USING KINDERGARTEN ENTRY DATA TO GUIDE TEACHERS' SKILLS, BELIEFS, SELF-EFFICACY, AND PRACTICES THROUGH INSTRUCTIONALLY FOCUSED, DATA-DRIVEN CONSULTATION

A Dissertation

Presented to

The Faculty of the Curry School of Education

University of Virginia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Sadie L. Hasbrouck, M.Ed.

May 2016

Department of Human Services
Curry School of Education
University of Virginia
Charlottesville, Virginia

APPROVAL OF THE DISSERTATION

This dissertation, "Using Kindergarten Entry Data to Guide Teachers' Skills, Beliefs, Self-Efficacy, and Practices through Instructionally Focused, Data-Driven Consultation," has been approved by the Graduate Faculty of the Curry School of Education in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

____
Dr. Jason T. Downer (Chair)

____
Dr. Amanda P. Williford

____
Dr. Marcia Invernizzi

____
Dr. Daphna Bassok

______________________Date

.

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the guidance and support of my research advisor, Jason Downer. Over the past four years, you have facilitated my development not only as a researcher and clinician, but also as a better person. I was beyond fortunate to have an advisor that was willing to mentor me from a recent college graduate who liked working with young children to a scholar in early childhood development. Your continuous patience and thoughtful feedback was often undeserved but greatly appreciated. Thank you.

Thank you to Amanda Williford for serving as my unofficial advisor throughout graduate school. You were incredibly generous to allow me to design an intervention for VKRP. You have taught me how to get to the point succinctly and be clear in my expectations of myself and others.

Thank you to Marcia Invernizzi and Daphna Bassok for being a part of my dissertation committee. You both provided expertise and thoughtful recommendations that were unparalleled. I would also like to thank my CASTL colleagues for providing feedback at the various stages of the intervention.

Thank you to the National Center on Quality Teaching and Learning and the Virginia Department of Education for funding my time at the Curry School of Education. I am beyond grateful for this support.

Thank you to Jaclyn Russo for serving as a consultant for the intervention. Not only did you support me as a colleague, but also as a friend.

Thank you to the entire Virginia Kindergarten Readiness Program, who provided feedback and support throughout the entire intervention. To Wanda and Genna, for your incredible organization and never-ending help coordinating our project through thick and thin. And to the VKRP teachers, for your patience and willingness to participate.

Thank you to Pilar for your patience in helping me with Mplus. I would also like to thank my lab-mates, Lauren, Catherine, Krishtine, Sam, Shannon, Michelle, and Cathy for helping me through the many stages of this work.

Lastly, I would like to thank my parents, Laurie and Michael Keilholz, and my grandparents, Nancy and Ronald Hasbrouck, for their continued emotional and financial support throughout my education. To my mom, who taught me to be inquisitive and strong. To Mike, my father, who showed me how to be dedicated to whom and what you love. And to my grandparents, who have done so much for me since day 1 that I will forever be indebted. Thank you all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER I

Statement of the problem

Assessments of students' academic achievement have become a staple in state- and country-wide initiatives of educational accountability given their reliance on testing data for assessing students', teachers', schools', and districts' performance (Cooper, 2009; Council of Chief State School Officers [CCSSO], 2000a). Since the 1980's, forty-eight states and several countries have adopted standardized testing mandates that utilize scores as a measure of progress (Brown, 2004a; Choi, 1999; CCSSO, 2000a). This emphasis on assessments, often coined the standards-based movement (SBM; Gullo & Hughes, 2011; Schmoker & Marzano, 1999), emerged from decades of research demonstrating student achievement across content areas was lower than that of similarly developed nations (National Commission on Excellence in Education, 1983; No Child Left Behind Act, 2001). To remedy this, political and educational leaders promoted assessment data as a measure of students', teachers', and schools' achievement by offering reinforcement for higher scores. This surge in the prevalence of testing has led educational institutions to include assessment-related skills as a principle competency for teachers; for example, the American Federation of Teachers (1990) detailed the importance of teachers possessing competency in assessment administration and interpretation to comply with the new era of accountability.

Although the initial rise in educational testing originated from the use of these assessments as outcome measures of learning, more recently the emphasis has shifted to the benefits of using assessment data to inform instructional practice (Ball & Gettinger, 2009; Stecker, Fuchs, & Fuchs, 2005; Strand & Cerna, 2010). This idea is often referred to as using assessment *for* learning, in that assessments can be used to determine where students are at in their learning so that instruction can be tailored to their needs (Elwood & Klenowski, 2002; Firestone, Mayrowetz, & Fairman, 1998). In comparison to assessment *of* learning, the more traditional method of accountability where students' scores are the outcomes, using assessment for learning emphasizes a process by which teachers alter their practices based on data. When implemented in this way, assessments are a valuable tool for teachers to analyze the real-time achievement of their students and to inform their teaching practices based on students' current skill levels.

Using assessment to understand students' baseline skills so that teachers can inform their instruction is particularly important at kindergarten entry as this grade sees the most varied incoming abilities (Heaviside, 1993; Snow, 2011). Children's knowledge and skills at kindergarten entry coalesce into the construct of "readiness" (Meisels, 1996, 1999): how ready is the child to participate in formalized schooling? Readiness gained attention at a national level when the first national education goal (National Education Goals Panel, 1991) was released, emphasizing that "all children will start school ready to learn" (pg. 1). This goal and emphasis on readiness more generally resulted from empirical evidence demonstrating that abilities at kindergarten entry predict both short- and long-term achievement (Badian, 1988; Goldstein, Eastwood, & Behuniak, 2014; Meisels, 1999). For instance, proficiency of cognitive, academic, and social skills at the beginning of the school year are negatively related to kindergarten retention (Goldstein et al., 2014) and positively related to reading abilities at the end of third and eighth grade

(Badian, 1988; Gibson & Levin, 1975; Neuman & Roskos, 2005; Snow, Burns & Griffin, 1998). Furthermore, this pattern holds across cultural and socioeconomic backgrounds (De Feyter & Winsler, 2009; Quirk, Grimm, Nylund-Gibson, Furlong, & Swami, 2015), suggesting that kindergarten entry data have the potential to inform types of support necessary to ensure children's success across the increasingly diverse landscape of public schools in the United States.

As a result of the importance of children's abilities at kindergarten entry, national educational organizations promote the importance of measuring readiness. For example, in 2011 nine states received Race to the Top-Early Learning Challenge grants (RTT-ELC; U.S. Department of Education & U.S. Department of Health and Human Services [U.S. DOE & U.S. DHHS], 2014), which aim to augment the quality of early education. As a part of these grants, states were required to utilize kindergarten entry assessments (KEAs) to assess the cognitive, academic, and social status of children entering kindergarten. It is not enough to simply measure these abilities; the data must then be used to inform teachers' practices, if children are to benefit. Literature supports the implementation of assessment for learning in early grades; when used to inform instructional practices, assessments are useful for helping students achieve at higher rates (Ball & Gettinger, 2009; Stecker et al. 2005; Strand & Cerna, 2010). Consequently, national organizations, such as the National Association for the Education of Young Children (NAEYC), now emphasize using readiness data as imperative for providing instruction responsive to children's unique needs (Kagan & Kauerz, 2006; National Association for the Education of Young Children [NAEYC], 2003).

Despite the documented benefits of using assessments for learning, student assessments are rarely implemented in this way. Reviews of teachers' use of assessment across educational settings demonstrate few teachers implement assessments before or during learning to inform instruction (Cizek, Fitzgerald, & Rachor, 1995; Stecker et al., 2005; Wesson, King, & Deno, 1984). So why do assessments for learning not translate into classroom practice? Interviews of teachers reveal logistics of testing are the most salient obstacle; teachers report they do not have enough time or the proper training to implement assessments effectively (Looney, 2005; Wesson et al. 1984). Empirical evaluations of assessments used in the classroom have revealed that although logistics are a challenge, the management of assessments is not the most salient barrier (Stecker et al., 2005). Rather, research implicates teacher-level factors such as their skills in, beliefs around using, and self-efficacy for data in the classroom as the significant obstacles affecting implementation (Brown, 2004a; 2009; Stecker et al., 2005; Tschannen-Moran & McMaster, 2009). Foremost, teachers lack comprehensive knowledge around and skills to interpret data and translate these interpretations to instructional practices (Fuchs et al., 1992; Fuchs, Fuchs, Hamlett & Stecker, 1991; Stecker et al., 2005). A review of assessments used to monitor students' progress during the academic year found that while most teachers collected data accurately, they were unable to make instructional decisions using these data and thus their practice went unaffected (Stecker et al., 2005). When provided with a specific strategy based on the data, however, they were able to change their instructional practices, ultimately improving student learning (Al Otaiba, Conner, Folsom, Greulich, Meadows, & Li, 2011; Stecker et al., 2005). This discrepancy suggests the missing piece is teachers' data-interpretation skills, specifically in selecting relevant instructional strategies based on student data. As such, interventions aiming to improve teachers' practices through use of assessment data should focus on building teachers' capacity to interpret data and select strategies.

In addition to their skills in using data, teachers' beliefs about the use of data in classrooms have the potential to influence their instructional practices (Brown, 2004a; 2009). As is consistent with more general theories describing the connections between beliefs and subsequent behaviors (e.g., Bandura, 1986), teachers implement and use the data from assessments in ways that align with their beliefs (Brown, 2004a; 2009). When teachers do not believe that assessments are useful for informing teaching, they are less likely to make data-driven decisions that impact the implementation of practices. Consequently, an intervention with the goal of improving teachers' practices related to using assessment data should also consider how to affect teachers' beliefs about assessment.

Finally teachers' sense of self-efficacy has the potential to influence their data-use practices (Allinder, 1994; Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998). Research demonstrates the connection between self-efficacy and effective instruction (Allinder, 1994; Tschannen-Moran, et al., 1998) and preliminary work suggests this mechanism operates similar for data-driven practices (Ross, 1992; Tschannen-Moran & McMaster, 2009). As such, self-efficacy may be a crucial component in improving teachers' use of assessment data, and should therefore be a focus of intervention.

Consultee-centered consultation offers an avenue for affecting teachers' skills in, beliefs around, and self-efficacy for assessment data-use with the goal of influencing their data-driven practices in the classroom. This type of consultation allows for mutual problem-solving between the consultant and consultee around a client. In the current, school-based model, the consultant and consultee (teacher) collaborated around clients' (individual students or the classroom) data to determine appropriate instruction based on that data. To do this, principles of two existing consultations models were crossed: data-driven and instructional consultation. Data-driven consultation, a model that focuses on "trying to understand the 'root causes' behind the numbers" (Slavin, Cheung, Holmes, Madden, & Chamberlain, 2011, p. 4), aims to build teachers' skills in interpreting and using data accurately. Data-driven consultation uses five steps, beginning with obtaining, reflecting, and designing an intervention based on the data, then implementing and receiving feedback on the intervention. Through these steps, teachers hone skills in data-analysis, which contributes to alleviating the first barrier of data-interpretation skills. To help teachers link these data to specific instructional practices, principles of a second model, instructional consultation (IC), were included. IC aims to enhance instruction by identifying the problem, designing an instructional plan, implementing the plan, and then evaluating the effects of the instruction on the problem (Rosenfield, 2004; 2013). In this way, IC provides a focus on identifying appropriate input necessary to help teachers select appropriate instruction and evaluate the effects of that instruction. When successful in improving learning, these types of experiences improve teachers' feelings of self-efficacy (Tschannen-Moran et al., 1998; Tschannen-Moran & McMaster, 2009). Furthermore, IC principles have been demonstrated to affect teachers' beliefs about instruction (Rosenfield, 2008; Rosenfield, Gravois, & Silva, 2014), thereby providing an avenue through which to affect teachers' beliefs about data-use practices. By combining these two models, we created a consultation model aiming to build teachers' skills in beliefs around, and self-efficacy for using assessment data to inform their instructional practice: instructionally focused, data-driven consultation (IFDDC).

Through chapter II, I review literature that explores kindergarten entry and how data around children's incoming skills can be used to inform subsequent instruction, how assessment data are utilized more generally to improve instruction, the barriers teachers experience that limit the efficacy of using assessment data to inform their practices, and how consultation, specifically IFDDC, can be implemented to remedy these barriers. This literature coalesces into a theory of change model and subsequent aims and hypotheses of an IFDDC randomized controlled trial study at the end of Chapter II. Chapter III describes the methods used to conduct the evaluation of IFDDC on teachers' skills, beliefs, self-efficacy and practices in a sample of 72 kindergarten teachers as part of the voluntary rollout of a new KEA program in the Commonwealth of Virginia. The results of this evaluation are examined in Chapter IV, followed by interpretation of the results and practical implications discussed in Chapter V.

CHAPTER II

Literature Review

Kindergarten Readiness and Instruction

The kindergarten year is a crucial period in children's academic and social development; the National Association for the Education of Young Children (NAEYC) described kindergarten as the time when "children's critical learning patterns begin to be established" (Kagan & Kauerz, 2006, p. 163). This emphasis on kindergarten originates from longitudinal research demonstrating the effects of these early academic skills on children's trajectories; for example, analyses of the Early Childhood Longitudinal Study-Kindergarten Cohort (ECLS-K, U.S. Department of Education, National Center for Statistics, 2009), which followed nearly 20,000 children from kindergarten through the end of eighth grade, consistently relate skills gained in kindergarten to both short- and long-term academic and social outcomes (Mulligan, McCarroll, Flanagan, & Potter, 2014; Walston, Rathbun, & Germino Hausken, 2008). In combination with other longitudinal studies demonstrating similarly strong predictive models (e.g., Lloyd & Hertzman, 2009; McClelland, Acock, & Morrison, 2006; Pagani, Fitzpatrick, Archambault, & Janosz, 2010), learning during the kindergarten year is now widely accepted as a significant contributor to children's overall development.

Now that the importance of kindergarten has been firmly established, research efforts have shifted to understanding how to help children thrive during this year

(NAEYC, 2003; U.S. Department of Education, 2013). Many factors that influence a child's learning during kindergarten are unchangeable; for instance, age at kindergarten entry and gender both predict learning outcomes (Elder & Lubotsky, 2009; Gullo & Burton, 1992; Huang & Invernizzi, 2012; Matthews, Ponitz, & Morrison, 2009; Stipek & Byler, 2001). Of malleable influences on kindergarten learning, the skills children arrive with at entry relate consistently to how children learn throughout the year (Meisels, 1999; Rock & Stenner, 2005). Consistent with Heckman's (2008) theory that skills beget skills, this indicates that children who enter with more advanced skills will learn more readily because "[a]ll capabilities are built on a foundation of capacities that are developed earlier" (p.289). Kindergarten is often viewed as the foundational year for first through twelfth grades, but in reality, a child's capacity for learning begins forming even earlier. Every child arrives on the first day of kindergarten with a unique profile of strengths and weaknesses, which teachers can use to tailor instruction to a developmentally appropriate level. By individualizing instruction to build upon children's incoming knowledge and skills, teachers can facilitate a successful beginning to the foundation of their education. To understand how to best help teachers understand and capitalize upon incoming abilities for instructional purposes, first kindergarten readiness and its implications for instruction throughout the year must be explored in greater depth.

