and math achievement gaps and trends. *Educational Evaluation and Policy Analysis, 34*(2), 209-231. doi: 10.3102/0162373711431604

Louis, K. S., & Marks, H. (1998). Does professional community affect the classroom? Teachers' work and student experiences in restructuring schools. *American Journal of Education, 106*(4). Retrieved from http://login.ezproxy.lib.umn.edu/login?url=http://search.proquest.com/docview/62494944?accountid=14586

Maheady, L., Jabot, M., Rey, J., & Michielli-Pendl, J. (2007). An early field-based experience and its impact on preservice candidates' teaching practice and their pupils' outcomes. *Teacher Education and Special Education, 30*(1), 24-33. doi: 10.1177/088840640703000103

Maxwell, G. M., Huggins, K. S., & Scheurich, J. J. (2010). How one historically underperforming rural and highly diverse high school achieved a successful turnaround. *Planning and Changing, 41*(3), 26. Retrieved from http://login.ezproxy.lib.umn.edu/login?url=http://search.proquest.com/docview/964184963?accountid=14586

Mercier Smith, J.L., Fien, H., Basaraba, D., & Travers, P. (2009). Planning, evaluating, and improving tiers of support in beginning reading. *TEACHING Exceptional Children, 41*(5), 16-22. Retrieved from http://login.ezproxy.lib.umn.edu. ezp1.lib.umn.edu/login?url=http://search.proquest.com.ezp1.lib. umn.edu/docview/61877119?accountid=14586.

Murphy, C. U. & Lick, D. W. (2005). *Whole-faculty study groups: Professional learning communities that target student learning*. (3rd ed.). Thousand Oaks , CA : Corwin Press.

National Center for Education Statistics (2011). The Nation's Report Card: Reading 2011(NCES 2012–457). Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

National Commission on Excellence in Elementary Teacher Preparation for Reading Instruction. (2003). An executive summary of the national commission on excellence in elementary teaching preparation for reading instruction. International Reading Association, Newark: DE.

Nilson, L.B. (2003). *Teaching at its best: A research-based resource for college instructors.* (2nd ed.). San Francisco, CA: Anker Publishing, Jossey-Bass.

No Child Left Behind Act of 2001, 20 U.S.C. 70 § 6301 et seq. (2002).

Phillips, J. (2003). Powerful learning: Creating learning communities in urban school reform. *Journal of Curriculum and Supervision, 18*(3), 240-258. Retrieved from http://login.ezproxy.lib.umn.edu/login?url=http://search.proquest.com/docview/62 230290?accountid=14586

Salvia, J., Ysseldyke, J.E., & Bolt, S. (2007). *Assessment in special and inclusive education*. (10th ed.). Boston, MA: Houghton Mifflin Company.

Stecker, P. M., & Fuchs, L. S. (2000). Effecting superior achievement using curriculum-based measurement: The importance of individual progress monitoring. *Learning Disabilities: Research & Practice, 15*(3), 128-134. Retrieved from http://login.ezproxy.lib.umn.edu/login?url=http://search. proquest.com/docview/62319860?accountid=14586

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement t to improve student achievement: Review of research. *Psychology in the Schools, 42*(8), 795-819. Retrieved from http://login.ezproxy.lib.umn. edu/login?url=http://search.proquest.com/docview/61995865?accountid=14586

Strahan, D. (2003). Promoting a collaborative professional culture in three elementary schools that have beaten the odds. *The Elementary School Journal, 104*(2), 127-146. Retrieved from: http://www.jstor.org/stable/3202983

Supovitz, J. A. (2002). Developing communities of instructional practice. *Teachers College Record*, *104*(8), 1591-1626. Retrieved from http://login.ezproxy.lib. umn.edu/login?url=http://search.proquest.com/docview/62216328?accountid=145 86

Supovitz, J. A., & Christman, J. B. (2003). *Developing communities of instructional practice: Lessons from Cincinnati and Philadelphia. CPRE policy briefs. RB-3 39.* ().Consortium for Policy Research in Education. University of Pennsylvania, 3440 Market Street Suite 560, Philadelphia, PA 19104. Retrieved from http://login.ezproxy.lib.umn.edu/login?url=http://search.proquest.com/docview/62000382?accountid=14586

Taylor, B.M., Pearson, P.D., Peterson, D.S., Rodriguez, M.C. (2003). Reading growth in high-poverty classrooms: The influence of teacher practices that encourage cognitive engagement in literacy learning. *The Elementary School Journal, 104*(1), 3-28. Retrieved from http://login.ezproxy.lib.umn.edu/?url=http://search.proquest.com/docview/62182778?accountid=14586

Thompson, B., Diamond, K.E., McWilliam, R., Snyder, P. (2005). Evaluating the quality of evidence from correlational research for evidence-based practice. *Exceptional Children, 71*(2), 181-194. Retrieved from http://education.uoregon.edu/grantmatters/pdf/DR/Correlational.pdf

U.S. Department of Education, Office of Planning, Evaluation and Policy Development. (2007). *Teachers' use of student data management systems to improve instruction* (Contract number ED-04-CO-0040/0002). Retrieved from http://ctl.sri.com/publications/downloads/TeacherDataUseBrief09_10_07.pdf

U.S. Department of Education, Office of Planning, Evaluation and Policy Development. (2008). *Teachers' use of student data systems to improve instruction: 2005 to 2007* (Contract number ED-04-CO-0040/0002). Retrieved from http://www2.ed.gov/rschstat/eval/tech/teachers-data-use-2005-2007/teachers-data-use-2005-2007.pdf

U.S. Department of Education, Office of Planning, Evaluation and Policy Development. (2009). *Implementing data-informed decision making in schools: Teacher access, supports and uuse* (Contract number ED-01-CO-0040 Task 0002). Retrieved from www2.ed.gov/.../data-informed-decision/data-informed-decision.doc

U.S. Department of Education, Office of Planning, Evaluation, and Policy Development. (2010). *Use of education data at the local level from accountability to instructional improvement* (Contract number ED-01-CO-0040 Task 0002). Retrieved from http://www2.ed.gov/rschstat/eval/tech/use-of-education-data/use-of-education-data.pdf

U.S. Department of Education, Office of Planning, Evaluation and Policy Development. (2011). *Teachers' ability to use data to inform instruction: Challenges and supports* (Contract number ED-01-CO-0040/0002). Retrieved from www2.ed.gov/rschstat/eval/data-to-inform-instruction/report.doc

Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education, 24,* 80-91. doi: 10.1016/j.tate.2007.01.004.