Kindergarten readiness. The abilities children possess at kindergarten entry are often referred to as readiness skills (Blair, 2002; Meisels, 1999; Rafolth, Buchenauer, Crissman, & Halko, 2004; Scott-Little, & Niemeyer, 2001; U.S. DOE & U.S. DHHS, 2014). In the context of kindergarten, 'readiness' aims to describe a child's competencies for formalized schooling before beginning kindergarten (Meisels, 1999; Pianta, Rimm-

Kaufman, & Cox, 1999; Rock & Stenner, 2005). While operational definitions of this construct vary (e.g., Kagan, 1990; Kohlberg & Mayer, 1972; Meisels, 1999), the main tenants of readiness across literature surround possessing developmentally appropriate pre-academic, social, and approaches to learning skills necessary for active engagement in kindergarten classrooms (Blair, 2002; Meisels, 1996, 1999; Rafolth et al., 2004).

As previously mentioned, kindergarten readiness consistently predicts short- and long-term outcomes in that children who possess higher abilities at kindergarten entry demonstrate better achievement in kindergarten (Goldstein et al., 2014) and through high school (Badian, 1988; Gibson & Levin, 1975; Neuman & Roskos, 2005; Snow, Burns & Griffin, 1998) than their peers with lower initial abilities. Unfortunately, children of lower socioeconomic backgrounds and those of minority racial ethnicities are less likely to be 'ready' than their affluent, white counterparts (Fryer Jr. & Levitt, 2004; Janus & Duku, 2007; Williford, Downer, Hamre, & Pianta, 2014). Research demonstrates that these initial gaps in learning remain present throughout K-12 education (Fryer Jr. & Levitt, 2004; Isaacs, 2012; Johnson, 2002) suggesting that addressing this divide at kindergarten entry may alleviate later racial and ethnic achievement gaps (Fryer & Levitt, 2004; Ramey & Ramey, 2004). As a result of the importance and potential power of readiness, measuring and understanding children's abilities at kindergarten entry has gained attention of national policymakers aiming to improve educational outcomes (NAEYC, 2003; National Education Goals Panel, 1991; North Carolina Ready for School Goal Team, 2000; U.S. DOE & U.S. DHHS, 2014). This attention in part allocated resources with which kindergarten readiness and instruction have been explored in great detail, providing theoretical and empirical support for how to best understand and use readiness data.

Domains of kindergarten readiness. Nuanced evaluations of the operationalization of readiness reveal that the construct is not global; that is, several readiness domains exist and a child may be “ready” in one domain while not in others (Ackerman & Barnett, 2005; Meisels, 1999; Sabol & Pianta, 2012; U.S. DOE & U.S. DHHS, 2014; Williford et al., 2014). The skills necessary for children’s success in kindergarten initially emerged as falling under the content areas of language/literacy and mathematics (U.S. Department of Health and Human Services, 2002; Neuman & Roskos, 2005), but more recently non-academic skills, such as social skills (Raver & Zigler, 1997; Webster-Stratton & Reid, 2004) and approaches to learning (Duncan et al., 2007), have been emphasized as skill areas essential to a child’s scholastic success.

Evaluations of incoming kindergarteners have supported the distinct nature of learning domains; for example, a 2014 evaluation of nearly 2,000 kindergarteners in Virginia revealed that while one-third of children were ‘not ready’ in at least one learning domain, only 1.6 percent lacked adequate skills across literacy, math, social skills, and approaches to learning (Williford et al., 2004). As a result of the differences in children’s abilities across learning domains, national organizations of early childhood development now consider readiness a child’s unique combination of skills across learning areas (Bierman et al., 2008; U.S. House of Representatives and Senate, 2007; U.S. DOE & U.S. DHHS, 2014). Each domain and associated skills expected at kindergarten entry must be understood independently to provide a full picture of the vast potential for variance in children's profile, each of which has implications for teachers' instructional decisions.

Language and literacy readiness. The historical emphasis on young children's academic skills has typically surrounded their language and literacy development, which is likely due to the strong relationship between early language/literacy skills and long-term reading outcomes (Gibson & Levin, 1975; Neuman & Roskos, 2005; Snow et al., 1998). As a result of this ongoing attention, experts in this field have reached a relative consensus on the necessary early skills that lay the foundation for reading (National Reading Panel Report, 2000). Consequently, language and literacy readiness largely focuses on these basic competencies: lexical knowledge or receptive language, phonological awareness, letter knowledge, print conventions, and expressive language (Hall, 1987; Administration of Children and Families, 2015; Invernizzi, Justice, Landrum, & Booker, 2004; Neuman & Roskos, 2005). Receptive language is one of the most apparent early skills children develop, and during the preschool years children learn as many as seven new words per day (Snow et al., 1998). Children's exposure to, and subsequent acquisition and use of, words is essential to their ability to read and write (HSELOF, 2015; Neuman, 2001). Although there is no specific number of words children should know before entering kindergarten, children who possess larger and more complex incoming vocabularies demonstrate better learning outcomes across domains (HSELOF, 2015; Invernizzi et al., 2004; Snow et al., 1998). Lexical knowledge, or "vocabulary breadth and depth" (Verhoeven, van Leeuwe, Irausquin, & Segers, 2016, p. 591), is a precursor to developing phonological awareness, a crucial contributor to reading ability (Bus & van IJzendoorn, 1999; Invernizzi, 2003; Stanovich, Cunningham,

& Cramer, 1984). Phonological awareness represents the ability to discern between units of language (Bear, Invernizzi, Templeton, & Johnson, 2011; National Reading Panel Report, 2000), and at the end of preschool this most often surrounds isolating syllables and sounds in word (e.g., “man” is made up of the sounds “m” and “an”). Understanding of phonemes helps children to develop more complex understanding of words, creating a feedback loop between the development of lexical knowledge and phonological awareness during early childhood (Ehri & Roberts, 2005). With scaffolding, children begin to recognize and associate letters with these sounds and words, building their letter knowledge (Bowles, Pentimonti, Gerde & Montroy, 2013; Ehri, 1979; Gibson & Levin, 1975).

In addition to understanding word structure and growing their vocabularies, kindergarteners are also expected to be able to reproduce words through print conventions and expanded expressive language (Day & Day, 1984; Huang & Invernizzi, 2014; Moats, 2000). These two skills, while distinct, represent the generation of language. To do this, children need to have a base of letter knowledge and an elementary understanding of print conventions, such as reading from left-to-right (Huang & Invernizzi, 2014; Neuman & Roskos, 2005). Simultaneously, children are expanding their expressive language capacity by incorporating their newfound lexical, phonological, and print knowledge into the words they produce to communicate (HSELOF, 2015; Williams, 1997). Together, these interwoven skills contribute to a child’s readiness for language and literacy instruction throughout the kindergarten year.

Mathematics readiness. Early math skills are commonly referenced as a critical area of children’s development due to their association with later academic success

(Duncan et al., 2007; National Council of Teachers of Mathematics [NCTM], 2002; Siegler et al., 2012; Stevenson & Newman, 1986). Recent reviews of pre-kindergarten and state readiness mathematics standards found that early skills were generally divided into numbers/operations, geometry/spatial relations, and algebra/data analysis (NAEYC/ NCTM, 2010; Neuman & Roskos, 2005). Numerical knowledge is one of the first academic skills attained (Kilpatrick et al., 2001), but experts have been careful to discern that the critical skill associated with counting involves an abstract understanding of number relations, not simply reciting numbers as one would do with the alphabet (Bowman, Donovan, & Burns, 2000; Clements, 2001). Counting tends to be a large emphasis of prekindergarten curricula (HSELOF, 2015; Lembke & Foegen, 2009), so many children enter kindergarten with at least a basic foundation of numeracy awareness (Clement, 2001). Geometry skills also tend to develop early as children begin to recognize basic shapes by age four (Clements, 2001). More advanced skills in this area revolve around reasoning with and comparing objects (Clements, Sarama, & DiBiase, 2004), as well as understanding more complex spatial relationships (Rauscher & Zupan, 2000). The last group of skills, algebra and data-use, involve applying the first two mathematics capacities to solve real-world problems (Neuman & Roskos, 2005). For young children, this often surrounds recognizing patterns and using numbers with physical representations to solve problems (Ginsburg, 1989, National Association for the Education of Young Children/National Council of Teacher of Mathematics [NAEYC/NCTM], 2002/2010). In tandem, these three skill areas coalesce into a child's mathematics readiness, allowing him or her to the expected math skills during the kindergarten year.

Social readiness. As previously stated, the influence of social development for a child's kindergarten success has only recently gained political attention (U.S. Department of Education, 2013), but social skills have been consistently proven imperative in a child's ability to thrive in the classroom setting (Meisels, 1999; Raver & Zigler, 1997; Webster-Stratton & Reid, 2004). Interpersonal or social skills represent children's capacity to negotiate social situations with children and adults (Hart, Ladd, & Burleson, 1990; Webster-Stratton & Reid, 2004). At the age when children enter kindergarten, developmental theories attest that they should be able to communicate their basic needs sufficiently and empathize with others (Fabes, Gaertner, Popp, McCartney & Phillips, 2006; Vygotsky, 1978). Specifically, children should be able to maintain secure attachments to their teachers, demonstrate interest in peers, and interact with peers appropriately at kindergarten entry (HSELOF, 2015; McClelland & Morrison, 2003). Age-appropriate social skills are central for playing and working with others, which are common formats of learning in kindergarten. For children entering formalized schooling for the first time, their kindergarten classmates may be the first children they have had to cooperate with consistently (Fabes et al., 2006; Whiting, Edwards, & Edwards, 1992). Children coming from formal preschools, on the other hand, are likely to have experienced systematic interactions with peers and to have begun building the capacity to facilitate these interactions positively (Fabes et al., 2006; Hamre & Pianta, 2007). As such, kindergarten teachers often see significant variation in their students' incoming social skills; one study found that one-fifth of kindergarteners lacked the social capacities deemed necessary to succeed in the classroom at the beginning of the year (Williford et al., 2014). Given the observed discrepancies and the necessity of these skills for learning, many teachers prefer social over academic readiness and emphasize this domain early in the academic year (Ackerman & Barnett, 2005; Lin, Lawrence, & Gorrell, 2003).

Approaches to learning readiness. The final domain necessary for success in kindergarten surrounds how a child engages in, or approaches, learning activities (Bronson, 2006; Duncan et al., 2006). The construct of 'approaches to learning' includes the various non-academic skills necessary for learning, such as attention, persistence, inhibition, and curiosity (Gijbels & Dochy, 2006). These capacities fall under many names depending on the literature referenced, such as self-regulation (Eisenberg, Valiente, & Eggum, 2010; Zimmerman, 1990), effortful control (Blair & Razza, 2007), or work-related skills (Cooper & Farran, 1988; McClelland, Morrison, & Holmes, 2000). Regardless of the name, skills implicated in this domain are those that relate to an individual's ability to engage in classroom activities effectively. For many children, kindergarten is the first setting in which they are asked to approach learning tasks systematically and research demonstrates that children's abilities to attend to and participate in these types of educational formats is essential for their success during the kindergarten year and beyond (Cooper & Farran, 1988; Duncan et al., 2006; Meisels, 1996, 1999). For example, children's ability to inhibit immediate responses and persist when faced with frustration at the beginning of school predicts end of year math and reading achievement (McClelland et al., 2007). In this way, a child's approaches to learning abilities at the beginning of kindergarten lay the foundation necessary for learning throughout schooling.

At kindergarten entry, children are considered adequately prepared for classroom learning if they can stay on task for approximately ten to fifteen minutes (Barkley, 2006;

Murphy-Berman, Rosell, & Wright 1986), filter out minimal distractions (Cooper & Farran, 1988), listen and follow two- to three-step instructions (Alloway & Alloway, 2010; Barkley, 2006), and organize their work materials and efforts with minimal-to-moderate scaffolding (Cooper & Farran, 1988). Unfortunately, many students do not arrive at kindergarten with the necessary approaches to learning skills; evaluations of children's incoming attention, inhibition, and general self-regulation capacities show between 11 and 16 percent of children are rated as below the threshold for readiness (Blair & Razza, 2007; Williford et al., 2014). Children who do not possess the aforementioned self-regulatory capacities require additional scaffolding from adults to engage appropriately in the classroom, meaning teachers are spending time providing these supports as opposed to instruction. As such, teachers often highlight self-regulatory capacities equal or more essential to kindergarten success as compared to incoming academic skills (Ackerman & Barnett, 2005; Lin et al., 2003).

Variation in kindergarten readiness. Readiness covers several domains which operate independently and consequently, there is often variation in a child's incoming profile of skills and knowledge across content areas (U.S. DOE & U.S. DHHS, 2014). Differences in readiness skills are observed between- and within-children; the former relates to diversity in groups of children's readiness, whereas the latter refers to differences in a specific child's readiness by domain. Kindergarten teachers must consider both between- and within-children variation when tailoring instruction to the incoming skills of their classroom's readiness landscape.

Between-child variation. Unlike more advanced grades, children enter kindergarten classrooms with an array of experiences ranging from structured, center-based preschool to family-only exposure (Hamre & Pianta, 2007; Meisels, 1999; U.S. DOE & U.S. DHHS, 2014). For some children, the kindergarten year represents their introduction to schooling, or where they are socialized to learning (Meisels, 1999; Pianta, et al., 1999), whereas others have participated in structured preschool programs which emphasize academic and social development (Hamre & Pianta, 2007; Meisels, 1999). As such, a kindergarten classroom of twenty students may consist of students with vastly different experiences prior to the kindergarten year. Although evaluations of kindergarten readiness have observed one-third of children to lack adequate preparation in at least one learning domain (Boyer, 1991; Williford et al., 2014), these same studies highlight the variation that exists within classrooms across children. That is, some classrooms house children who are all "ready," whereas in others few if any children are ready. A 2014 evaluation of kindergarten classrooms at the beginning of the year across Virginia demonstrated that 27 percent of classrooms had nearly all students ready, 61 percent had between three- and four-fifths of students ready, whereas the remaining 11 percent had very few students who were ready to participate in kindergarten instruction (Williford et al., 2014). The differences in students' preparedness highlights the need for measuring and understanding these skills at kindergarten entry so teachers can adequately prepare and tailor instruction to be consistent with best practices (Middendorf, 2007; NAEYC, 2003; Sarama & Clements, 2006; U.S. DOE & U.S. DHHS, 2014)

Within-child variation. Readiness also varies within-children as skills can be unevenly developed across learning domains (U.S. DOE & U.S. DHHS, 2014; Williford et al., 2014). Attention and associated funding brought forth by the national educational emphasis on readiness, like the RTT-ELC grants, allowed for several large-scale evaluations of children's incoming skills at the beginning of kindergarten. Averages differ across studies but generally fall within the following range of children noted as 'not ready' in respective domains: eight to 14 percent in math, eight to 11 percent in literacy, nine to 20 percent in social skills, and 16 to 17 percent in approaches to learning (Davies, Janus, Duku, & Gaskin, 2015; Fantuzzo, Bulotsky-Shearer, McDermott, McWayne, & Frye, 2007; Maryland Department of Education, 2009; North Carolina Ready for School Goal Team, 2009; Williford et al., 2014; U.S. DOE & U.S. DHHS, 2014). Few children are unprepared across all domains, however; one study found that only 1.6 percent of incoming kindergarteners lacked readiness across math, literacy, social, and self-regulation skills (Williford et al., 2014). This rate is in contrast to kindergarten teachers' perceptions of children, as surveys demonstrate teachers rate nearly twenty percent of children as robustly unprepared (Heaviside, 1993; Rimm-Kaufman, Pianta, Cox, 2000). The discrepancy in children's abilities and teachers' perceptions of these abilities highlights the necessity of measuring readiness domains independently and presenting these results to teachers in an accessible manner so they can tailor their instruction accordingly.