Wayman, M., Wallace, T., Wiley, H. I., Ticha, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *Journal of Special Education, 41*(2), 85-120. doi:http://dx.doi.org/10.1177/00224669070410020401

Wideen, M., Mayer-Smith, J., & Moon, B. (1998). A critical analysis of the research on learning to teach: Making the case for an ecological perspective on inquiry. *Review of Educational Research, 68*(2), 130-178. doi:http://dx.doi.org/10.3102/00346543068002130

Wilson, S.M., Floden, R.E., & Ferrini-Mundy, J. (2002). Teacher preparation research: An insider's view from the outside. *Journal of Teacher Education*, *53*(3), 190-204. doi: 10.1177/0022487102053003002

Appendix A: Quality indicators for experimental and quasi-experimental research designs taken directly from Gersten et al., 2005

Essential Quality Indicators

Quality Indicators for Describing Participants

1. Was sufficient information provided to determine/confirm whether the participants demonstrated the disability(ies) or difficulties presented?

2. Were appropriate procedures used to increase the likelihood that relevant characteristics of participants in the sample were comparable across conditions?

3. Was sufficient information given characterizing the interventionists or teachers provided? Did it indicate whether they were comparable across conditions?

Quality Indicators for Implementation of the Intervention and Description of Comparison Conditions

4. Was the intervention clearly described and specified?

5. Was the fidelity of implementation described and assessed?

6. Was the nature of services provided in comparison conditions described?

Quality Indicators for Outcome Measures

7. Were multiple measures used to provide an appropriate balance between measures closely aligned with the intervention and measures of generalized performance?

8. Were outcomes for capturing the intervention's effect measured at the appropriate times?

Quality Indicators for Data Analysis

9. Were the data analysis techniques appropriately linked to key research questions and hypotheses? Were they appropriately linked to the unit of analysis in the study?

10. Did the research report include not only inferential statistics but also effect size calculations?

Desirable Quality Indicators

11. Was data available on attrition rates among intervention samples? Was severe overall attrition documented? If so, is attrition comparable across samples? Is overall attrition less than 30%?

12. Did the study provide not only internal consistency reliability but also test–retest reliability and interrater reliability (when appropriate) for outcome measures? Were data collectors and/or scorers blind to study conditions and equally (un)familiar to examinees across study conditions?

13. Were outcomes for capturing the intervention's effect measured beyond an immediate posttest?

14. Was evidence of the criterion-related validity and construct validity of the measures provided?

15. Did the research team assess not only surface features of fidelity implementation (e.g., number of minutes allocated to the intervention or teacher/interventionist following procedures specified), but also examine quality of implementation?

16. Was any documentation of the nature of instruction or series provided in comparison conditions?

17. Did the research report include actual audio or videotape excerpts that capture the nature of the intervention?

18. Were results presented in a clear, coherent fashion?

Appendix B: Quality indicators for correlational research taken directly from Thompson et al., 2005

Measurement

1. Score reliability coefficients are reported for all measured variables, based on induction from a
prior study or test manual, with explicit and reasonable justifications as regards comparabilities of (a) sample compositions and (b) score dispersions.

2. Score reliability coefficients are reported for all measured variables, based on analysis of the data in hand in the particular study.

3. Evidence is inducted, with explicit rationale, from a prior study or test manual that suggests scores are valid for the inferences being made in the study.

4. Score validity is empirically evaluated based on data generated within the study.

5. The influences of score reliability and validity on study interpretations are explicitly considered in reasonable derail.

Practical and Clinical Significance

6. One or more effect size statistics is reported for each study primary outcome, and the effect statistic used is clearly identified.

7. Authors interpret study effect sizes for selected practices by directly and explicitly comparing
study effects with those reported in related prior studies.

8. Authors explicitly consider study design and effect size statistic limitations as part of effect interpretation.

Avoiding Some Common Macro-Analytic Mistakes

9. GLM weights (e.g., beta weights) are interpreted as reflecting correlations of predictors with outcome variables only in the exceptional case that the weights indeed are correlation coefficients.

10. When noteworthy results are detected, and the origins of these effects are investigated, the interpretation includes examination of structure coefficients.

11. Interval data are not converted to nominal scale, unless such choices are justified on the extraordinary basis of distribution shapes, and the consequences of the conversion are thoughtfully considered as part of result interpretation.

12. Univariate methods are not used in the presence of multiple outcome variables.

13. Univariate methods are not used post hoc to multivariate tests.

14. Persuasive evidence is explicitly presented that the assumptions of statistical methods are sufficiently well-met for results to be deemed credible.

CIs for Reliability Coefficients, Statistics, and Effect Sizes

15. Confidence intervals are reported for the reliability coefficients derived for study data.

16. Confidence intervals are reported for the sample statistics (e.g., means, correlation coefficients) of primary interest in the study.

17. Confidence intervals are reported for study effect sizes.

18. Confidence intervals are interpreted by direct and explicit comparison with related Cls from prior studies.

Appendix C: Credibility measures and quality indicators for qualitative research taken directly from Brantlinger et al., 2005

Credibility Measures

1. Triangulation—search for convergence of, or consistency among, evidence from multiple and varied data sources (observations/interviews; one participant & another; interviews/documents).

   • Data triangulation—use of varied data sources in a study.

   • Investigator triangulation—use of several researchers, evaluators, peer debriefers.

   • Theory triangulation—use of multiple perspectives to interpret a single set of data.

   • Methodological triangulation—use of multiple methods to study a single problem.

2. Disconfirming evidence—after establishing preliminary themes/categories, the researcher looks for evidence inconsistent with these themes (outliers); also known as negative or discrepant case analysis.

3. Researcher reflexivity—researchers attempt to understand and self-disclose their assumptions, beliefs, values, and biases (i.e., being forthright about position/perspective).

4. Member checks—having participants review and confirm the accuracy (or inaccuracy) of interview transcriptions or observational field notes.

• First level—taking transcriptions to participants prior to analyses and interpretations of results.

• Second level—taking analyses and interpretations of data to participants (prior to publication) for validation of (or support for) researchers' conclusions.

5. Collaborative work—involving multiple researchers in designing a study or concurring about conclusions to ensure that analyses and interpretations are not idiosyncratic and/or biased; could involve interrater reliability checks on the observations made or the coding of data. (The notion that persons working together will get reliable results is dependent on the "truth claim" assumption that one can get accurate descriptions of situational realities.)
6. External auditors—using outsiders (to the research) to examine if, and confirm that, a researcher's inferences are logical and grounded in findings.
7. Peer debriefing—having a colleague or someone familiar with phenomena being studied review and provide critical feedback on descriptions, analyses, and interpretations or a study's results.
8. Audit trail—keeping track of interviews conducted and/or specific times and dates spent observing as well as who was observed on each occasion; used to document and substantiate that sufficient time was spent in the field to claim dependable and confirmable results.
9. Prolonged field engagement—repeated, substantive observations; multiple, in-depth interviews; inspection of a range of relevant documents; thick description validates the study's soundness.

10. Thick, detailed description—reporting sufficient quotes and field note descriptions to provide evidence for researchers' interpretations and conclusions.
11. Particularizability—documenting cases with thick description so that readers can determine the degree of transferability to their own situations.