Measuring kindergarten readiness. Given the importance of and variation in kindergarten readiness across learning domains, accurately measuring these skills has emerged as an educational priority (Maxwell & Clifford, 2004; North Carolina Ready for School Goal Team, 2000; Snow, 2011). In 1998 the National Education Goal Panel (NEGP) released a report entitled "*Principles and Recommendations for Early Childhood Assessments*" (Shepard, Kagan, & Wurtz, 1998), which highlighted the necessity of testing for school readiness before or directly after kindergarten entry. At the time of publication, few schools utilized KEAs, but since then the quantity of KEAs has risen dramatically. By 2000, a nationally representative survey orchestrated in part by the National Center for Early Development and Learning (NCEDL) found almost every state was using some type of entry assessment (Saluja, Scott-Little, & Clifford, 2000). By 2014, twenty statewide KEAs were administered as a part of RTT-ELC grants and 23 additional states were developing or adapting KEAs for rollout (U.S. DOE & U.S. DHHS, 2014)

Despite the emerging frequency of readiness measures, quality of these assessments vary dramatically (Costenbader, Rohrer, & DiFonzo, 2000; Niemeyer & Scott-Little, 2001; Saluja et al., 2000; Scott-Little, & Niemeyer, 2001; U.S. DOE & U.S. DHHS, 2014). For example, a review of the 37 RTT-ELC applications received in 2011 demonstrated that only 12 were employing KEAs with established norms for comparison (Hanover Research, 2013). Best practices in school readiness assessment consistently warn against erroneous assessment systems or the misuse of data (Niemeyer & Scott-Little, 2001; Saluja et al., 2000; Scott-Little, & Niemeyer, 2001), further highlighting the need for reliable measures. As such, SERVE, an educational organization funded by the U.S. Department of Educations' Office of Educational Research and Improvement, published a comprehensive review of 39 commercially available readiness assessments (Niemeyer & Scott-Little, 2001). This report, or compendium, specifies broad information about the assessments in a user-friendly manner meant to be accessible to school divisions and educational policy makers (Niemeyer & Scott-Little, 2001). The compendium was revolutionary, because it was the first publication that went beyond highlighting the need for KEAs by providing inclusive, accessible information. Since that time, so many other measures have been published and validated that the National Conference on State Legislature provides yearly reports on the availability and use of school readiness measures (Sterdon & Berger, 2010).

Although the original purpose behind many KEAs was to collect outcome data on early childhood program efficacy (Hanover Research, 2011; Sterdon & Berger, 2010), the use of these data to tailor instruction to children's needs throughout the kindergarten year is becoming more prominent (U.S. DOE & U.S. DHHS, 2014). This increase in data-driven instructional decisions can be partly attributed to the SBM, which emphasizes the use of data in schools, and also to research demonstrating the benefits of using assessment data to help children learn (Ebbeler, Poortman, Schildkamp, & Pieters, 2016; Scott-Little & Niemeyer, 2001; Shepard et al., 1998; Stecker et al., 2005). The theory and research behind this purpose are explored in greater detail beginning on page 39, but first the expectations of and practices in kindergarten instruction are discussed to frame the potential avenue through which KEA data can be used to inform instructional decisions.

Kindergarten instruction: expectations, practices, and barriers to execution. Although teachers at any level are expected to vary and individualize instruction based on students' needs (Howes, 1970; U.S. DHHS, 2002), kindergarten teachers are often faced with the most variable set of student skills at the beginning of the year on which to base this individualization. As such, these teachers typically require support in translating readiness data into classroom practices, creating a prime target for professional development (PD). Before considering how to create a PD model that helps teachers use entry data to inform their instruction, instructional expectations, existing practices, and potential contextual barriers of kindergarten must be explored.

Expectations of kindergarten instruction. The theories surrounding kindergarten instruction consistently highlight the need for learning through experience and play regardless of incoming skills (Bryant, Clifford, & Peisner, 1991; Middendorf, 2007; Pianta et al., 1999). This emphasis on experiential learning for young children originates from theories of learning, such as those described by Piaget (1926) and Vygotsky (1978), which promote developmentally-appropriate instruction. During the kindergarten year, children are at a stage of exploration in which they are using their imagination and intuition to understand the world (Piaget, 1926; Vygotsky, 1978). Age-appropriate instruction for four-through-six-year-olds, therefore, involves learning through doing (Bryant et al., 1991; Piaget, 1926; Pianta, et al., 1999). In other words, children are facilitating their learning through experiences rather than through listening alone. The latter describes teacher-centered practices, which typically involve lecture and subsequent question/answer formats common to more advanced grades (Brown, 2003; Cuban, 1983; Schuh, 2004). In contrast, many kindergarten classrooms provide child-centered, hands-on learning experiences that include collaborative activities with an emphasis on connecting learning to prior knowledge (Bryant et al., 1991; Pyle & DeLuca, 2013). Research supports the use of these child- or learner-centered practices, as experiential learning has been shown to produce larger gains in kindergarteners' learning than teacher-centric instruction (Brown, 2003; Heroman & Copple, 2006, Kagan & Kauerz, 2006). Furthermore, recent investigations highlight the importance of proximal processes, or the interactions that occur between children and settings over time (Bronfenbrenner & Morris, 1998), as most crucial in children's early learning (Hamre & Pianta, 2007; Pianta, 2006). Effective proximal processes for learning at this age include emotionally and instructively supportive interactions with teachers and peers, which occur most often through planned, engaging activities which beckon high-order thinking (Bredekamp & Coople, 1997; Pianta, La Paro & Hamre, 2004). The type of child-centered activities that facilitate effective interactions differ by content area, indicating teachers must be instructionally flexible to achieve a developmentally-appropriate classroom based on incoming needs across language/literacy, math, and social/approaches to learning domains (Bredekamp & Coople, 1997; Copple, Bredekamp, Koralek, & Charner, 2014; Gullo & Hughes, 2011).

Expectations of language and literacy instruction. As was stated earlier, language and literacy concepts acquired during kindergarten lay the foundation for future reading skills (National Reading Panel Report, 2000; Snow et al., 1998). Early instruction in this area, therefore, should focus on how to best help children develop foundational skills, including lexical knowledge, phonological awareness, letter knowledge, and print conventions. Best practices in language and phonological awareness often involve repeating and extending children's language (Invernizzi, 2003; Strickland, 2006), particularly that occurring during play (Bryant et al., 1991; Rog, 2011). Teachers can help develop students' lexical base by introducing and using new vocabulary across content areas (Eller, Pappas & Brown, 1988; Strickland, 2006). Additionally, kindergarten instructional manuals frequently cite language scaffolding during imaginary play (NAEYC, 2003; Rog, 2011); for instance, while students are pretending to plant a garden, their teacher is modeling words, such as 'root' and 'seedlings.' In addition to providing words, it is often recommended that teachers prompt children with language-based questions (Invernizzi, 2003; National Governors Association Center for Best

Practices/Council of Chief State School Officers [NGACBP/CCCSO], 2010a), thereby necessitating children to engage their receptive and expressive language skills simultaneously. Letter and print knowledge generally require more direct instruction (Bergan, 2008; Blackwell-Bullock, Invernizzi, Drake, & Howell, 2009) so kindergarten teachers are encouraged to implement lesson plans around these explicit skills by engaging in whole- or small-group reading/writing activities (Blackwell-Bullock et al., 2009; Rog, 2011). In sum, kindergarten teachers are expected to scaffold children's language and literacy skills through a combination of play-based and directed learning activities tailored to the differing skills of students at entry.

Expectations of kindergarten math instruction. Although early math skills are crucial to children's problem-solving skills (Duncan et al., 2007; NCTM, 2002), a growing body of literature reveals kindergarten teachers are uncomfortable with and subsequently avoid math instruction (Copley, 2004; Hart, 2002; National Mathematics Advisor Panel, 2003). As a result, a joint effort between NAEYC and the National Council of Teachers of Mathematics (NCTM; NAEYC/NCTM, 2002/2010) established evidence-based content and instruction standards for kindergarten math. In addition to detailing the key components of early math skills (numbers/operations, geometry/spatial relations, and algebra/data analysis), the NAEYC/NCTM (2002/2010) standards suggest several methods of implementation aiming to help teachers with math instruction. Appropriate practices, as detailed in the standards and extended elsewhere (e.g., Samara & Clements, 2006), include combining verbal and visual methods of counting, sorting/making patterns with objects, and creating pictures with shapes (NAEYC/NCTM, 2002/2010). These practices are consistent with a recent national reevaluation of curriculum standards, which stated “in [k]indergarten, instructional time should focus on two critical areas: (1) representing and comparing whole numbers, initially with sets of objects; (2) describing shapes and space” (NGACBP/CCCSO, 2010b, p. 6). Evaluations of kindergarten math suggest children are best able to digest this instruction in one-on-one or small-group instruction (Clements, Sarama, & DiBase, 2004) suggesting teachers must first assess where children are in their math abilities and then decide how to use these incoming skills to scaffold them through un-mastered areas of learning individually or in small groups.

Expectations of kindergarten social skills and approaches to learning instruction. Instruction aimed at enhancing children’s social and approaches to learning skills is recognized as critical to a child’s success in the kindergarten and beyond due to their significant effects on children’s daily participation in the classroom (Bronson, 2006; Raver & Zigler, 1997). Together, teaching practices in these domains are typically referred to as social-emotional learning (SEL; Webster-Stratton & Reid, 2004; Zins & Elias, 2006), in that they aim to further a child’s development in social and approaches to learning skills. School-based SEL can be achieved through two avenues: universal strategies and explicit instruction. Universal, classroom-based SEL strategies are considered the first step in promoting development in these areas (Durlak, Weissberg, Symnicki, Taylor, & Schellinger, 2011; Zins & Elias, 2006). At the classroom-level, teachers who create an environment that fosters social-emotional development help facilitate children’s models of relationships (Bronfenbrenner, 1990; Bronson, 2006; Howes, 1997; Howes & Matheson, 1992). Evidence-based strategies that promote this type of emotional support in kindergarten classrooms include emphasizing problem-solving and positive reinforcement of prosocial interactions (Durlak et al., 2011; Gershenson, Lyon & Budd, 2010; Stevahn, Johnson, Johnson, Oberle, & Wahl, 2000). Teachers can also use universal strategies to help children develop positive work-related skills by creating classroom environments that are predictable and organized (Ostrosky, Jung, Hemmeter, & Thomas, 2007; Pianta et al., 2008). In classrooms such as these, children learn how to be patient, respect authority, follow directions, and obey rules (Jeynes 2006; Ostrosky et al., 2007). Classroom-wide strategies demonstrated to be effective in improving children's approaches to learning include implementing consistent routines and providing visual schedules (Hayes & Creange, 2001; Pianta et al., 2008).

In addition to implementing universal strategies to create emotionally supportive and organized classrooms, teachers can include specific instruction targeting social-emotional development. Explicit social skills instruction typically involves social stories and role plays that emphasize interpersonal problem-solving skills (Webster-Stratton & Reid, 2004). Teachers can also scaffold positive interactions with peers and learning tasks through direct verbal guidance (Bronson, 2006; Crockenberg, Jackson, & Langrock, 1996). In this way, teachers can use a combination of universal and targeted strategies to improve children's social and self-regulatory skills.

Kindergarten instruction in practice. Despite these established best practices, research examining kindergarten classrooms consistently demonstrates instruction does not align with the expectations reviewed above (Bredekamp & Copple, 1997; Early et al., 2005; Goldstein, 2007). Data collected from over 700 kindergarten classrooms as part of a multi-state evaluation highlight that 40 percent of a child's day is spent in non-instructional time, 28 percent was dedicated to language/literacy instruction, and only 11 percent was spent focusing on math instruction (Early et al., 2005). Furthermore, when instruction is implemented, the interactions between teachers and students generally focus on lower-level skills and are ineffective in promoting higher learning (Early et al. 2005; Hamre & Pianta, 2007). These less-than-ideal practices are consistent across learning domains, suggesting a wide array of targets for intervention.

Language/literacy instruction in practice. Evaluations of language and literacy instruction occurring during the kindergarten year highlight the diversity of practices during these instructional periods (Stipek & Byler, 2004 Xue & Meisels, 2000). Using the ECLS-K dataset, Xue and Meisels (2000) illustrated that teachers typically reported implementing a combination of systematic phonics/letter instruction and literature-based reading to varying degrees. Stipek and Byler (2004), however, found that teachers were not reliable reporters of their language and literacy instructional practices. Although kindergarten teachers in their sample reported emphasizing contextualized learning, such as small-group readings, more than didactic instruction, they were much more likely to implement didactic lessons. Teachers are able to demonstrate more effective language and literacy practices when provided with support in implementing evidenced-based instruction, however, (Baker & Smith, 1999; Cabell & Downer, 2011; Schuele et al., 2008), indicating the potential for targeted PD.

Math instruction in practice. In 2008 a nationally representative review found there was a lack of math instruction occurring during the early years of schooling (Cross, Woods, & Schweingruber, 2009). Specifically nearly one-third of kindergarten classrooms receive less than 30 minutes of math instruction every day (Cross et al., 2009). Furthermore, the math instruction that occurred was often facilitated through a higher percentage of didactic, teacher-directed instruction than other instructional periods (Bodovski & Farkas, 2007; Stipek, 2004; Stipek & Byler, 2000). These teacher-centric practices are especially common during basic skill instruction such as knowing math rules and steps (Stipek & Byler, 2004), a common feature of kindergarten math learning (Baroody, Eiland, & Thompson, 2009; Dyson, Jordan, & Glutting, 2013). Furthermore, Stipek (2004) noted interaction between didactic math instruction and SES, in that teachers serving students of lower SES were more likely to use teacher-centric practices for math instruction than those teaching higher SES students. Similar to the mechanism described for language and literacy domain, kindergarten teachers are able to improve their math instruction with relatively little in-service support (Chard et al., 2008). For example, Chard and colleagues (2008) found that kindergarten teachers were able to implement a math curriculum focused on number sense after a half-day PD training, which increased students' math achievement. Results from this and similar models (e.g., Clements & Sarama, 2003; Ginsburg, Galanter & Morgenlander, 2004) suggest teachers' math instruction is malleable with relatively low dosage PD.

Social skills and approaches to learning instruction in practice. Few studies are available on the quality of social skills and/or approaches to learning instruction in kindergarten, which is at least partly attributable to the relative recency of the initiative to understand and measure the social-emotional side of classrooms (U.S. DOE, 2013). As such, only three states had adopted SEL-related benchmarks into their educational standards of learning (Dusenbury, Weissberg, Green, & Domitrovich, 2014). When specific expectations are established for social and approaches to learning skill acquisition, teachers are noted as engaging in SEL-related instruction more frequently and students demonstrate better skill development in these domains (Payton et al., 2008; Zinsser, Weissberg & Dusenbury, 2013). On the other hand, preliminary evidence suggests few kindergarten teachers implement robust social-emotional instruction without these standards or assistance from a SEL-specific curriculum (Dusenbury et al., 2014; Kramer Caldarella, Christensen, & Shatzer, 2010).

The lack of social skills and approaches to learning emphasis in kindergarten classrooms could also stem from teachers' discomfort with SEL. A survey of pre-service teachers revealed they felt more unprepared to implement social-emotional instruction as compared to any academic domain (Cook, 2002). Other explanations for this lack of SEL practices could relate to the reluctance of schools and educators to dedicate instructional periods to non-academic domains (Kramer et al., 2010; Seifer, Gouley, Miller, & Zakriski, 2004). Regardless of the explanation, teachers' lack of SEL instruction is easily supplemented with relatively minor interventions (Bierman et al., 2008; Kramer et al., 2010), creating a prime target for professional development aiming to align instruction with evidence-based practices.

Contextual variation as a barrier to effective instruction. Given the differences between expectations and practices during kindergarten instruction, a natural question arises: what causes these discrepancies? In addition to the challenges associated with tailoring instruction to the between- and within-child difference in skills, structural and demographic variation may also contribute to the theory-practices gap in kindergarten instruction.

Structural variation of kindergarten. The logistics of kindergarten classrooms, which differ from that of first through twelfth grade (Bredekamp & Coople, 1997;

Goldstein, 2007), may influence teachers' ability to implement effective instructional practices. Only 15 states mandate kindergarten attendance (Samuels, 2014), indicating policies surrounding kindergarten vary within the United States. Despite the lack of a national mandate, the 2010 census found that 99 percent of five-year-old children were enrolled in kindergarten (Davis, & Bauman, 2013). Three-fourths of kindergarteners were enrolled in a full-day program, defined as necessitating that children arrive and leave school at times similar to those in older grades (Davis & Bauman, 2013). Although this represents a significant increase from the national average of 37 percent in 1987 (Davis & Bauman, 2013), nearly one-fourth of children who attend kindergarten in the U.S. are still receiving partial-day instruction. This distinction is important, as full-day programs have been consistently associated with better learning outcomes. A statewide study of children in kindergarten classrooms across Ohio found that children enrolled in full-day programs outperformed children in half-day programs concurrently and through first grade (Cryan, Sheehan, Wiechel, & Bandy-Hedden, 1992). Furthermore, within the classroom, the number of kindergarteners present also varies greatly. While data from the nearly 1,000 classrooms sampled via ECLS-K revealed the average class size to be 20, kindergarten classrooms varied from two to 52 children (Milesi & Gamoran, 2006). These structural variations unique to kindergarten may make it challenging to implement child-centric, evidence-based practices due to time and/or logistical constraints.

Demographic variation of kindergarten. The demographic diversity of kindergarten classrooms may also impact teachers' enactment of effective instructional practices. Of the four million children enrolled in kindergarten in 2011, thirty-three percent originated from non-white backgrounds and twenty-one percent lived in households where the primary language spoken was not English (Davis & Bauman, 2013). Although all public school levels house diverse landscapes of students (Center for Public Education, 2012; Olsen, 1997), kindergarten is the first experience of formalized schooling for many children of non-white backgrounds (Burchinal, Vandergrift, Pianta, & Mashburn, 2010; Early et al., 2005). This creates unique challenges for kindergarten teachers, as they not only have to balance curricular and learning demands, but also differences in cultural and linguistic expectations. Given this ever-increasing diversity, teachers need to understand individual differences that children bring with them into the classroom, and an effective way to achieve this is through using data systematically (Schmoker, 2003; Slavin et al. 2011).

In summary, there are significant differences between what is expected of kindergarten instruction and what teachers implement, a discrepancy which can be partly attributed to the contextual variation endured by kindergarten teachers. This variation makes it challenging to enact best practices without added supports. One such avenue for supporting teachers in selecting evidence-based instructional practices is utilizing assessments of children's incoming skills to make data-driven decisions. Using data to inform classroom practices has the potential to help teachers understand where children are in their learning and how to facilitate further learning regardless of variation (Slavin et al., 2011). Next, I will introduce the theory behind use of assessments in education and provide an overview of teachers' data-use practices to understand how to help overcome barriers to using KEA data to inform kindergarten instruction.

Student Assessments

Educational assessments include a wide range of methods for evaluating performance across individuals in schools (Gipps, 1994; Nitko, 2001). To understand the robust potential of these assessments, resulting data must be seen as contributing to a broader theory of education, such as that put forth by Ramaprasad (1983). In this model, Ramaprasad (1983) describes three pillars essential to effective teaching and learning: 1) establishing where the learners are in their learning, 2) establishing where they are going, and 3) establishing what needs to be done to get them there. The first pillar of this theory is typically ascertained through formal and/or informal assessments of students' learning, whereas the second and third are achieved through the interpretation and application of the assessment data.

Purpose of assessment implementation. Student assessments exist in a variety of forms (Lonigan, Allan, & Lerner, 2011; McMillan, 1997; McMillan, Myran, & Workman, 2002) and decisions around which assessments to use are driven by the purpose or theory behind implementation. Assessment implementation can be distinguished by its relation to learning (Elwood & Klenowski, 2002); that is, was the assessment used to evaluate or to promote learning? Assessments *of* learning aim to measure what a student has learned, whereas assessments *for* learning enable learning through instructional decisions based on assessment data. In other words, the purpose of an assessment has direct implications for how the data are used afterward (Black & Wiliam, 2009; Firestone et al., 1998). Consequently, it is essential to understand the theories that drive decision-making around assessment implementation and the subsequent instructional practices when evaluating teachers' use of data in the classroom.

Assessment of learning. As aforementioned, the construct 'assessment *of* learning' refers to testing implemented to capture what a student has already learned (Elwood & Klenowski, 2002). Most often, results of this type of assessment are used to hold an individual and/or institution accountable for what was or was not learned based on schooling experiences. Accountability in regards to educational assessment is often associated with the SBM, but in reality covers a much broader concept. By definition, accountability refers to a call to justify one's beliefs or actions (Lerner & Tetlock, 1999), and within schools this typically refers to justifying what and how students are learning. In this way, assessments of learning are applied in a multitude of scenarios to gauge learning in order to hold one or more parties accountable.

The theory behind assessment of learning. Evaluating learning is not a new phenomenon; early pedagogical theories dating back to the seventeenth century promote the need for students to demonstrate their knowledge (Katz, 1968; Pellegrino, 2004; Schneider & Hutt, 2014). Although traditional assessment of learning practices (i.e., grading) have been conducted at the student-level, the call to justify learning can occur at any level of education. Based in motivational theory, assessment used for this purpose aims to encourage students, teachers, schools, districts, or states through external avenues (Herman, 2004; Supovitz, 2009). At the student-level, assessment of learning is meant to ensure students have put forth the effort necessary to learn and provides them with feedback about their performance (Guthrie, 2002). Assessment of learning aggregated to the teacher-, school-, district-, or state-level shifts the burden of learning and potential rewards/consequences from the student to the larger entity (Firestone et al., 1998). In other words, did the teacher, school, district, or state use resources effectively to educate students? In this way, assessments of learning fulfill the first of Ramaprasad's (1983) pillars of teaching and learning in that they establish where students are in their learning.

Assessment of learning in practice. By nature, assessments of learning must occur after instruction, because the central goal of these assessments is to evaluate what a student has learned (Elwood & Klenowski, 2002; Firestone et al., 1998), typically through assigning a rating or grade based on norms (Figlio & Loeb, 2011; Schneider & Hutt, 2014). Individual grading scales originated as a means to heighten student motivation for learning in the late 1700s (Grant, 2003; Schneider & Hutt, 2014). As K-12 enrollment increased through the early twentieth century, educational scholars sought a more standard method for comparing achievement across students (Parker, 1902; Smallwood, 1935). This culminated in the well-known A-through-F grading scale, which has undergone few changes since its introduction in the 1940s (Schneider & Hutt, 2014).

More recently assessments of learning have been implemented using measures of teacher-, school-, district-, and state-level efficacy as a part of the SBM, which has led to mixed results. Several researchers have demonstrated positive gains in older students' achievement after the installation of accountability measures (Dee & Jacob, 2009; Figlio & Loeb, 2011; Neal & Shanzenbach, 2010). For example, by comparing the achievement gains of fourth and eighth graders in states who had long-standing accountability incentives with those from states who did not prior to the No Child Left Behind Act (2001), Dee and Jacob (2009) demonstrated a positive association between accountability standards and learning gains. In addition to student-level outcomes, the increase in testing mandates has led to a greater availability of specific tests, as well as curricula expansion and more standardized teaching practices across states and districts (Au, 2006; Firestone et al., 1998; Grant, 2003; Monfils et al., 2004).

Accountability measures have also been associated with negative outcomes (Herman, 2004; Stecher & Barron, 2001), especially when evaluated within the context of young students' learning (NAEYC, 1998, 2003). One of the most salient drawbacks is the possibility that allocation of educational funds based on test scores may incentivize teaching in narrow areas (Booher-Jennings, 2005; Hatch, 2005; NAEYC, 2002). This is particularly problematic for lower grades, as research demonstrates decreases in time allotted for developmentally appropriate, child-centric exploration in classrooms participating in high-stakes accountability (NAEYC, 1998, 2003). In addition, there is evidence to suggest students' knowledge base is reduced and instruction is more likely to be fragmented by subjects in schools with strict accountability procedures (Au, 2006; NAEYC, 2003; Yeh, 2005). Instructional fragmentation is detrimental to students who struggle in specific content areas, such as students with disabilities (Morison, McLaughlin & McDonnel, 1997) or English language learners (Palmer & Lynch, 2008; Thomas & Collier, 1997), because they are not able use context clues or cross-content connections to compensate for other deficits (Entwistle & Ramsden, 2015). Informed by this literature, NAEYC's (2003) official stance is against the use of accountability measures in kindergarten, indicating they have led to a "focus on a limited range of skills, causing teachers to narrow their curriculum and teaching practices… especially when stakes are high" (p.4).

Assessment for learning. In contrast to assessments of learning, assessment *for* learning is meant to enable students to learn through the process of testing (Elwood &

Klenowski, 2002). In this model, assessments are used either directly to promote learning, or indirectly to promote teaching practices (Black & William, 2009).

The theory behind assessment for learning. Although the use of assessments for learning is a relatively new concept compared to the previously presented theory, the concept has gained significant theoretical and practical support over the past three decades (Black & Wiliam, 2009; Elwood & Klenowski, 2002). In contrast to assessments of learning, assessments used for learning build upon traditional testing practices by using the data gathered to inform instructional decisions (Black & Wiliam, 2009; Elwood & Klenowski, 2002). To actualize this theory, an assessment must first describe student performance along a continuum, leaving room for improvement, thereby establishing the first of Ramaprasad's (1983) pillars. From there, teachers interpret the data with the intent of informing their practice, establishing the third pillar of what needs to be done to get the student to the ultimate learning goal (Ramaprasad, 1983).

Assessment for learning in practice. Unlike assessments of learning that occur after instruction, assessments for learning are administered before or during instruction to help in promoting learning (Elwood & Klenowski, 2002). Assessments for learning that are administered before instruction provide data about a student's knowledge or skills at one time point with the intent of helping the teacher individualize content (State Education Resource Center [SERC], 2012). Assessments administered before learning takes place are often referred to as screeners, because they are typically used for diagnostic screening in or out of mainstream classrooms (Gredler, 1997; SERC, 2012). Consequently, the term screener often connotes determining children at the low end of the learning continuum, while in reality screener assessments simply aim to catch a snapshot of abilities at one point in time. After obtaining this snapshot, data from the assessments are used to inform the necessary instruction.

With the recent emphasis on understanding students' skills upon kindergarten entry, screeners have been used more widely to capture incoming ability levels at this time point (Costenbader et al., 2000; Niemeyer & Scott-Little, 2001; Saluja et al., 2000; Scott-Little & Niemeyer, 2001). As such, KEAs have become the most widely used measure of assessment before learning next to diagnostic screeners for special education eligibility (Sterdon & Berger, 2010).

Assessments for learning can also be implemented as a part of, as opposed to before, the learning process. This can be achieved by directly improving learning through formative assessments or indirectly by informing practice through progress monitoring tests. In a formative assessment, the instructor provides feedback until a student has mastered the concepts covered by an assessment (Black & Wiliam, 2009). Through this process, a teacher works with a student on an assessment until he or she has achieved the expected level of learning. This technique is rarely used in early elementary school (McMillan, 1997; McMillan et al., 2002), likely as a result of the complex components necessary for effective formative assessment. Black & Wiliam (2009) detail three crucial elements for formative assessments to be cogent: external (teacher's feedback), internal (student's conceptualization of learning) and the interaction of the two, which requires articulate verbalization. Although kindergarteners have the capacity to internalize and conceptualize aspects of learning, it is often challenging for them to do so verbally (Coates & Hartup, 1969; Piaget, 1926). Consequently, the external (teacher's feedback) and internal (student's learning) are both functioning, but the interaction between the two is missing, limiting the utility of this assessment approach in kindergarten.

Progress monitoring tests also allow for assessment for learning during instruction, but differ from formative assessments in that they aim to affect teachers' practices rather than directly impacting learning (SERC, 2012; Stecker, Lembke & Foegen, 2008). Progress monitoring involves uniform tests given on multiple occasions to assess how much a child has learned since the last installment (Stecker et al., 2008). In this way, learning and assessments are often intermingled so that the assessment can continuously inform the teacher's practice.

Research evaluating assessments for learning. When assessments are used either before or during learning, students' outcomes are consistently improved (Ball & Gettinger, 2009; Black & Wiliam, 2009; Strand & Cerna, 2010; Stecker et al., 2005; Wiliam, 2011). Meta-analyses have found that when evaluated within the scope of students' academic achievement, assessments for learning are associated with effect sizes ranging from .2 to 1.10 (Black & Wiliam, 2009; Dempster, 1991; Wiliam, 2011). Preliminary evaluations of KEAs have provided similar results (Costenbader et al., 2000; Rock & Stenner, 2005; Maxwell & Clifford, 2004; Meisels, 1999). It is important to note that these types of assessments only provide data that must then be interpreted and applied to inform teaching practices. A review of progress monitoring tests based in curriculum highlights that teachers must actually *change* practice based on assessment data in order for students in their classrooms to demonstrate the most significant gains across content areas (Stecker et al., 2005; Wiliam, 2011). Although most studies included in Stecker, Fuchs, and Fuchs' (2005) review evaluated upper elementary classrooms, three studies (Fuchs, Fuchs, Hamlett, & Ferguson, 1992; Fuchs et al., 1997; Fuchs, Fuchs, Hamlett, & Stecker, 1991) established positive effects on math and reading when assessments were used to inform practice in first grade. Other studies have now replicated similar effects in kindergarten classrooms; Ball and Gettinger (2009) demonstrated that when data from literacy assessments were paired with instructional feedback for teachers around their teaching, kindergarteners made greater improvement in literacy outcomes compared with children of teachers who were not given feedback about the data. From these and comparable investigations (e.g., Black & Wiliam, 2009; Strand & Cerna, 2010), measuring incoming skills and monitoring progress are now viewed as essential components in guiding teachers' instruction across all grade levels (Jitendra, Dupuis & Zaslofsky, 2014; NAEYC, 2003).

Content of student assessments. Across assessment purposes, measures now exist in a variety of forms that span most conceivable content areas. The most traditional form of assessments measure a student's academic knowledge and/or skills across math, language, history, science, etc. More advanced content assessments not only assess a student's knowledge, but also measure the ability to apply those skills to novel problems (SERC, 2002). Content assessments are typically used to gauge what a student has learned and performance is heavily related to a child's future content knowledge (McMillian, 1997; Perrone, 1991); that is, the best predictor of a student's future math functioning is his or her current math functioning. At kindergarten entry, math and language/literacy are typically the content areas of focus (Costenbader et al., 2000; Meisels, 1999), though recent initiatives have also supported the importance of science knowledge at this age (Kim et al., 2014; Kinzie, et al., 2014).