Quality Indicators for Types of Qualitative Studies

Interview Studies (or Interview Components of Comprehensive Studies)

1. Appropriate participants are selected (purposefully identified, effectively recruited, adequate number, representative of population of interest).
2. Interview questions are reasonable (clearly worded, not leading, appropriate and sufficient for exploring domains of interest).
3. Adequate mechanisms are used to record and transcribe interviews.
4. Participants are represented sensitively and fairly in the report.
5. Sound measures are used to ensure confidentiality.

Observation Studies (or Observation Components of Comprehensive Studies)

1. Appropriate setting(s) and/or people are selected for observation.
2. Sufficient time is spent in the field (number and duration of observations, study time span).
3. Researcher fits into the site (accepted, respected, unobtrusive).
4. Research has minimal impact on setting (except for action research, which is purposely designed to have an impact).
5. Field notes systematically collected (videotaped, audiotaped, written during or soon after observations).

6. Sound measures are used to ensure confidentiality of participants and settings.

Document Analysis

1. Meaningful documents (texts, artifacts, objects, pictures) are found and their relevance is established.
2. Documents are obtained and stored in a careful manner.
3. Documents are sufficiently described and cited.
4. Sound measures are used to ensure confidentiality of private documents.

Data Analysis

1. Results are sorted and coded in a systematic and meaningful way.
2. Sufficient rationale is provided for what was (or was not) included in the report.
3. Documentation of methods used to establish trustworthiness and credibility are clear.
4. Reflection about researchers' personal position/perspectives are provided.
5. Conclusions are substantiated by sufficient quotations from participants, field notes of observations, and evidence of documentation inspection.
6. Connections are made with related research.

Appendix D: Pretest Measurement Tool – Student Copy

Data Based Decision Making Pretest **ID # ___________**

Adapted from U.S. Department of Education Office of Planning, Evaluation, and Policy Development (Means et al., 2011)

SCENARIO

SCENARIO: Suppose that this is the fourth week of school and that you're a fourth-grade teacher planning your instruction for the remainder of this term. As shown below, you have scores from the state reading test given last spring and from a sight reading assessment and a passage comprehension test that you've had your students take during the first three weeks of school.

Table 1. Student Performance on State and Classroom Reading Tests

			2011-2012 State Achievement Test Scale				Fall 2012 Class Tests	
Student		**Race**	**Total Reading**	**Comprehension**	**Vocabulary**		**Sight Reading**	**Text Comprehension**
Andrew	M	6	398	380	415		18	7
Alexis	F	6	535	515	555		26	9
Brooklyn	F	3	503	510	495		24	10
Benjamin	M	6	533	520	545		28	11
Catrina	F	4	650	665	635		30	14
Chantal	F	6	518	520	515		22	12
Caleah	F	4	578	565	590		26	12
Donovan	M	3	593	571	615		22	8
Jordan	M	3	560	555	565		27	12
Katherine	F	6	546	558	533		28	11
Micah	M	6	415	400	430		18	7
Neven	M	3	698	683	705		32	13
Patty	F	6	421	405	437		22	9
Raul	M	4	568	585	550		28	10
Sunshine	F	3	485	505	465		24	12

Total Possible			705	705	705		32	14
Class Average			533	529	537		25	10

Race:
1= American Indian or Alaska Native
2= Asian
3= Black or African American
4= Hispanic/Latino
5= Native Hawaiian or Pacific Islander
6= White

1. Read the statement below and say whether you agree or disagree and the reasons why.

   The class average for vocabulary is 537. At least half of the students are scoring below the class average.

2. What, if anything, do these data tell you about how you might want to differentiate instruction for different students in your class?

3. Which group would you put Andrew into? *OR* Which approach would you use for Andrew? Why?

4. Which group would you put Donovan into? What is your reason for that decision?

5. What group would you put Sunshine into? Why?

Appendix E: Pretest Measurement Intended Answer Key

1. Read the statement below and say whether you agree or disagree and the reasons why.

   The class average for comprehension is 537. At least half of the students are scoring below the class average.

Component	**Skill or Concept**	**Evidence for Presence**	**Did they get this component?**	**Extension?**
Data location	Finds relevant data in a complex table or graph	Examines comprehension scores and counts the number of students that scored below 537.		
Data comprehension	Moves fluently between different representations of data	Disagrees. Seven students are scoring below the class average.		

2. What, if anything, do these data tell you about how you might want to differentiate instruction for different students in your class?

Component	Skill or Concept	Evidence for Presence	Did they get this compo nent?	Extension?
Data use	Uses subscale and item data	Discusses looking at comprehension performance, sight reading (fluency), and vocabulary scores separately to decide what to emphasize with each group and within each group, what each student is struggling with, and how to accommodate their needs.		
Data use	Understands concept of differentiating instruction based on data	Articulates a coherent, data-informed rule for grouping (i.e., includes test data but data need not be the sole criterion).		

3. Which group would you put Andrew into? *OR* Which approach would you use for Andrew? Why?

Component	Skill or Concept	Evidence for Presence	Did they get this component?	Extension?
Data use	Uses subscale and item data	Discusses comprehension performance, sight reading (fluency), and vocabulary scores separately to decide what to emphasize with		

		each group. Indicates a desire to look more closely at Andrew's performance on different items or standards to support a diagnosis of his needs.		
Question posing	Appreciates value of multiple measures	Discusses comprehension scores on both the state test and the classroom test. Indicates a desire to obtain additional information, such as a reading specialist's assessment of Andrew or notes from his second-grade teacher.		
Data use	Understands concept of differentiating instruction based on data	Discusses an individualized instructional plan for Andrew based on his assessment results.		

4. Which group would you put Donovan into? What is your reason for that decision?

Component	Skill or Concept	Evidence for Presence	Did they get this component?	Extension?
Data use	Uses subscale and item data	Notes the discrepancy between Donovan's high scores on the state test and his below-average scores on the in-class tests of reading comprehension and sight reading. Indicates desire to have more detailed information about Donovan's		

		performance on particular test items or standards. Discusses issues around possible differences in the content covered by statewide and in-class tests.		
Data interpretation	Understands the concept of measurement error and variability	Notes that Donovan's score on the in-class test of reading comprehension could be an aberrant result.		
Question posing	Appreciates value of multiple measures	Suggests getting another assessment of Donovan's reading comprehension and sight reading, either formally or through in-class observation. Mentions the potential to make a group assignment but keep the groups fluid and keep assessing and regrouping kids.		

5. What group would you put Sunshine into? Why?

Component	**Skill or Concept**	**Evidence for Presence**	**Did they get this component?**	**Extension?**
Data use	Uses subscale and item data	Notes the discrepancy between Sunshine's relatively low scores on the state test last spring and her strong performance on the in-class tests this fall.		

		Indicates desire to have more detailed information about Sunshine's performance on particular test items or standards. Discusses issues around possible differences in the content covered by statewide and in-class tests.		
Data interpretation	Understands the concept of measurement error and variability	Notes that Sunshine's state test performance last fall might underrepresent her skills.		
Question Posing	Appreciates value of multiple measures	Suggests getting additional information about Sunshine's reading ability (e.g., report card, second-grade end of semester reading test). Mentions the potential to make a group assignment but keep the groups fluid and keep assessing and regrouping kids.		

Appendix F: Posttest Measurement Tool - Student Copy

Data Based Decision Making Posttest **ID # ___________**

Adapted from U.S. Department of Education Office of Planning, Evaluation, and Policy Development (2011)

SCENARIO

SCENARIO: Suppose that this is the third week of school and that you're a second-grade teacher planning your instruction for the remainder of this term. As shown below, you have scores from the state reading test given last spring and from a sight reading assessment and a passage comprehension test that you've had your students take during the first two weeks of school.

Table 1. Student Performance on State and Classroom Reading Tests

			2011-2012 State Achievement Test Scale Score				Fall 2012 Class Tests	
Student		**Race**	**Total Reading**	**Comprehension**	**Vocabulary**		**Sight Reading**	**Text Comprehension**
Alex	M	6	403	385	420		23	12
Andrea	F	6	540	520	560		31	14
Brookie	F	3	508	515	500		29	15
Brock	M	6	538	525	550		33	16
Caitlin	F	4	655	670	640		35	19
Carly	F	6	523	525	520		27	17
Catherine	F	4	583	570	595		31	17
Daniel	M	3	598	576	620		27	13
Jeff	M	3	565	560	570		32	17
Katrina	F	6	551	563	538		33	16
Maxwell	M	6	420	405	435		23	12
Nathan	M	3	703	688	710		37	18
Paige	F	6	426	410	442		27	14
Roger	M	4	573	590	555		33	15

Susan	F	3	490	510	470		29	17
Total Possible			710	710	710		37	19
Class Average			538	534	542		30	15

Race:
1= American Indian or Alaska Native
2= Asian
3= Black or African American
4= Hispanic/Latino
5= Native Hawaiian or Pacific Islander
6= White

1. Read the statement below and say whether you agree or disagree and the reasons why.

   The class average for vocabulary is 542. At least half of the students are scoring above the class average.

2. What, if anything, do these data tell you about how you might want to differentiate instruction for different students in your class?

3. Which group would you put Alex into? *OR* Which approach would you use for Alex? Why?

4. Which group would you put Daniel into? What is your reason for that decision?

5. What group would you put Susan into? Why?

Appendix G: Posttest Measurement Intended Answer Key

1. Read the statement below and say whether you agree or disagree and the reasons why.

   The class average for comprehension is 542. At least half of the students are scoring above the class average.

Component	Skill or Concept	Evidence for Presence	Did they get this component?	Extension?
Data location	Finds relevant data in a complex table or graph	Examines comprehension scores and counts the number of students that scored below 542		
Data comprehension	Moves fluently between different representations of data	Agrees. Eight students are scoring above the class average.		

2. What, if anything, do these data tell you about how you might want to differentiate instruction for different students in your class?

Component	Skill or Concept	Evidence for Presence	Did they get this component?	Extension?
Data use	Uses subscale and item data	Discusses looking at comprehension performance, sight reading (fluency), and vocabulary scores separately to decide		

		what to emphasize with each group and within each group, what each student is struggling with, and how to accommodate their needs.		
Data use	Understands concept of differentiating instruction based on data	Articulates a coherent, data-informed rule for grouping (i.e., includes test data but data need not be the sole criterion).		

3. Which group would you put Alex into? *OR* Which approach would you use for Alex? Why?

Component	Skill or Concept	Evidence for Presence	Did they get this component?	Extension?
Data use	Uses subscale and item data	Discusses comprehension performance, sight reading (fluency), and vocabulary scores separately to decide what to emphasize with each group. Indicates a desire to look more closely at Alex's performance on different items or		

		standards to support a diagnosis of his needs.		
Question posing	Appreciates value of multiple measures	Discusses comprehension scores on both the state test and the classroom test. Indicates a desire to obtain additional information, such as a reading specialist's assessment of Alex or notes from his second-grade teacher.		
Data use	Understands concept of differentiating instruction based on data	Discusses an individualized instructional plan for Alex based on his assessment results.		

4. Which group would you put Daniel into? What is your reason for that decision?

Component	Skill or Concept	Evidence for Presence	Did they get this component?	Extension?
Data use	Uses subscale and item data	Notes the discrepancy between Daniel's high scores on the		

		state test and his below-average scores on the in-class tests of reading comprehension and sight reading. Indicates desire to have more detailed information about Daniel's performance on particular test items or standards. Discusses issues around possible differences in the content covered by statewide and in-class tests.		
Data interpretation	Understands the concept of measurement error and variability	Notes that Daniel's score on the in-class test of reading comprehension could be an aberrant result.		
Question posing	Appreciates value of multiple measures	Suggests getting another assessment of Daniel's reading comprehension and sight reading, either formally or through in-class observation. Mentions the potential to make a group assignment		

		but keep the groups fluid and keep assessing and regrouping kids.		

5. What group would you put Susan into? Why?

Component	**Skill or Concept**	**Evidence for Presence**	**Did they get this component?**	**Extension?**
Data use	Uses subscale and item data	Notes the discrepancy between Susan's relatively low scores on the state test last spring and her strong performance on the in-class tests this fall. Indicates desire to have more detailed information about Susan's performance on particular test items or standards. Discusses issues around possible differences in the content covered by statewide and in-class tests.		
Data interpretation	Understands the concept of measurement error and variability	Notes that Susan's state test performance last fall might underrepresent her skills.		

Question Posing	Appreciates value of multiple measures	Suggests getting additional information about Susan's reading ability (e.g., report card, second-grade end of semester reading test). Mentions the potential to make a group assignment but keep the groups fluid and keep assessing and regrouping kids.		

Appendix H: Data Based Decision Making - Social Validity and Confidence Tool

On a scale of 1-5 (**where 1 is strongly disagree and 5 is strongly agree**) please rate yourself.

1. Data based decision making is valuable for making instructional decisions.

   1 2 3 4 5

2. I am able to find the right data to use (data location) (e.g., I can find relevant data in a complex table or graph).

   1 2 3 4 5

3. I am able to figure out what the data say about student performance (data comprehension) (e.g., From reviewing the data, I understand how the student performed).

   1 2 3 4 5

4. I am able to make meaning from the data (data interpretation) (e.g., I understand the impact of extreme scores on the mean).

   1 2 3 4 5

5. I am able to apply the data to plan instruction (data use) (e.g., I am able to differentiate instruction based on data).