In addition to measuring academic knowledge and application, assessments can be used to determine other skills related to learning, such as a child's social skills and approaches to learning. Social skills assessments measure a student's ability to interact with others at age-appropriate levels, which are typically rated by teachers (Demaray, Ruffalo, Carlson, Busse & Olson, 1995; McFall, 1982). Approaches to learning assessments aim to capture a student's methods of engaging in learning, such as their curiosity, attention, and persistence (Gijbels & Dochy, 2006). Assessments of social skills and approaches to learning are readily used in early elementary classrooms, as skills in these areas at a young age are related to future academic and social outcomes (Gijbels et al., 2005; Watkins, 2001).

Validity of student assessments. Regardless of the purpose behind or content of the assessment, teachers need to know they can trust the results and to do so, the assessments must demonstrate reliability and validity (Cozby, 2001; Northwest Regional Educational Lab [NREL], 1994). Reliability refers to the consistency of scores expected over time, whereas validity refers to the ability of the assessment to be interpreted meaningfully (Downing, 2001; NREL, 1994). For an assessment to be valid, it must first be reliable, in that teachers should expect similar results from the same children taking the test under the same conditions (Cozby, 2001; NREL, 1994). Once reliable, the validity of an assessment can be evaluated via criterion-related and construct validity, each of which has particular implications for KEAs. Criterion-related validity refers to the ability of the test to reveal current or predict future performance (Cozby, 2001). KEAs are typically implemented for their criterion-related validity, as the purpose of such assessments is to diagnose where students are at and where they are going.

Construct validity, which has been argued as the most imperative for knowledge-based assessments (Downing, 2003), refers to the measurement of what is intended (i.e., the construct) and nothing else (Cozby, 2001; NREL, 1994). For example, to achieve construct validity, a KEA of math must measure only a student's incoming math abilities and not their reading and/or writing abilities. When an assessment is psychometrically sound, in that it is reliable and has adequate criterion-related and construct validity, teachers can feel confident in using the results to inform their practice.

When the type, content, and validity of assessment have been given the necessary weight and are employed accurately, research has consistently demonstrated that classroom-based student assessments can be used to effectively promote students' learning (Black & Wiliam, 20099; Strand & Cerna, 2010; Stecker et al., 2005).

Teachers' assessment practices. If assessments have been demonstrated to be effective in informing appropriate instruction and improving children's learning, why is this not occurring regularly in classrooms? When teachers are surveyed, they often provide responses surrounding logistics; that is, a combination of lacking time, money, and/or resources affects their ability to use assessments to inform their practice (Looney, 2005; Wesson et al., 1984). Although these managerial aspects certainly contribute, evaluations have isolated barriers to effective assessment down to three features: teachers' ability to use assessment data effectively (Datnow, et al., 2007; Stecker et al., 2005; Wayman & Jimerson, 2014), beliefs about assessment more generally (Brown, 2004a; 2009), and their feelings of self-efficacy with relation to data-use (Pierce & Chick, 2011). More so than logistics, which can be addressed through educational policy, the effects of teachers' ability to interpret data effectively, beliefs about assessments, and feelings of self-efficacy require more direct intervention.

The ability to use assessment data effectively. Research consistently highlights the difficulty teachers experience when they are asked to use assessment data in their classrooms (Chick & Pierce, 2013; Stecker et al., 2005; Wayman & Jimerson, 2014). Using data in classrooms consists of three compounding skills: collecting reliable data, interpreting those data accurately, and applying those interpretations to instructional decision-making.

Collecting data. Collecting accurate and useful data on students' learning requires knowing what data to collect, obtaining a reliable and valid measure of that data, and administering that measure in an appropriate manner (Bernhardt, 2013; Johnson, 2002). Teachers generally have little control over the type or manner of data they are collecting (Bernhardt, 2013; Johnson, 2002), leaving the latter for consideration. Research indicates teachers are generally accurate test administrators when provided with the necessary supports, such as training and classroom coverage (Kauffman, Moore Johnson, Kardos, Lui, & Peske, 2002; Limbos & Geva, 2001; Stecker et al., 2005). These data suggest it is not data collection that is challenging for teachers, but rather interpreting and/or extrapolating these data to actual practices. Indeed, Stecker et al.'s (2005) review found that teachers accurately collected data, but experienced difficulty understanding and applying the data to shifts in instructional methods, which ultimately resulted in little change in student learning.

Interpreting data. After collecting reliable data, teachers must then interpret these data accurately to utilize them appropriately in their classroom practices. To do this, teachers must possess an elementary level of ‘statistical literacy’ which allows for understanding relationships among quantitative data (Chick & Pierce, 2013; Downing, 2003; Mills, 2000; Pierce & Chick, 2011). Building off of Curcio’s (1987) and Wild and colleagues’ (2011) concepts of inference, Pierce and Chick (2011) present a “framework for professional statistical literacy” (p. 633) which denotes three hierarchical skills included in interpreting data: reading values, comparing values, and analyzing the data as a whole. The first includes understanding basic statistical terminology (e.g., ‘average’) and reading graphical presentations (Curcio, 1987; Pierce & Chick, 2011). This capacity is necessary to move to more complex aspects of data interpretation, like comparing values. At this level, teachers must evaluate trends among different values thereby making initial inferences (Chick & Pierce, 2013; Pierce, Chick, Watson, Les, & Dalton, 2014). The final conclusions come from considering the data set as a whole by aggregating these inferences and making global conclusions about the individual comparisons (Pierce & Chick, 2011; Pierce et al., 2014).

Evaluations of teachers’ data interpretation show robust deficits in their abilities (Fuchs et al., 1990; Pierce & Chick, 2013; Roth, McDuffie, & Morrison, 2008). These challenges begin at the first two levels of data interpretation: reading and comparing values. Unfortunately, preservice curricula allocate little time helping soon-to-be teachers to master these skills (Mandinach, Friedman, & Gummer, 2015; Roth et al., 2008). Consistent with this, assessments of teachers’ capacity to read and compare data values are low (Fuchs et al., 1992; Pierce & Chick, 2011; Roth et al., 2008). It is not surprising, therefore, that teachers struggle to compare these values and make global inferences (Stecker et al., 2005).

Applying data to classroom instruction. The final step in applying assessment data to classroom practices necessitates extrapolating the inferences made from the previous steps to instructional decisions (Desimone, 2009; Ebbeler et al., 2016). To do this, teachers must relate a particular inference to a specific instructional strategy. For example, if through data collection and interpretation it was determined that Student A had a deficit in numerical awareness but was strong in vocabulary knowledge, a teacher could implement a lesson that involved counting within a story to use the students' verbal strength to build their numeracy skills.

Research has demonstrated that even when provided with the accurate inferences about a student's learning, teachers struggle to select and apply relevant instructional strategies (Bergan & Schnaps, 1983; Fuchs et al., 1992; Fuchs et al., 1991; MacSuga & Simonsen, 2011). Similarly, Stecker et al.'s (2005) review found the largest effects on student outcomes occurred when teachers were provided with instructional practices relating to their students' data.

Professional development focused on building teachers' data-related competencies. Given the robust deficits in teachers' data interpretation and extrapolation skills, scholars have built and tested professional development (PD) models aiming to build teachers' capacities in these areas (Al Otaiba et al., 2011; Datnow et al., 2007; Hamilton et al., 2009; Wayman & Jimerson, 2014). Several existing projects have utilized professional learning communities (Ebbeler et al., 2016; Kerr, Marsh, Ikemoto, Darilek, & Barney, 2006; Lachat, & Smith, 2005), which focus on collaborative learning through group reflection centered on a topic (Vescio, Ross, & Adams, 2008). Other models have used client-centered consultation, described later, to build teachers' capacities in this domain. Regardless of the format, PD aiming to improve the translation of data to instruction cites a similar goal: to support teachers' data-related competencies. Wayman and Jimerson (2014) reviewed the components necessary to reach this goal, and isolated six competencies: "(1) asking the right questions; (2) integrating data-use with curriculum, instruction, and assessment; (3) analyzing and interpreting data; (4) linking data to classroom practice; (5) computer skills; and (6) collaborating around data" (p. 28). As such, any PD model aiming to improve teachers' assessment-related practices should strive for building these competencies.

***Teachers' beliefs around student assessments*.** In addition to their limitations in using data, teachers' beliefs around the use of student assessments contribute to their testing practices (Brown, 2009; Calderhead, 1996; Fives & Buehl, 2012; Heaton, 1975; Pyle & DeLuca, 2013; Warren & Nisbet, 1999). Although this construct takes on many names such as "stance," "goals," "beliefs," and "conceptions," the larger concept remains focused on the ideas that teachers hold about assessments' use in learning that influence their classroom practice (Brown, 2004a; Calderhead, 1996; Mansour, 2009; Pajares, 1992). This is consistent with more general theories describing the relationship between beliefs and behaviors, which emphasize that people most often act in accordance with their beliefs (Ajzen, 2005; Bandura, 1986; Heaton, 1975; Pajares, 2002; Warren & Nisbet, 1999). Therefore, teachers' beliefs offer a framework for understanding their global and daily classroom choices (Brown, 2004a; Marton, 1981).

Evaluations of this framework reveal teachers' beliefs about assessment purposes factor into four primary categories that relate to their classroom practices: school accountability, student accountability, improvement of teaching and learning, and irrelevance (Brown, 2004a, 2009). School and student accountability refer to the belief that assessments are implemented to hold schools or students accountable for learning. These two categories of beliefs map onto the theory of assessments *of* learning, in that they establish where students are in their learning after instruction has occurred. These teachers employ more norm-based, externally created measures (Brown 2009), which provide a quantifiable method to determine if students meet learning benchmarks. The third belief, that assessments should be used to improve teaching and learning, pertains to the theory of assessments *for* learning; that is, this belief surrounds using assessments to promote learning before or during instruction. Teachers who hold the belief that assessments should be used to improve teaching and learning are more likely to use teacher-created or informal assessment techniques (Brown, 2009), reporting that they are better able to build learning using these measures. The final category of beliefs, that assessments are irrelevant, represents the idea that assessments are not useful (Brown, 2004a). Teachers who endorse this belief employ practices similar to those favoring the 'improvement' belief, but report believing they hold no value for their day-to-day classroom interactions (Brown, 2004a; 2009).

As aforementioned, teachers' beliefs around assessment follow more traditional theories describing the relationship between beliefs and subsequent behaviors (Bingimlas & Hanrahan, 2010; Pajares, 1992), which include the potential for holding contradictory beliefs. Several evaluations of beliefs about assessment indicate teachers can hold multiple and even contradictory beliefs (Brown, 2004a; Cizek et al., 1995). This is partly due to the plurality of uses of assessments; assessments can hold multiple purposes in the classroom. For instance, a teacher could believe that assessments are useful for holding students accountable, while also believing that he or she could use the assessments to inform practices. In support of this, Brown (2004a) found the four factors of beliefs were moderately or highly correlated with the exception of the association between ‘school accountability’ and ‘irrelevance,’ which only produced a weak relationship. This suggests that most teachers hold some combination of beliefs about the purpose of assessments in classrooms (Brown, 2004a). The intercorrelations between these factors have been replicated (Brown, Chaudhry, & Dhamija, 2015; Brown, Lake, & Matters, 2011) lending support to the hypothesis that teachers’ beliefs should not be perceived as mutually exclusive. Instead, their beliefs about assessment uses in the classroom entail a delicate combination of perspectives on improvement and accountability (Brown, 2004a; Brown et al., 2011), both of which may contribute to their classroom assessment practices.

A unique aspect about teachers’ beliefs are driven by their experiences, indicating teachers form schemas about teaching practices from their daily interactions in the classroom (Fang, 1996; Mansour, 2009). This suggests their beliefs can shift with changes in experiences. Changing teachers’ practice is not simple, however; several researchers have noted that classroom changes only result from coordinated efforts between the teacher and a trusted colleague (Batten, 1991; Richards, Gallo, & Renandya, 2001; Vonk, 1991). In a survey of 341 teachers, a third of respondents indicated having changed their practice and beliefs through collaboration with a colleague or consultant (Richards et al., 2001). In addition, research demonstrates teachers’ beliefs are subject to their content and pedagogical knowledge (Brighton, 2001), suggesting that if their knowledge about teaching was to change, their beliefs about teaching may also change. Although a complex process, it is possible to change teachers’ beliefs around their practices, including the use of assessments, through collaboration and increasing content knowledge.

Factors affecting teachers' beliefs. Several variables have been established to influence teachers' belief systems more generally (Brown, Kennedy, Fok, Chan, & Yu, 2009; Fang, 1996; Hamilton et al., 2007). As such, it is important to understand and consider the potential effects of these factors when attempting to shift teachers' beliefs about assessment data-use.

Factors outside of teachers' control seem to be most influential on their belief systems (Brown et al., 2009; Hamilton et al., 2007). That is, the larger political contexts in which teachers operate contribute significantly to conceptions of their practices (Brown et al., 2009; Hamilton et al., 2007). With regards to assessment utility, teachers' beliefs have been found to differ by adjacent localities which house discrepant educational mandates (Hamilton et al., 2007). Although evaluations of these differences are quasi-experimental, they provide correlational evidence of a connection between educational policies and the beliefs teachers endorse (Brown et al., 2009; Hamilton et al., 2007). This association is hypothesized to be mediated by the proximal interactions that occur within schools as a result of these mandates (Hallinger, Bickman, & Davis, 1996; Ingram, Lewis, & Schroeder, 2004; Thapa, Cohen, Guffey, & Higgins-D'Alessandro, 2013). For example, school personnel operating in districts which promote data-driven decisions are more likely to collaborate around data-use (Hamilton et al., 2007). The interactions that occur within schools coalesce into the construct of school climate, or the norms, beliefs, and values that represent the majority of staff in the school (Anderson, 1982; Thapa et al., 2013). Positive, data-focused school climates have been shown to promote beliefs about the use of assessment for learning (Goddard & Goddard, 2001; Hipp & Bredeson, 1995; Lee, Dedick, & Smith, 1991) and subsequently higher use of data-driven practices in the classroom (Ingram, Lewis, & Schroeder, 2004; Massell, 2001). When evaluating teachers' beliefs, therefore, it is crucial to consider the context of schools' climate and its potential influences on teachers' conceptions and related actions.

There is less consistent evidence linking demographic factors to teachers' beliefs about assessment. For example, analyses from Cizek and colleagues' (1995) survey of 143 elementary, middle, and high school teachers revealed that although teachers' practices differed by their years of experience and gender, qualitative interviews revealed teachers' beliefs about assessment were not affected by these same demographics. Examinations of more general beliefs systems teacher hold indicate demographic variables are influential (Fang, 1996; Pajares, 1992), however, suggesting there is the potential for associations with beliefs about assessment uses. As such, scholars of teachers' beliefs recommend considering demographic influences when measuring belief-related constructs (Brown, 2004a, Cizek et al., 1995 Parajes, 1992).

In sum, the research presented above indicates it is prudent to consider the potential influences of school and individual characteristics when evaluating teachers' beliefs about their practices and the behavioral results of those beliefs.

PD aimed at shifting teachers' beliefs about assessment. Teachers' beliefs in any form affect their practice (Brown, 2004a), suggesting beliefs are crucial components to understanding and affecting teachers' practices. Consequently, understanding and influencing belief systems around assessment purpose arises as a possible avenue for changing assessment practices in the classroom. Clark and Peterson (1998) proposed that

PD aiming to affect teachers' beliefs must include explicit exploration of these belief systems through reflection and questions. Kagan (1992) reiterated this point, noting PD with this emphasis must "require them to make their preexisting personal beliefs explicit" and "challenges the adequacy of those beliefs" (p. 77). This theory is supported by empirical evidence demonstrating changes in belief schemas when individuals are scaffolded to reflect on these schemas (Brown, Chaundry & Dhamija, 2015; DeBacker & Crowson, 2006; Luft, 1999; Mansour, 2009; Romme & van Seggelen-Damen, 2015). Few studies have examined this tactic's utility in changing teachers' beliefs, but there is indication that reflective questions operates similarly with this population by illuminating beliefs that were previously unexplored (Brown et al., 2015; Romme & van Seggelen-Damen, 2015). Although this is only preliminary evidence, these results suggest teachers can begin to change their beliefs through reflection that occurs during the implementation of assessments for learning.