   1 2 3 4 5

6. I am able to figure out questions that will generate useful data (question posing) (e.g., I can align questions with purpose and data).

1	2	3	4	5

Appendix I: Demographic Survey

Participant ID # ___________

Demographic Information

Age: __________

Gender: _____ Female _____ Male

Race (please check all that apply):
_____ American Indian or Alaska Native
_____ Asian
_____ Black or African American
_____ Native Hawaiian or Other Pacific Islander
_____ White

Appendix J: Consent Form

CONSENT FORM

Effects of a Collaborative Intervention on the Quality

of Preservice Teachers' Data Based Decision Making

You are invited to be in a research study of preservice teachers' quality of data based decision making. You were selected as a possible participant because of your enrollment in EPSY 5616. We ask that you read this form and ask any questions you may have before agreeing to be in the study.

This study is being conducted by: Jennifer Wilson, Educational Psychology.

Background Information:

The purpose of this study is: to investigate the effects of a collaborative data-based decision making intervention on the quality of preservice teachers' data based decision-making.

Procedures:

In class you will be doing the following:

- Completing a pre-and posttest on data based decision making (approximately 20 minutes each)
- Completing a survey on the social validity of data based decision making and your confidence levels to make these decisions (approximately 5 minutes)
- Participating in a collaborative work group. The work group will be audio taped and occur during class time.

If you agree to be in the study you are agreeing the researcher can analyze the data from the pre- and posttest and survey.

Risks and Benefits of being in the Study:

The study has no known risks or benefits.

Compensation:

You will receive your daily class participation points. In addition, you will receive a $5 gift card to the Purple Onion for your participation. At the completion of the study you will receive the gift card if you have participated in all areas (pretest, posttest, survey, intervention group).

Confidentiality:

The records of this study will be kept private. In any sort of report we might publish, we will not include any information that will make it possible to identify a subject. Research records will be stored securely and only researchers will have access to the records. Study data will be encrypted according to current University policy for protection of confidentiality. The tape recordings will be used to ensure the mentors implemented the study as requested. They will then be erased.

Voluntary Nature of the Study:

Participation in this study is voluntary. Your decision whether or not to participate will not affect your current or future relations with the University of Minnesota. If you decide to participate, you are free to not answer any question or withdraw at any time without affecting those relationships.

Contacts and Questions:

The researcher conducting this study is Jennifer Wilson. You may ask any questions you have now. If you have questions later, **you are encouraged** to contact them at wils1313@umn.edu. You may also contact Jennifer's advisor, Dr. Susan Hupp, at shupp@umn.edu

If you have any questions or concerns regarding this study and would like to talk to someone other than the researcher(s), **you are encouraged** to contact the Research Subjects' Advocate Line, D528 Mayo, 420 Delaware St. Southeast, Minneapolis, Minnesota 55455; (612) 625-1650.

You will be given a copy of this information to keep for your records.

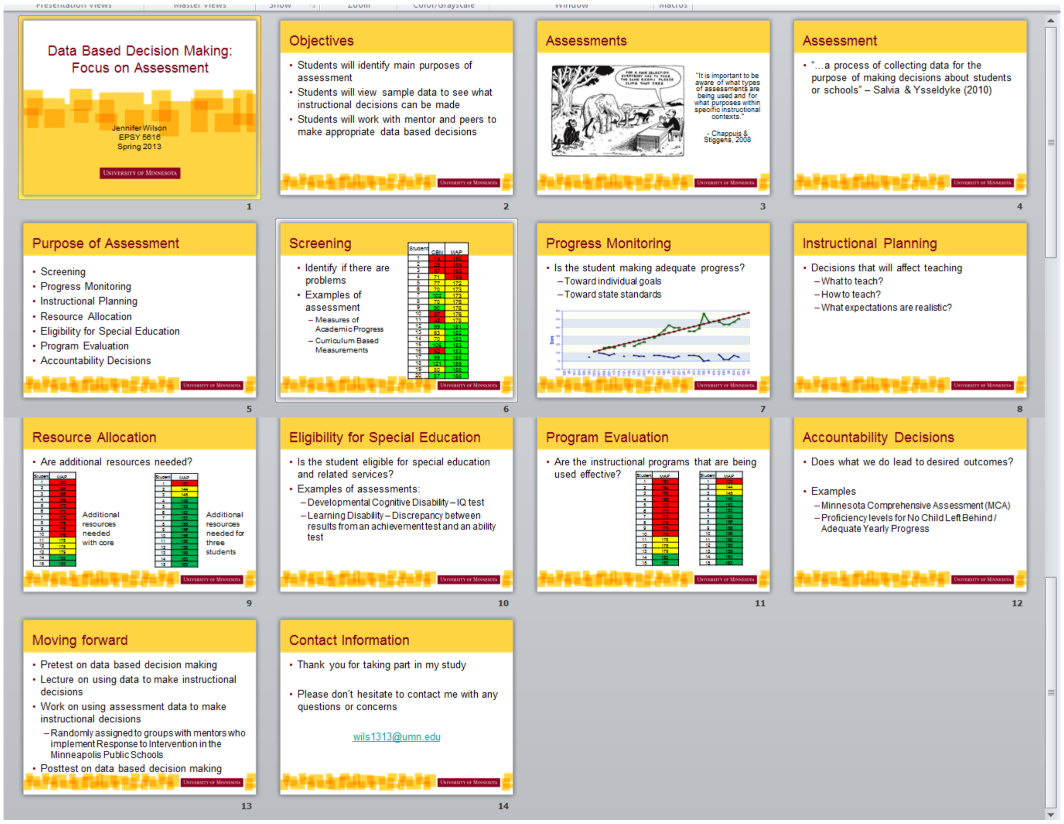
Statement of Consent:

I have read the above information. I have asked questions and have received answers. I consent to participate in the study.

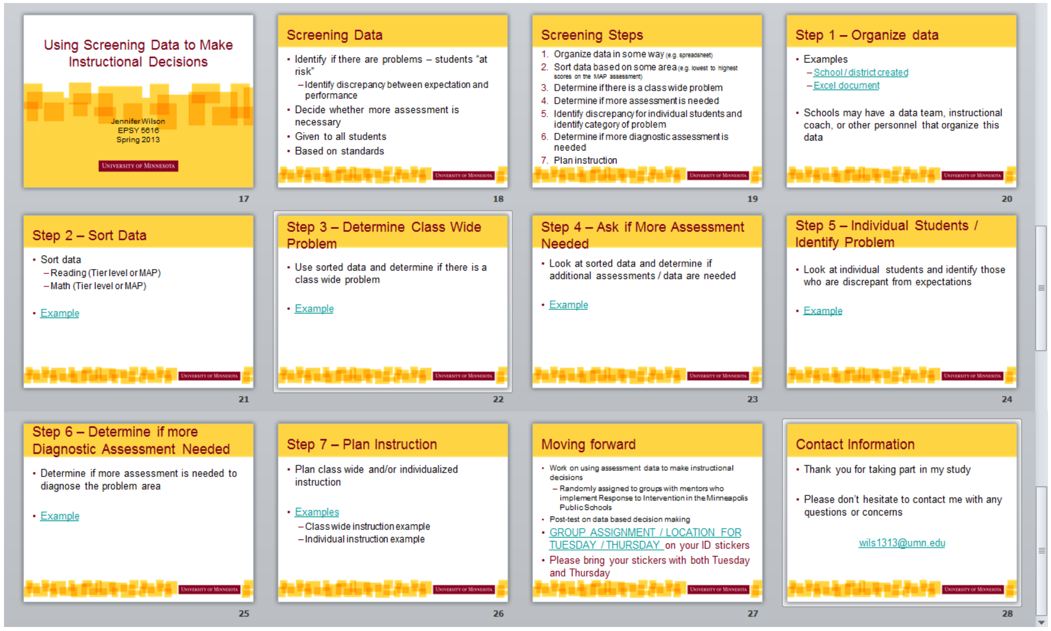
Signature:____ Date: ______________

Signature of Investigator:___________________________ Date: ______________

Appendix K: Data Based Decision Making Power Point Presentation



Appendix L: Steps in Making Data-Based Decisions Power Presentation



Appendix M: Adapted Case Study Approach

Intervention - Semi-Scripted Collaborative Work Time

Thank the participants for being in the study and introduce yourself briefly! Ask them to introduce themselves also (note which participants are in your group and remember to record their participation levels).

Tell the participants to read the scenario along with you and to take about 30 seconds to look over the data (Participants will have the scenario, data, and questions).

SCENARIO:

You are a third grade teacher and it is the third week of school. You are planning your instruction for the next six weeks and want to use the data below to help drive your instruction. As shown below, you have scores from the state reading test given last spring and from two assessments you've had your students take during the first two weeks of school (which they also took last spring).

- CBM stands for Curriculum Based Measurement and is an oral reading fluency measure; students are tested individually by the classroom teacher or support staff.
- F & P stands for Fountas and Pinnell and is a diagnostic assessment that the teacher administers individually to students; it allows the teacher to determine areas of strength and need.
- MAP stands for Measures of Academic Progress and is a nationally standardized computer reading assessment that provides strand data (e.g., word recognition, narrative comprehension). Many school districts in MN give this assessment.

GR	FIRST NAME	Gender	Race	CBM Spring 2011	CBM Fall 2012	F&P Spring 2011	F&P Fall 20 12	MAP-Spring 2012
3	ANTONIO	M	6	28	60	I	J	165
3	ARRIELLA	F	6	109	139	Q	S	205
3	BRANDI	F	3	62	87	M	O	193
3	BRANDON	M	6	70	91	K	P	205
3	CHELSEA	F	4	29	67	I	L	192
3	CHLOE	F	6	88	99	L	M	194
3	CRYSTAL	F	4	63	87	J	L	194
3	DANTE	M	3	32	47	L	N	200
3	JACKSON	M	3	69	90	M	O	204
3	KAYLA	F	6		126		P	203
3	MITCH	M	6	85	106	P	R	201
3	NATHAN	M	3	21	39	I	J	173
3	PHOEBE	F	6		105		O	189
3	RYAN	M	4	38	45	I	M	166
3	SYDNEY	F	3	70	91	M	N	175

Race:
1= American Indian or Alaska Native
2= Asian
3= Black or African American
4= Hispanic/Latino
5= Native Hawaiian or Pacific Islander
6= White

Third grade fall expectations:

	Tier 1	Tier 2	Tier 3
F & P	N or >	M	L or <
CBM	89 or >	72-88	71 or <
MAP	192 or >	184-191	183 or <

Time to START this problem – no later than 2:40

1. PROMPT: Listen to the statement below and say whether you agree or disagree and the reasons why.

The class average for the MAP is 191. At least half of the students are scoring above the class average.

Response Analysis

Component	Skill or Concept	Evidence for Presence	Additional Prompts	Did they get this component?
Data location	Finds relevant data in a complex table or graph	Examines MAP scores and counts the number of students that scored above 191.	a. How could you find out if half of the students score above the class average?	
Data comprehension	Moves fluently between different representations of data	Agrees. Ten students are scoring above the class average.	a. If they say a number other than 10 or disagree say Maybe you miscounted, could you try that again? b. If there are 15 students in the class and 10 scored above the class average, is the statement true?	

Time to START this problem – no later than 2:45

2. PROMPT: What data would you look at as you're planning your instruction?

Response Analysis

Component	Skill or Concept	Evidence for Presence	Additional Prompts	Did they get this component?

Question posing	Understands value of multiple measures	Notes the advantage of looking at more than one assessment. Statewide tests (MAP) usually have stronger technical quality (reliability) but may not match the local curriculum and may not include content above Grade 2. Also, the classroom tests were given more recently.	a. Why would it be important to look at both the statewide tests and the assessments done by the classroom teacher? b. What are advantages of the statewide tests? c. What are advantages of the classroom tests? d. Did you know tests such as the MAP have been given to hundreds of children and that it has stronger technical qualities such as reliability? e. When were the tests given? Why may that matter?	

Time to START this problem – no later than 2:55

3. PROMPT: What, if anything, do these data tell you about how you might want to differentiate instruction for different students in your class? How would you group students and how would your instruction vary for the different groups?

RESPONSE ANALYSIS

Component	**Skill or Concept**	**Evidence for Presence**	**Additional Prompts**	**Did they get this component?**
Data use	Uses subscale and item data	Discusses looking at CBM, F&P, and MAP scores separately to decide what to emphasize with each group and within each group, what each student is struggling with, and how to accommodate their needs.	a. Why would it be important to look at the CBM, F & P, and MAP scores separately? b. What can the CBM data tell you? c. What can the F & P data tell you? d. What can the MAP data tell you?	
Data use	Understands concept of differentiating instruction based on data	Articulates a coherent, data-informed rule for grouping (i.e., includes test data but data need not be the sole criterion).	a. How may you want to group students differently based on their data? b. Would you put a student who is tier 1 (has all ‘green’ scores) in the same group as someone who is struggling (‘red’ scores)? Why or why not?	

Data use	Understands concept of differentiating instruction based on data	Articulates different content or pedagogy for the groups or for individual students consistent with their score profiles (e.g., assign different books or provide different amounts of direct teaching or guided reading to different groups).	a. What may you do differently for students who are in the different groups? How would your instruction change? b. Can you tell me about the amount of time you may spend with the students who are struggling readers compared to those who are proficient? c. What materials may you use that are different for the students who are struggling readers compared to those who are proficient?	