Teachers' self-efficacy in using data to inform practice. Self-efficacy, or an individual's expectation of the quality of their behavior (Bandura, 1986), also has the potential to affect changes in teachers' classroom practices. It is well-established that feelings of efficacy are strongly related to motivational action (Bandura, 1993, 1986), which suggests that self-efficacy would be a necessary precursor to any change expected. The relationship between self-efficacy and related actions has been demonstrated with teachers, in that teachers with higher self-efficacy implement more effective instruction (Allinder, 1994; Tschannen-Moran et al., 1998) and facilitate higher achieving student learning (Ashton & Webb, 1986).

Consequently, it is essential to first build teachers' feelings of efficacy with respect to data-use before expecting changes in practices. Although no specific data about teachers' feelings of efficacy with data-use are available at the time of writing, we can assume this type of self-efficacy is relatively low given teachers' low reported self-efficacy surrounding assessment and instructional decisions more generally (Tschannen-Moran et al., 1998; Tschannen-Moran & Woolfolk Hoy, 2001).

Factors affecting teachers' self-efficacy. Several influences on teachers' self-efficacy have been identified and have implications for the mechanisms through which their self-efficacy can be augmented. At the most basic level, the resources teachers are allotted affect their perception of how effective they are (Tschannen-Moran et al., 1998). That is, teachers who are provided with greater educational capital like implementable curricula and immediate classroom support/aid report higher feelings of efficacy (Tschannen-Moran et al., 1998). These resources often relate to school climate, in that schools with more positive climates provide greater levels of support to their teachers (Thapa et al., 2013). As a result, positive school climates and effective leadership are related to higher rates of self-efficacy in teachers (Goddard & Goddard, 2011; Hipp & Bredeson, 1995; Lee et al., 1991). As was true for beliefs, therefore, it is essential to consider school climate when examining how to facilitate teachers' self-efficacy.

PD aimed at improving teachers' self-efficacy around assessment data-use. Research has shown coaching can have an impact on teachers' self-efficacy across instructional practices (Ross, 1992; Tschannen-Moran & McMaster, 2009). In a review of four studies of PD models aiming to improve teachers' self-efficacy, Tschannen-Moran and McMaster (2009) found the effects were dependent on two components: (1)

opportunities for mastery and (2) subsequent analysis about this mastery. Mastery experiences had been previously established as essential to affecting feelings of efficacy (Ross, 1992; Woolfolk-Hoy & Burke-Spero, 2005), and these effects were enhanced when teachers were asked to analyze these experiences (Timperley & Phillips, 2003; Tschannen-Moran & McMaster, 2009). To facilitate changes in self-efficacy, with the ultimate goal of affecting practices, PD must provide teachers with a space for mastery experiences and supplement this with helping them analyze these experiences.

Consultation as an Avenue for Targeted Professional Development

Consultation offers a method of PD through which teachers' skills, beliefs, and self-efficacy surrounding assessment data-use can be supported to improve their classroom practices. Consultation, or the process of mutual problem-solving between a consultee and consultant (Dougherty, 2009; Kratochwill & Bergan, 1990), is a common form of school-based PD that allows for targeting specific teaching practices. Although many models of consultation exist (e.g., Kratochwill & Bowman, 1990; Lowman, 2002), educational settings most typically engage in consultee-centered consultation (Caplan, 1970; Meyer, 1975). Consultee-centered consultation involves three parties: the consultant (individual facilitating problem-solving), consultee (individual directly receiving consultation), and client (the focus of the consultation). This type of consultation requires the consultant to work with the consultee around problem solving around the client (Dimmitt, Carey, & Hatch, 2007; Caplan, 1970). In school settings, the client is the student or classroom, and the consultee (teacher) and the consultant work together to help the client. This model has been shown to be effective in reducing challenging behaviors for individual children (Hsieh, Hemmeter, McCollum, & Ostrosky, 2009; Powell, Dunlap, & Fox, 2006), as well as improving instruction at the classroom level (Conroy, Sutherland, Algina, Wilson, Martinez, & Whalon, 2015; Sutherland, Conroy, Vo, & Ladwig, 2015). Educational consultants range in their title but typically consist of school psychologists, organizational consultants, or behavioral specialists (Dougherty, 2009; Kratochwill & Bergan, 1990; Rosenfield, 2013). Regardless of the background of the consultant, the aim of consultee-centered consultation remains: helping teachers to problem-solve around their classroom or students.

Several consultee-centered consultation models exist and selecting the appropriate protocol is dependent on the desired outcome (Borko, 2004; Dimmitt et al., 2007). When considering the ultimate outcome of selecting effective instructional practices based on data, two established models are relevant: instructional (Rosenfield, 2004; 2013) and data-driven (Dimmitt et al., 2007; Slavin et al., 2011) consultation.

Instructional consultation. Consultee-centered consultation that focuses on students as the client is often referred to as instructional consultation (IC), as the basic goal is to enhance the instruction of, and ultimately achievement of, students (Hicks, 1999; Rosenfield, 2004; 2013). This model originated from an ecological framework emphasizing the interaction between students, instruction, and task demands (Gravois & Rosenfield, 2002; Rosenfield, 2004). By considering the interaction between these classroom components, IC allows teachers to focus on identifying the most appropriate instruction for a particular student at a specific moment (Bergan & Schnaps, 1983; Rosenfield, 2004). There are five distinct, linear stages to IC: (1) entry and contracting, (2) problem identification and analysis, (3) intervention design and planning, (4) intervention implementation and evaluation, and (5) closure (Rosenfield, 2013). The first stage, entry and contracting, involves introducing teachers to the concept of consultation and agreeing to the extent of the intervention (Bergan & Schnaps, 1983; Rosenfield, 2004). The next phase typically occurs two-to-three weeks into the consultative process when the teacher is ready to identify a problem in student-instruction match (Kaiser, Rosenfield, & Gravois, 2009). At this point, several methods are presented for consultants to help analyze the problem, all of which aim to facilitate collaborative-problem solving (CPS; Pugach & Johnson, 1988). CPS involves simultaneous and collaborative brainstorming in which the consultee and the consultant use each other's ideas to find a solution to a problem (Idol, Nevin, & Paolucci-Whitcomb, 1994; Pugach & Johnson, 1988). After identifying an appropriate instructional intervention, teachers implement the agreed-upon instruction in the classroom. This can be done on his or her own or with the consultant's help (Rosenfield, 2004). Consultants assist teachers in assessing the efficacy of the instruction, by revaluating any changes in students' learning (Hicks, 1999). Steps two through four can be repeated many times, and often IC is delivered throughout the academic year allowing multiple instructional foci (Bergan & Schnaps, 1983; Rosenfield, 2004). Toward the end of the academic year or consultation process, consultants are expected to provide closure-related activities (Rosenfield, 2013), rounding out the teacher's experience of consultation.

The efficacy of this model for improving teachers' instructional practice has been established for decades (Gravois & Rosenfield, 2006; Kaiser et al., 2009; Neef, Iwata, & Page, 1980; Ray, 2005; Ysseldyke & Christenson, 1993). More recently, empirical efforts shifted to explore the mechanism though which IC affects teachers' skill acquisition and generalization (Knotek, Rosenfield, Gravois, & Babinski, 2003). Through interviews with teachers who had participated in an IC intervention, Knotek, Rosenfield, Gravois, and Baninski (2003) established a theory of change in which teachers reported more efficacy and skills in evaluating data for problem-solving instructional challenges. Quantitative studies have corroborated this model (e.g., Kaiser et al., 2009), indicating IC has the potential to augment practices surrounding the use of data to inform their instruction by building their efficacy and skills in instructional data-use.

Data-driven consultation. While IC's focus on improving instruction has more recently occurred through data-use, other consultation models emphasize building teachers' capacity for using data first and generalizing these skills to all aspects of their practice (Dimmitt et al., 2007; Halverson et al., 2006). Consultation ascribing to this goal, often referred to as data-driven (DD) consultation (Dimmitt et al., 2007), originates from the SBM's call for data-driven accountability (Berndhardt, 2003; Halverson et al., 2006; Schmoker, 2003; Slavin et al., 2011). Sometimes coined the DD reform, this sub-movement emphasizes "obtaining timely, useful information, trying to understand the 'root causes' behind the numbers, and designing interventions targeted to the specific areas most likely to be inhibiting success" (Slavin et al., 2011, p. 4). Frequently this involves collecting, interpreting, and disseminating data with the purpose of informing classroom-based decisions (Berndhardt, 2003; Dimmitt et al., 2007; Slavin et al., 2011).

With these goals in mind, the Center for Data-Driven Reform in Education (CDDRE), funded by the U.S. Department of Education, published a DD consultation model based on empirically effective practices in prompting educational data-use (Slavin et al., 2011). In this and comparable models of DD consultation (e.g., Berndhardt, 2003; Dimmitt et al., 2007; Halverson et al., 2006), consultants aim to help school staff understand and act upon the results of benchmark assessments administrated throughout the year (Slavin et al., 2011). A series of five steps are promoted to achieve this goal: (1) data acquisition, (2) data reflection, (3) program design, (4) program alignment, and (5) formative feedback (Dimmitt et al., 2006; Slavin et al., 2011).

Given the centrality of data for these interventions, the first stage of any DD consultation involves reliable and valid (1) data acquisition. Many models use standardized assessments that establish benchmarks students are expected to meet (e.g., Slavin et al., 2011; May & Robinson, 2007), because data from this type of testing provides a basis for comparison across classrooms, schools, and districts (CCSSO, 2000b; Moon, Brighton, & Callahan, 2003). The results of these assessments are used to facilitate the second step, (2) data reflection, by comparing achievement to the established benchmarks. Various models take different approaches to helping consultees make these comparisons; for example, a review of four DD interventions funded by the RAND Corporation found that often reflection occurred at the consultant level (Marsh et al., 2005), whereas in Slavin and colleagues' (2011) intervention consultants walked through results with administrators and teachers to synthesize data. There is debate over whether consultees should participate in the data reflection process or whether systems accomplish this synthesis for them (Chick & Pierce, 2013; Pierce & Chick, 2011; Stecker et al., 2005). Chick and Pierce (2013) argue it is essential for school personnel to possess at least an elementary level of statistical literacy, described on page 52, so that effects of DD interventions can be generalized to other classroom practices.

Regardless of the method of reflection, DD models dictate the synthesized data should be used in the next stage of (3) program design (Halverson et al., 2006; Marsh et al., 2005). Similar to the third step of IC involving intervention design and planning (Rosenfield, 2013), this stage of DD consultation models typically aims to use data to make informed, practical decisions (Dimmitt et al., 2007; Marsh et al., 2005). The success of this stage relies on the structure of designed programs; they must provide concrete, actionable steps that are reasonable in nature for the consultee (Marsh et al., 2005). Once an intervention is proposed, DD consultants engage consultees in (4) program alignment, to ensure the plans fit within the larger goals or context of the school (Dimmitt et al., 2007; Halverson et al., 2006). Reviews of school-based interventions consistently demonstrate that programs which are misaligned with existing initiatives are unsuccessful (Franks et al., 2015; Murray et al., 2007; Nelson, Martella, & Marchand-Martella, 2002), highlighting the necessity of this often overlooked step. In this way, it is the DD consultant's job to ensure program design does not veer far from the intended target.

After the data have been collected, synthesized and used to design an intervention that fits within the larger academic framework, consultees are expected to implement the agreed upon program. Following this implementation, consultants provide (5) formative feedback around the success of the intervention (Halverson et al., 2006). This final step is essential in DD consultation, as it allows for evaluation of success of the first four steps (Halverson et al., 2006). "Feedback processes translate… outputs into useful information that guides subsequent input behaviors" (Halverson et al., 2006, p. 10), facilitating the continuation of DD decisions.

Although the majority of DD reform interventions have primarily targeted administers' use of data (Carlson, Borman, & Robinson, 2011; Slavin et al., 2011;

Wohlstettre, Datnow, & Park, 2008), there has been a recent push to promote this model at the teacher-level. Evaluations of this approach have demonstrated that teachers are able to acquire skills in synthesizing data and implementing associated interventions with PD (Ebbeler et al., 2016; Pierce et al., 2014), suggesting applying the steps as outlined in DD consultation has the potential to positively influence teachers' data-use practices.

The intersection of instructional and data-driven consultation: instructionally focused, data-driven consultation. In the current study, the principles of instructional and data-driven consultation were crossed to create an instructionally focused, data-driven consultation (IFDDC) model that aims to help teachers use data to inform their instructional practices. A visual representation of this crosswalk can be seen in Table 1. The first central component derived from the crossing of these two models involves (1) identifying the problem. This covers the second step of IC, problem identification and analysis (Rosenfield 2013), and the first and second steps of data-driven consultation, data acquisition and reflection, respectively (Dimmitt et al., 2007; Halverson et al., 2006; Slavin et al., 2011). Through this component, IFDDC consultants will help teachers to identify the 'problem' or data point on which they would like to intervene, thereby ascribing to the method of data synthesis in which the consultee participates (Chick & Pierce, 2013). Consequently, this step targets teachers' data interpretation skills through the first three of Wayman and Jimerson's (2014) competencies. To do this, Chick and Pierce's (2013) model for promoting statistical literacy was used.

The second component of IFDDC is (2) designing a 'solution' for this 'problem.' This covers the third step of IC, intervention design and planning (Rosenfield 2013), and the third and fourth steps of data-driven consultation, program design and alignment, respectively (Dimmitt, Carey, & Hatch, 2007; Halverson et al., 2006; Slavin et al., 2011). To do this, consultant-teacher dyads will engage in CPS revolving around evidence-based teaching strategies aligned with best-practices in kindergarten learning. In keeping with literature detailing the necessity of these plans to be concrete, this step will be completed by creating actionable plans from the 'solutions' designed during CPS. Through this component, IFDDC continues to target teachers' data interpretation skills through the fourth of Wayman and Jimerson's (2014) competencies by linking data to classroom practice.

The final IFDDC component surrounds providing (3) evaluative feedback. To do this, teachers must first implement the designed intervention, which allows for direct changes in their classroom practices. After, consultants scaffold teachers through evaluating the efficacy of this practice and provide additional feedback. This addresses part of IC's fourth step, intervention evaluation (Rosenfield, 2013), and the fifth data-driven consultation principle, formative feedback (Dimmitt et al., 2007; Halverson et al., 2006; Slavin et al., 2011). During this step, IFDDC allows for a positive experience of implementing an evidence-based practice and subsequent evaluation necessary for augmenting teachers' self-efficacy. Furthermore, the reflective questions around the practice have the potential to change teachers' beliefs. Together, these three steps combine best practices from IC and DD consultation models to create a consultation model that emphasizes the use of teachers' data for informing classroom practices. The implementation of IFDDC is explained further on page 81 and in Appendix 4.

Implementation fidelity as an essential element of IFDDC. In addition to the theoretical components detailed above, implementation fidelity is a necessary overarching element to ensure effective delivery of an individualized intervention, such as IFDDC, when conducted in an applied setting (Dane & Schneider, 1998; Hulleman & Cordray, 2009; Moncher & Prinz, 1991). The emphasis on fidelity in applied contexts, such as schools, originates from research demonstrating the significant impact of nonadherence/compliance on intervention effects in non-standardized conditions (Dhillon, Darrow, & Meyers, 2014; Hulleman & Cordray, 2009; Moncher & Prinz, 1991). Although there are many theoretical and empirical explanations of fidelity components, one of the more prominent operationalizations originates from Moncher and Prinz's (1991) review of treatment fidelity in outcome evaluations. In this article, the authors differentiate between several aspects of treatment fidelity that fall into two domains: content and process fidelity (Moncher & Prinz, 1991). Content fidelity captures program adherence, or "the extent to which specified program components were delivered as prescribed" (Dane & Schneider, 1998, p. 44). Measuring and assessing content fidelity allows for between-group comparisons of intervention-related experiences. Analysis of this construct is essential to establishing the amount of adherence to the intervention protocol in the treatment group, and to ensure the control group did not have similar intervention-related experiences to the treatment group (Chalmers et al., 1981; Dane & Schneider, 1998; Moncher & Prinz, 1991).

Process fidelity surrounds the occurrences or process of the content of the intervention (Moncher & Prinz, 1991). To do this, process fidelity data are only collected on participants who received the intervention. In this way, process fidelity measures provide information on within-group differences in intervention experiences. Together, content and process fidelity provide an overview of how the intervention was delivered and how that delivery may impact observed effects. As such, the effects of IFDDC will be viewed within the context of the content and process fidelity implemented.

Theory of Change

A theory of change (TOC) model describing the hypothesized effects of IFDDC on teachers' skills, beliefs, self-efficacy, and practices was developed from the literature reviewed above. This TOC model is described below and presented in Figure 1.

Given research linking improvements in students' learning when teachers' change their practice effectively (Richardson, 1990; Stecker et al., 2005), the central goal of the intervention is to influence teachers' practices. Specifically, IFDDC is theorized to have a positive impact on teachers' implementation of data-use practices through direct and indirect methods. The direct effect of the intervention on practices, displayed by the solid line between IFDDC and Data-Use Practices in Figure 1, is expected to result from the actionable plans teachers will implement after the second component, designing a solution.

IFDDC is hypothesized to produce indirect effects on teachers' data-use practices through augmentation of their skills, beliefs, and self-efficacy related to data-use, following research demonstrating the connection of these constructs to classroom practices (Brown et al, 2015; Roth, et al., 2008; Pierce & Chick, 2011; Tschannen-Moran et al., 1998; Tschannen-Moran & Woolfolk Hoy, 2001). These indirect effects are displayed by dashed lines passing through Skills, Improvement Belief, and Data-Use Efficacy in Figure 1 and described below.

The first and second components of IFDDC were designed to improve teachers' data-interpretation skills through building their statistical literacy (Chick & Pierce, 2013; Pierce et al., 2011) and data-related competencies (Wayman & Jimerson, 2014). As such, the intervention is hypothesized to have a direct effect on teachers' abilities to interpret data, as portrayed by the solid line between IFDDC and Skills in Figure 1. By improving these skills, IFDDC is expected to have an indirect effect on teachers' classroom practices related to data-use through building their capacity for effective data-use, as has been demonstrated by the generalized effects of statistical literacy (Ebbeler et al., 2016; Pierce et al., 2011).

The third component of the intervention in which consultants provide evaluative feedback is theorized to have direct effects on teachers' beliefs about and self-efficacy surrounding assessment data-use. Specifically, it is hypothesized that by participating in IFDDC, teachers will endorse beliefs relating to the potential for assessment data to inform their classroom practices, or Improvement Beliefs. This change is hypothesized given research demonstrating the impacts of reflective questioning on teachers' beliefs about their practices (DeBacker & Crowson, 2006; Romme & van Seggelen-Damen, 2015). By endorsing these beliefs related to the benefits of assessment for informing teaching, teachers will be more likely to implement classroom practices aligning with this belief (Bingimlas & Hanrahan, 2010; Pajares, 1992). Figure 1 highlights this hypothesized direct effect of IFDDC on teachers' Improvement Beliefs with a solid line, and the hypothesized indirect effect of IFDDC on teachers' practices through the dashed line from IFDDC to Data-Use Practices that passes through Improvement Belief.

The third component of the intervention is also expected to impact teachers' self-efficacy with regards to data-use. This effect is theorized from research highlighting the effects on self-efficacy after participating in an associated mastery experience (Ross, 1992; Woolfolk-Hoy & Burke-Spero, 2005). In IFDDC, the implementation of strategies designed in the second portion of the intervention is expected to result in a mastery experience given the evidence-based nature of the interventions which will be used. The intervention is expected to have an indirect effect on practices through increases in self-efficacy, as is consistent with theories linking self-efficacy to action (Bandura, 1993). Consequently, these effects are demonstrated in Figure 1 by the solid line connecting IFDDC to Data-Use Self-Efficacy (direct effect) and the dashed line between IFDDC and Data-Use Practices passing through Data-Use Self-Efficacy (indirect effect).

Aims and Hypotheses

Using this TOC, two central aims were derived for evaluating the current consultation intervention: to examine the efficacy of the intervention, and to explore the effects of implementation fidelity on the intervention. These aims, described in detail below, were examined using a randomized controlled trial in which 72 teachers from one school division in the Commonwealth of Virginia were assigned to either the treatment (n = 38) or business-as-usual (BAU; n = 34) condition. The participants and procedure for the study are detailed beginning on page 76.

Aim 1: To evaluate the efficacy of the intervention as proposed. The first aim addresses whether the intervention functioned as proposed. That is, questions housed within this aim evaluate if teachers who participated in the intervention demonstrate more accurate data interpretation skills and report higher levels of beliefs, self-efficacy, and practices relating to data-use as compared to the business-as-usual (BAU) teachers at posttest. This produces five research questions pertaining to changes in skills, beliefs, self-efficacy, practices, and the accuracy of the TOC model in describing the relationship among these constructs.

Research question 1. The first research question addresses the potential effects of IFDDC on data interpretation skills: (1) are treatment teachers more accurate in their data interpretation skills at the classroom- and individual student-level than BAU teachers at posttest after controlling for pretest responses? We hypothesized that treatment teachers would be more accurate in their data interpretation skills across levels than BAU teachers given the emphasis on building data-related capacities using frameworks that have demonstrated positive effects on teachers' skills in these areas (Ebbeler et al., 2016; Pierce & Chick, 2011).

Research question 2. The second research question concerns teachers' beliefs about assessment: (2) does the treatment have an effect on teachers' beliefs about assessment? Based on the focus of using assessment data to improve their teaching, we hypothesized that teachers receiving consultation would endorse higher rates of 'improvement of teaching' beliefs than BAU teachers. We also hypothesized there would be no discernable effect of the intervention on teachers' 'school accountability' or 'irrelevance' beliefs given the lack of emphasis of these constructs in the intervention.

Research question 3. The third question addresses the hypothesized effect of the intervention on teachers' self-efficacy with data-use: (3) do treatment teachers endorse higher rates of self-efficacy with using assessment data in their classroom as compared to BAU teachers at posttest after controlling for pretest responses? Given this demonstrated effect in other PD models (Allinder, 1994; Tschannen-Moran et al., 1998: Tschannen-Moran & Woolfolk Hoy, 2001), we hypothesized treatment teachers would endorse higher feelings of self-efficacy with using assessment data in their classroom as compared to BAU teachers.

Research question 4. The fourth question revolves around the hypothesized changes in teachers' assessment-related practices: (4) do treatment teachers endorse using more practices related to using assessment to inform their practice than BAU teachers at posttest after controlling for their pretest responses? Given previous studies demonstrating consultation surrounding data-use has the potential to influence teachers' practices (e.g., Dimmitt et al. 2007; Halverson et al., 2006; Slavin et al., 2011), we hypothesized that teachers participating in IFDDC would endorse more positive statements around using assessment data in the classroom than BAU teachers after the intervention.

Research question 5. The fifth and final research question of Aim 1 pertains to the hypothesized indirect effects of the intervention on data-use practices through skills, beliefs, and self-efficacy: (5) Do skills, beliefs, and self-efficacy uniquely mediate changes in the relationship between the intervention and classroom practices? Given the hypothesized direct effects and research demonstrating the connection of each of these constructs to classroom practices (Allinder, 1994; Bandura, 1988; Tschannen-Moran et al., 1998), we hypothesized skills, beliefs, and self-efficacy would uniquely mediate the relationship between the intervention and changes in classroom practices.

Together, these first five research questions coalesce to examine if the effects of the intervention functioned as proposed.

Aim 2: To explore the effects of implementation fidelity on the intervention. The second aim of the study, to explore the effects of implementation fidelity on the intervention, was generated based on evidence that fidelity of school-based intervention can significantly impact observed effects (Dhillon, Darrow, & Meyers, 2014; Hulleman & Cordray, 2009; Moncher & Prinz, 1991). To ensure accurate interpretation, intervention fidelity should be explored between- and within-groups (Moncher & Prinz, 1991) for a complete understanding of the potential influences of fidelity. As such, the first two questions of this aim concern the differences in fidelity across groups, while the latter two explore the experiences of the treatment group in greater detail.

Research question 6. The sixth research question of the study addresses the content fidelity of the intervention to ensure the groups had different experiences related to data-focused consultation. Content fidelity of the current study was operationalized as participating in a consultation session, detailed on page 93, and led to the following question: (6) were there differences in the percentage of teachers who received data-use consultation sessions by treatment condition? Due to the consultation session held as a part of IFDDC and research demonstrating few schools implement data-focused initiatives (Kerr et al., 2006; Marsh et al., 2005), we hypothesized that the treatment group would participate in consultation sessions around their classroom data at a higher rate than the BAU group.

Research question 7. The seventh research question pertains to the relative strength of the intervention after considering the content fidelity established in the previous research question: (7) what is the achieved relative strength of the intervention after considering between-group content fidelity? We hypothesized that achieved relative strength of the intervention would be higher than that of the actual strength of the intervention given the documented reduction of effects that occurs during school-based intervention (Dhillon et al., 2014; Greenberg, Domitrovich, Graczyk, & Zins, 2003; Hulleman & Cordray, 2009).

Research question 8. The eighth research question explores the experiences of the treatment group: (8) were there differences in intervention-related experiences within the treatment group? Again, given evidence indicating intervention fidelity is challenging to uphold in educational settings (Dhillon et al., 2014; Greenberg et al., 2003; Hulleman & Cordray, 2009), we hypothesized there would be differences in intervention-related experiences within the treatment group.

Research question 9. The final research question addresses treatment teachers' satisfaction with the intervention: (9) were treatment teachers satisfied with the consultation? We hypothesized treatment teachers would be satisfied with the consultation given research demonstrating positive attitudes towards similar PD models (Garet, Porter, Desimone, Birman, & Yon, 2001; Nir & Bogler, 2008; Tschannen-Moran & McMaster, 2009).

CHAPTER III

Method

Participants

The present study drew participants from the Virginia Kindergarten Readiness Program (VKRP), an initiative that involved 533 kindergarten teachers across the state of Virginia administering math, literacy, and social skills KEAs to all children enrolled in their classroom. Through a joint effort between the Virginia Department of Education and the Center for the Advanced Study of Teaching and Learning, 21 school divisions were recruited to participate. All kindergarten teachers within a participating division were expected to administer KEAs during the VKRP administration window which occurred between weeks four and six of the 2015-16 academic year.

For the current intervention, teachers from one large division totaling 116 classrooms were selected to participate. The division was selected based on a multi-level power analysis conducted for the intervention that suggested targeting a sample size of 122 (for more details on this power analysis, see Appendix 1). Only two divisions participating in VKRP met this sample size, and the division ultimately selected to participate in the current intervention demonstrated interest in providing professional development to teachers (hence forth described as ‘selected division’).

Recruiting efforts within the selected division occurred through direct contact during VKRP trainings and outreach by the division’s lead instructional specialist, who served as the liaison throughout VKRP. In total, 73 teachers across 21 schools were consented to participate in the current intervention. Thirty-eight teachers were randomly assigned to the treatment group and 35 were assigned to a business-as-usual (BAU) group. Over the course of the intervention, one teacher in the BAU group withdrew from the intervention due to going on maternity leave, leaving a total sample of 72 teachers ($N = 72$, $n^{TX} = 38$, $n^{BAU} = 34$).

The sample of participating teachers was mostly white (86%) with the remaining teachers identifying as African American (14%). Two (3%) of the teachers were male. All teachers had at least a Bachelor's degree, and 71 percent ($n = 48$) had a Master's degree or higher. The average age of teachers was 42 ($SD = 12.44$), and the average number of years of experience in K-12 schools was 13 ($SD =10.71$). There were no significant differences in demographic data between groups. In addition, *t*-tests comparing teachers participating in the current study to non-participating teachers in the selected division revealed no significant differences in demographic data. Demographics for the selected division, current study, treatment group, and BAU group are further illustrated in Table 2.

In comparison to the National Center for Education Information's (Feistritzer, 2011) report on elementary teachers' demographics, the current sample was on par with the national racial spread (84% white), age ($\bar{X} = 40$ years; SD = unlisted), and years of experience of teachers ($\bar{X} =$ 10 to 14, reported categorically). The most notable difference between the national teacher demographics and the current sample concerns gender- while only three percent of the current sample's participants were male, 16 percent of the national sample were male (Feistritzer, 2011).

Procedure

A visual representation of the procedure for the current study can be seen in Table 3. All kindergarten teachers in the selected division were required to attend the general VKRP training, which occurred during one of the division's professional development days in the summer of 2015. The training focused on implementation of the VKRP readiness assessments (described below and in Appendix 2) but also included a brief introduction on how to interpret the VKRP reports, a walk-through of instructional resources available through VKRP, and an explanation of how to contact VKRP through the hotline or online chat function. Additionally, teachers in the selected division received information about the current intervention, including the consultation timeline and consent procedures. Directly following the training, all teachers were provided with access to the VKRP website through a secure login. On the website, teachers were able to practice the assessments, enter assessment data during their VKRP administration window, access resources such as instructional strategies, and converse with a member of the VKRP team through the online chat function. In addition, teachers had access to the VKRP support hotline beginning after the training and lasting through the 2015-16 academic year. Although the majority of contact teachers in the selected division made with VKRP through the website or hotline surrounded technical problems, issues with administration, and/or difficulty accessing reports (95 percent), data about each caller and reason for the call were collected throughout the duration of the intervention and are included in the fidelity results when applicable to teachers enrolled in the study.

Directly after the training, the demographic survey was sent via email to all teachers in the division as a part of the larger VKRP initiative. The demographic survey was completed using Qualtrics© survey software (Qualtrics, 2016). Eighty percent (n =

96) of teachers in the selected division completed the demographic survey. Of the 72 consented teachers, 83 percent completed the demographic survey (completion rate: n^{TX} = 34; n^{BAU} = 26). All survey completion data can be seen in Table 2 for the selected division, treatment, and BAU teachers.

Approximately one week after the training, all 116 teachers in the selected division received an email detailing consent procedures for the current study. Over the course of four weeks, 73 teachers provided consent to participate. These consented teachers were then randomly assigned to the treatment or BAU group with the following procedure: consented teachers were placed in alphabetical order by school, and each school's list was compared to the output of the random binary generator. Names with a corresponding '1' were assigned to the treatment group whereas teachers with a corresponding '0' were assigned to the BAU group. Block randomization by school was used to ensure adequate variability in assignment at the school level. This produced 38 teachers assigned to the Treatment group and 35 assigned to the BAU group. As aforementioned, one teacher assigned to the BAU group withdrew during the course of the study, leaving a total sample size of 72 (n^{TX} = 38, n^{BAU} = 34). Teachers were not alerted to their intervention condition until after VKRP KEAs administration.

After randomization, all consented teachers were asked to complete the pretest measures. Teachers were contacted through an email which included a link to the survey. Surveys were created, distributed and completed using Qualtrics© survey software (Qualtrics, 2016). This survey was open to teachers during the first four weeks of school before the beginning of their division's VKRP administration window. Sixty-seven

(94%) of the consented teachers completed the pretest survey (pretest completion rate: $n^{TX} = 38$; $n^{BAU} = 29$). Completion rates of the pretest survey can be seen in Table 2.

Following the pretest survey, all teachers completed VKRP as usual; between the fourth and sixth weeks of school, they administered the *Happy Birthday: Early Mathematics Assessment System* (Birthday Party: EMAS; Ginsburg & Pappas, 2016), the *Phonological Awareness and Literacy Screening-Kindergarten* (PALS-K; Invernizzi, Swank, Juel, & Meier, 2003), and *Child Behavior Rating Scale* (CBRS; Bronson, Goodson, Layzer, & Love, 1990) to all children in their classroom. The Birthday Party: EMAS represents the math domain, the PALS-K represents the literacy domain, and the CBRS represents both the Social Skills and Self-Regulation domains. More information on the structure, reliability, and validity of the KEAs included as a part of VKRP can be found in Appendix 2. Given some unforeseen technical difficulties, not all teachers completed VKRP administration in the timeframe allotted; completion rates of the selected division were as follows: *Birthday Party: EMAS*- 53 percent, *PALS-K*- 100 percent, *CBRS*- 69 percent. Completion rates of the 72 consented teachers were as follows *Birthday Party: EMAS*- 61 percent, *PALS-K*- 100 percent, *CBRS*- 72 percent. As soon as a teacher completed an administration of one of the assessments for a child, the teacher could access that child's data through the VKRP report. If a teacher completed an assessment for all children in his/her classroom, the teacher had access to the classroom-level data for that assessment. For completed administrations, the VKRP report detailed individual- and classroom-level data for the four domains (math, literacy, social skills, and self-regulation), allowing for comparison across math, literacy, and social-emotional readiness at the individual- and classroom-level. An example VKRP report can be seen in Appendix 3.

Once the VKRP administration window closed after the sixth week of school, teachers assigned to the treatment condition received IFDDC as described below. After the six week intervention period, all consented teachers received the posttest survey via emails. The survey was again completed using Qualtrics© survey software (Qualtrics, 2016). Sixty-seven (94%) of the consented teachers completed the posttest survey (posttest completion rate: n^{TX} = 37 n^{BAU} = 30). Completion rates of the posttest survey can be seen in Table 2; all teachers provided data for at least one of the pre- or post-test survey, and the majority (83%) provided data at both time points.

Instructionally focused, data-driven consultation. The author of this dissertation (referred to as Consultant 1) and a fellow masters-level graduate student pursuing a doctoral degree in clinical psychology at the University of Virginia (referred to as Consultant 2) served as the consultants. Consultant 1 designed the protocol based on a crosswalk and synthesis of best practices in instructional (Rosenfield, 2013; Rosenfield et al., 2014) and data-driven (Halverson et al., 2006; Slavin et al., 2011) consultation models as shown in Table 1. This process revealed the following four steps for IFDDC: (1) entry and explanation; (2) data review and problem identification; (3) intervention design and planning; and (4) intervention evaluation/feedback. This protocol underwent a review/iteration process with the principal investigators (PIs) of VKRP before being finalized. Consultant 1 then trained Consultant 2 on the intervention protocol and the two met weekly during the four-week intervention period to discuss progress. The PIs of VKRP provided supervision to the consultants when necessary.

A VKRP research assistant emailed all treatment teachers directly following the VKRP window. The email described the nature of the consultation and requested preferred meeting times. Teachers were then assigned to consultants based on scheduling availability. Twenty-eight teachers were assigned to Consultant 1 and ten teachers were assigned to Consultant 2. Eight teachers were unable to be scheduled for an in-person meeting with either consultant. These eight teachers were offered a phone consultation session as an alternative, to which three agreed. Consequently, five teachers did not receive consultation sessions. In total, Consultant 1 completed 23 in-person and three phone consultation sessions, where Consultant 2 completed seven in-person sessions.

Although a summary of the instructionally focused, data-driven consultation (IFDDC) procedure is described below, a more detailed explanation of the intervention can be found in the VKRP Consultation Manual provided in Appendix 4.

Preparation. The preparation component was implemented to meet the first principle of IFDDC: entry and explanation. After a teacher was assigned to a consultant, the consultant emailed the teacher to introduce herself and the general consultation process. This initial email also contained a confirmation of the time and location for their consultation meeting and requested that the teacher review his/her VKRP report prior to the meeting.

Consultation meeting. The consultation meeting housed the second and third components of IFDCC: (2) data review and problem identification and (3) intervention design and planning. The meeting began with a review of the teacher's data in which the consultant guided the teacher with questions such as "What stands out most to you?" and "In which domain were your students most successful and which was most challenging for them?" From there, consultants engaged in a conversation about points of data that were most interesting to the teacher with questions such as "What do you think that score means?" and "Is that consistent with what you see in the classroom?" This type of data-review process is consistent with those used in other data-driven consultation models (Halverson et al., 2006; Slavin et al., 2011) and was meant to model/elicit the first and third of Wayman and Jimerson's (2014) data-related competencies: (1) asking the right questions and (3) analyzing and interpreting the data.

After teachers and consultants worked together to review the data and isolate a point on which to intervene the pair engaged in collaborative problem-solving (Pugach & Johnson, 1988; Idol et al., 1994). This involves simultaneous and collaborative brainstorming in which the consultee and the consultant use each other's ideas to find a solution to a problem (Pugach & Johnson, 1988; Idol et al., 1994). The "problem" in IFDDC surrounded the data point that was previously identified and the "solution" focused on identifying instructional strategies based on that data point, which aimed to meet the fourth of Wayman and Jimerson's data-related competencies: (4) linking data to classroom practice. To do this, pairs discussed the ideal "solution" and worked backwards to decide on realistic approaches to achieve this solution. During this portion of the meeting consultants provided information about relevant strategies offered on the VKRP and/or PALS website(s) and reviewed them with the teacher. The agreed-upon solution was not limited to these strategies; however, if no VKRP or PALS strategy was suitable to the teacher, consultants worked with teachers to design a new instructional strategy. The entirety of this process, starting with data problem identification and ending with instructional strategy selection, is referred to as an instructional cycle.

After completing an instructional cycle, consultants worked with teachers to complete the Action Plan document, presented in Appendix 5. This document intended to provide a summary of the instructional cycle by describing the identified data point(s) for intervention and selected instructional strategy. Furthermore, this document included a concrete plan for intervention, in which teachers specified the date, time, and target child/group for the instructional strategy described. By completing this document together, teachers and consultants completed the third component of IFDDC: (3) intervention design and planning.

This process of completing an instructional cycle and Action Plan was repeated up to three times during a consultation meeting. The number of cycles per consultation meeting varied depending on the teachers' ability to identify patterns in the data that suggest the need for intervention and the time left after the first completed Action Plan.

At the end of the consultation meeting, the consultant explained the upcoming communications that the teacher could expect, including a follow-up prompt and posttest survey. The consultant also reminded teachers of the VKRP consultation email address, indicating that he or she should reach out if any questions arose about the strategies or the consultation more generally.

Follow-up prompts. The day following the date specified on the Action Plan, teachers received a follow-up prompt via email inquiring about the use of their specific strategies. The prompts were individualized to each teacher, asking him or her to describe implementation and efficacy of each of the chosen strategies. A sample follow-up prompt can be seen in Appendix 6.

Measures

Covariates.

Teacher demographics. The following data about teachers were collected through the VKRP demographic survey: age, gender (male = 0, female = 1), race (white [reference group], black), years of teaching experience in kindergarten, years of teaching experience in total, highest degree of education (Bachelor's = 0, Master's = 1), and if the teacher worked in his/her current school during the previous academic year.

Perception of organizational culture around data-use. To account for the effects of school climate on teachers' beliefs and self, efficacy, data about teachers' perceptions of their school's culture towards assessment and data-use were gathered during the pretest data collection through an adapted version of a teacher survey developed for the Learning from Leadership project (Leithwood, Seashore Louis, Anderson & Wahlstrom, 2004). The original survey was created from existing instruments with the addition of new items to evaluate the relationship between principal-teacher relationships and teachers' instructional practices (Leithwood et al., 2004). Various studies utilized the original version of this teacher survey, which factors into five constructs (Seashore Louis, Dretze, & Wahlstrom, 2010; Wahlstrom & Seahore Louis, 2008): focused instruction, teacher's professional community, shared leadership, instructional leadership, and trust in principal. Of these factors, instructional leadership, or a teacher's perception of his/her leadership's attitudes towards instruction, is relevant for gauging teachers' perception of their school's climate towards data-use and therefore was the sole construct assessed in this study. The original items, which focused on instruction as a whole, were adapted (with the permission of the instrument's authors) to focus on instructional practices related specifically to assessment. For example, the item "my school administrator clearly defines standards for instructional practices" was altered to read "my school administrator clearly defines standards for using assessment to inform instructional practices." One item -- "How often in this school year has your school administrator buffered teachers from distractions to their instruction" -- was deleted because its intent was not relevant to data-use. A complete list of the original items and adaptations for the current study can be seen in Table 4.

All items were placed on a six-point Likert scale. One item asked teachers to report their agreement with a statement about their school administration's standards for data-use (1- Strongly Disagree to 6- Strongly Agree). The remaining five items asked teachers to report the frequency with which actions pertaining to data-use occurred in their school (continuous; 1- Never to 6- Very Often). An analysis of internal consistency demonstrated a Cronbach's alpha of 0.873, suggesting the adapted items showed similarly good fit to a one-factor model as the original items had in a previous analysis ($\alpha = 0.82$ as reported in Seashore Louis et al., 2004). Therefore, the six items were summed to create a Perception of Data-Use Climate composite, which was used as a covariate throughout analyses to account for school-level differences in expectations pertaining to data-use.

Outcome measures. All outcome measures were gathered during the pretest and posttest surveys and are presented in Appendix 7.

Data interpretation skills. To analyze their proficiency with interpreting data, teachers were asked to review a hypothetical data report and select relevant instructional strategies. Teachers were provided with 11 options for instructional strategies, all of which were drawn from the VKRP or PALS websites. Three strategies pertained to math instruction (one for Numeracy, one for Computation, and one for Patterns), three were relevant to literacy instruction, three were aimed at social skills instruction, and the remaining two focused on self-regulation instruction. Instructional strategies were not listed under a particular domain (i.e., “Math Strategy- Geometry”), but rather described the instructional plan (i.e., “Finding Patterns in the Environment- Encourage children to look for and draw patterns in the classroom and/or outside.”).

The hypothetical report was consistent with the VKRP reports; classroom-level data reported as percent that met the benchmark was displayed graphically across math, literacy, social skills, and self-regulation scores. On the following page, students’ individual scores for the four domains were displayed in a table; that is, students were listed in the first column on the left and the subsequent four columns listed the domains. Each row contained one student’s performance across math, literacy, social skills, and self-regulation, respectively. Next the teacher was provided with a similar chart detailing all students’ performance across the four math areas tests: Numeracy, Patterning, Computation, and Geometry. The final section of the report contained the individual math, social skills, and self-regulation scores for one child.

Teachers were first asked to select three instructional strategies that best served the needs of the whole classroom based on the data. In the hypothetical report provided, the classroom-level data demonstrated the following percent of students who met the corresponding benchmarks for readiness: 40 percent in math, 75 percent in literacy, 80 percent in social skills, and 50 percent in self-regulation. The next page demonstrated the same percentages, but separated out children’s individual scores across the four domains.

From these two pages, the teachers were expected to surmise that this classroom would benefit from instructional strategies targeting math or self-regulation. The third table revealed the classroom-level math scores as follows: 85 percent in Numeracy, 25 percent in Patterning, 70 percent in Computation, and 50 percent in Geometry. From here, teachers were expected to further specify the math strategies to be aimed at Patterning or Geometry (of which only Patterning was provided as an option). This allowed for three correct responses: the instructional strategy pertaining to Patterning and two strategies targeting self-regulation. The number of these three strategies that a teacher selected was summed and divided by three to create a ratio detailing his/her accuracy in selecting instructional strategies based on the classroom-level data, or the Classroom-Level Data Interpretation Ratio.

Teachers were then asked to select four instructional strategies that best served the needs of an individual student. The same 11 strategies were provided. This student's individual math, social-emotional, and self-regulation scores were detailed after the classroom-level data described above. His overall math data was described as "below meeting the benchmark for readiness" and the specific areas tested read as follows: 75 percent in Numeracy, 25 percent in Patterning, 25 percent in Computation, and 75 percent in Geometry. From this, the teacher was expected to surmise that strategies targeting Computation or Patterning would be beneficial. His social skills data was reported to be above the benchmark, but his self-regulation was reported to be lower than expected for an incoming kindergartener. The item-level data pertaining to his social skills and self-regulation report confirmed this. As a result, teachers were expected to select the two self-regulation strategies as well, allowing for four correct responses. The number of these four strategies that a teacher selected was summed and divided by four to create a ratio detailing his/her accuracy in selecting instructional strategies based on the individual student-level data, or the Data Interpretation Ratio-Student.

To determine the reliability of this measure, seven expert reviewers were asked to complete the measure after posttest data collection and a Cohen's kappa ($\varkappa$) was calculated based on agreement with the author. Using the Landis and Koch (1977) guidelines for interpreting the strength of the kappa statistic which are widely supported in applied literature (e.g., Pett, 2015; Sim & Wright, 2005; Suen & Ary, 2014), reviewers scores ranged between the 'substantial' to 'almost perfect' strengths (average $\varkappa = 0.87$; range: 0.72 - 1.00). Reviewers agreed slightly more with the author on strategy selection at the individual student-level, with all scores falling in the 'almost perfect' range (average $\varkappa = 0.91$; range: 0.82 - 1.00).

Teachers' beliefs about assessment. Teachers' beliefs about assessment were ascertained using the short form of the *Conceptions of Assessment, Third Edition* (COA-III, Brown, 2006). For this measure, teachers respond to 27 statements about the beliefs about or reasons for the use of assessment using a six-point Likert scale (1- Strongly Disagree to 6- Strongly Agree). These statements concentrate on beliefs related to the following uses of assessment: Improvement, Student Accountability, School Accountability, and Irrelevance. The 11 items of the Improvement factor surround beliefs about using data to inform practice and/or learning (Brown, 2004a). The three items associated with the Student Accountability factor pertain to the belief that assessments are useful for holding students accountable (Brown, 2004a). School Accountability includes three items about the use of assessment for holding schools liable (Brown, 2004a). The final 11 items describe assessment as irrelevant, and thus correspond to the Irrelevance belief (Brown, 2004a).

As is consistent with analyses of the full 51-item COA-III, the short form demonstrates good fit onto a four-factor model (RMSEA =0.058, TLI =0.938; Brown, 2004a). In previous evaluations, the Cronbach alpha scale reliabilities were acceptable for the short scale: Improvement α =0.85, Student Accountability α = 0.66, School Accountability α = 0.90 