Time to START this problem – no later than 3:05

4. PROMPT: Which group would you put Antonio into? *OR* Which approach would you use for
   Antonio? Why?

Response Analysis

Component	Skill or Concept	Evidence for Presence	Additional Prompts	Did they get this component?
Data use	Uses subscale and item data	Discusses looking at CBM, F&P, and MAP scores separately to	a. Why would it be important to look at the CBM, F & P, and MAP	

		decide what to emphasize with each group. Indicates a desire to look more closely at Antonio's performance on different items or standards to support a diagnosis of his needs.	scores separately? b. How would looking at the different strands of the MAP or areas of the F&P help determine what Antonio's needs may be?	
Question posing	Appreciates value of multiple measures	Discusses scores on both the state test (MAP) and the classroom tests (CBM and F&P). Indicates a desire to obtain additional information, such as a reading specialist's assessment of Antonio or notes from his second-grade teacher.	a. Look at the scores on the MAP, CBM, and F&P – how are they different? b. Is there additional information you'd like to know about Antonio's reading ability? c. What could you do to get additional information about Antonio's reading ability?	
Data use	Understands concept of differentiating instruction based on data	Discusses an individualized instructional plan for Antonio based on his assessment results.	a. What type of instruction might be best for Antonio? b. Could Antonio benefit from an individualized instructional plan? If yes, why?	

Time to START this problem – no later than 3:15

5. PROMPT: Which group would you put Dante into? What is your reason for that decision?

RESPONSE ANALYSIS

Component	Skill or Concept	Evidence for Presence	Additional Prompts	Did they get this component?
Data use	Uses subscale and item data	Notes the discrepancy between Dante's high scores on the state test (MAP) and his below-average scores on the in-class tests (CBM and F &P) Indicates desire to have more detailed information about Dante's performance on particular test items or standards (from the MAP). Discusses issues around possible differences in the content covered by statewide and in-class tests.	a. How are Dante's scores different or alike across the different assessments? b. What information might be helpful to know from the MAP? c. How are statewide and class tests different?	
Data interpretation	Understands the concept of measurement error and variability	Notes that Dante's score on the in-class CBM test of reading could be an aberrant result.	a. Which of the test scores does not seem to fit? b. Why may his CBM scores be so much different than the other assessment results?	

Question posing	Appreciates value of multiple measures	Suggests getting another assessment of Dante's reading, either formally or through in-class observation. Mentions the potential to make a group assignment but keep the groups fluid and keep assessing and regrouping kids.	a. Is there additional information you'd like to know about Dante's reading ability? b. What could you do to get additional information about Dante's reading ability? c. What type of group would you put Dante into? d. Would the groups stay the same until the end of the semester?	

Time to START this problem – no later than 3:25

6. PROMPT: What group would you put Sydney into? Why?

RESPONSE ANALYSIS

Component	Skill or Concept	Evidence for Presence	Additional Prompts	Did they get this component?
Data use	Uses subscale and item data	Notes the discrepancy between Sydney's relatively low scores on the state test (MAP) and her strong performance on the in-class tests. Indicates desire to have more detailed	a. How are Sydney's scores different or alike across the different assessments? b. What information might be helpful to know from the MAP?	

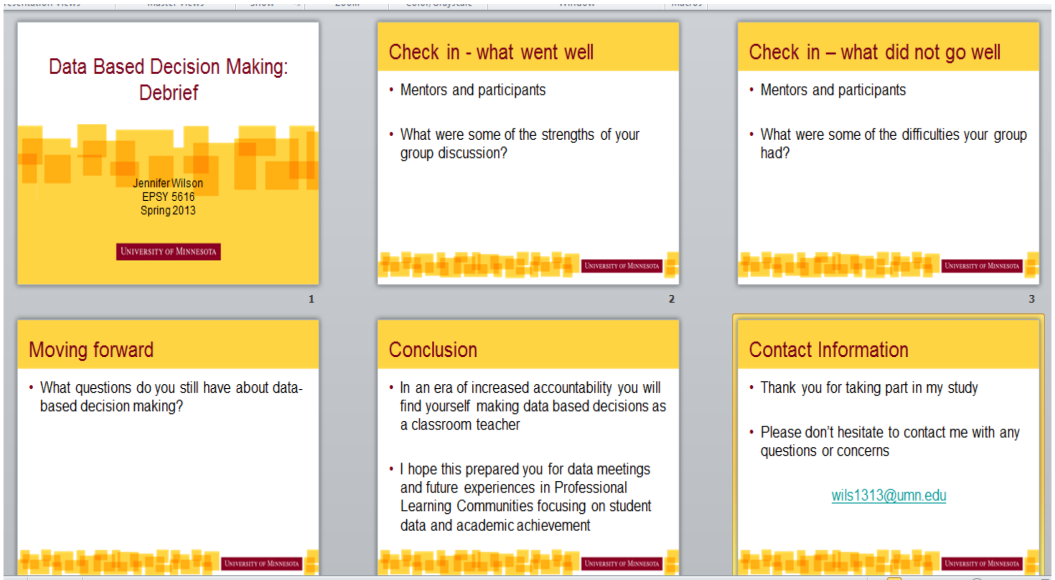
		information about Sydney's performance on particular test items or standards. Discusses issues around possible differences in the content covered by statewide and in-class tests.	c. How are statewide and class tests different? What is the difference in content covered?	
Data interpretation	Understands the concept of measurement error and variability	Notes that Sydney's state test performance might underrepresent her skills.	a. Why may Sydney's MAP score be so different? b. Could the MAP assessment under or over represent her skills?	
Question Posing	Appreciates value of multiple measures	Suggests getting additional information about Sydney's reading ability (e.g., report card, second-grade end of semester reading test). Mentions the potential to make a group assignment but keep the groups fluid and keep assessing and regrouping kids.	a. Is there additional information you'd like to know about Sydney's reading ability? b. What could you do to get additional information about Sydney's reading ability? c. What type of group would you put Sydney into? d. Would the groups stay the same until the end of the semester?	

Appendix N: Participation Recording Chart

To ensure that all participants are active members of the collaborative work group time please document (i.e., tally) how many times they respond to your question prompts. If it seems as though some members are not contributing, please ask them specifically to respond to some of the prompts. If it seems as though all participants are responding (i.e., you've documented that they've all supplied answers at least 5 times) you may discontinue using this form.

Participant Name	**Responses**

Appendix O: Debrief



Appendix P: Procedural fidelity of implementation of Lecture on Using Screening Data to Make Instructional Decisions

Lecturer ___________________________ Rater ___________________________

Step	Yes	No	N/A	Comments
1. Lecturer explains what screening data is	Y	N		
2. Lecturer briefly goes through all of the steps in using screening data to make instructional decisions	Y	N		
3. Step 1: Lecturer tells participants how to organize data	Y	N		
4. Lecturer shows two examples of how to organize data	Y	N		
5. Step 2: Lecturer tells participants how to sort the data	Y	N		
6. Lecturer shows participants how to sort data	Y	N		
7. Step 3: Lecturer tells participants how to determine if there is a class wide problem	Y	N		
8. Lecturer shows an example and non-example of class wide problems	Y	N		
9. Step 4: Lecturer tells participants how to determine if more assessment is needed	Y	N		
10. Lecturer shows an example and provides types of additional assessment that may be needed	Y	N		
11. Step 5: Lecturer tells participants how to determine if there is an individual student problem / discrepancy	Y	N		
12. Lecturer shows an example of individual students	Y	N		
13. Step 6: Lecturer tells participants how to determine if more diagnostic assessment is needed	Y	N		
14. Lecturer shows an example of a student who would need more diagnoistic assessment and gives examples	Y	N		
15. Step 7: Lecture tells participants how the data can help inform their instruction	Y	N		

16. Lecturer provides a class wide and individual instruction example	Y	N		
17. Lecturer uses a didactic presentation style throughout presentation	Y	N		
18. Lecturer allows for questions but does not pose questions to participants	Y	N		

______ of ______ steps were completed as a Y = ______ %

Do not count N/As in the total number of steps above

Comments:

Appendix Q: Fidelity Checklist

Data Based Decision Making
Intervention - Fidelity Checklist

Mentor ___________________________ Rater __________________________

Step	Yes	No	N/A	Comments
1. Mentor introduces herself and participants introduce themselves	Y	N		
2. Mentor tells participants to read the scenario along with her and to take about 30 seconds to look over the data	Y	N		
3. Mentor reads the entire scenario You are a third grade teacher and it is the third week of school. You are planning your instruction for the next six weeks and want to use the data below to help drive your instruction. As shown below, you have scores from the state reading test given last spring and from two assessments you've had your students take during the first two weeks of school (which they also took last spring). • CBM stands for Curriculum Based Measurement and is an oral reading fluency measure; students are tested individually by the classroom teacher or support staff. • F & P stands for Fountas and Pinnell and is a diagnostic assessment that the teacher administers individually to students; it allows the teacher to determine areas of strength and need. • MAP stands for Measures of Academic Progress and is a nationally standardized computer reading assessment that provides strand data (e.g., word recognition, narrative comprehension). Many school districts in MN give this assessment.	Y	N		
Question 1				
4. Mentor says: Listen to the statement below and say whether you agree or disagree and the reasons why. The class average for the MAP is 191. At least half of the students are scoring above the class average.	Y	N		
5. Mentor pauses for participants to do the task	Y	N		
6. Participants share their answer and mentor responds by asking additional prompting questions if the answer was not accurate.	Y	N		
7. Mentor does not proceed to next	Y	N		

question until the participants indicate that the answer is: Agrees – ten students are scoring above the class average				
Question 2				
8. Mentor says: What data would you look at as you're planning for instruction?	Y	N		
9. Mentor pauses for participants to answer the question	Y	N		
10. Participants share their answer and mentor responds by asking additional prompting questions if the answer was not accurate.	Y	N		
11. Mentor does not proceed to next question until the participants indicate the following answers: Notes the advantage of looking at more than one assessment. Statewide tests (MAP) usually have stronger technical quality (reliability) but may not match the local curriculum and may not include content above Grade 2. Also, the classroom tests were given more recently.	Y	N		
Question 3				
12. Mentor says: What, if anything, do these data tell you about how you might want to differentiate instruction for different students in your class? How would you group students and how would your instruction vary for the different groups?	Y	N		
13. Mentor pauses for participants to answer the question	Y	N		
14. Participants share their answer and mentor responds by asking additional prompting questions if the answer was not accurate.	Y	N		
15. Mentor does not proceed to next question until the participants indicate the following answers: Discusses looking at CBM, F&P, and MAP scores separately to decide what to emphasize with each group and within each group, what each student is struggling with, and how to accommodate their needs.	Y	N		

Articulates a coherent, data-informed rule for grouping (i.e., includes test data but data need not be the sole criterion). Articulates different content or pedagogy for the groups or for individual students consistent with their score profiles (e.g., assign different books or provide different amounts of direct teaching or guided reading to different groups).				
Question 4				
16. Mentor says: Which group would you put Antonio into? OR Which approach would you use for Antonio? Why?	Y	N		
17. Mentor pauses for participants to answer the question	Y	N		
18. Participants share their answer and mentor responds by asking additional prompting questions if the answer was not accurate.	Y	N		
19. Mentor does not proceed to next question until the participants indicate the following answers: Discusses looking at CBM, F&P, and MAP scores separately to decide what to emphasize with each group. Indicates a desire to look more closely at Antonio's performance on different items or standards to support a diagnosis of his needs. Discusses scores on both the state test (MAP) and the classroom tests (CBM and F&P). Indicates a desire to obtain additional information, such as a reading specialist's assessment of Antonio or notes from his second-grade teacher. Discusses an individualized instructional plan for Antonio based on his assessment results.	Y	N		
Question 5				
20. Mentor says: Which group would you put Dante into? What is your reason for that decision?	Y	N		
21. Mentor pauses for participants to answer the question	Y	N		
22. Participants share their answer and mentor responds by asking	Y	N		

additional prompting questions if the answer was not accurate.				
23. Mentor does not proceed to next question until the participants indicate the following answers: Notes the discrepancy between Dante's high scores on the state test (MAP) and his below-average scores on the in-class tests (CBM and F &P) Indicates desire to have more detailed information about Dante's performance on particular test items or standards (from the MAP). Discusses issues around possible differences in the content covered by statewide and in-class tests. Notes that Dante's score on the in-class CBM test of reading could be an aberrant result. Suggests getting another assessment of Dante's reading, either formally or through in-class observation. Mentions the potential to make a group assignment but keep the groups fluid and keep assessing and regrouping kids.	Y	N		
Question 6				
24. Mentor says: Which group would you put Sydney into? What is your reason for that decision?	Y	N		
25. Mentor pauses for participants to answer the question	Y	N		
26. Participants share their answer and mentor responds by asking additional prompting questions if the answer was not accurate.	Y	N		
27. Mentor does not finish the intervention until the participants indicate the following answers: Notes the discrepancy between Sydney's relatively low scores on the state test (MAP) and her strong performance on the in-class tests. Indicates desire to have more detailed information about Sydney's performance on particular test items or standards. Discusses issues around possible differences in the content covered by statewide and in-class tests. Notes that Sydney's state test	Y	N		

performance might underrepresent her skills. Suggests getting additional information about Sydney's reading ability (e.g., report card, second-grade end of semester reading test). Mentions the potential to make a group assignment but keep the groups fluid and keep assessing and regrouping kids.				
Other				
28. Mentor makes sure throughout the intervention time that all participants are engaged by answering questions	Y	N		

______ of ______ steps were completed as a Y = ______ %

Notes:

- Do not count N/As in the total number of steps above

Comments: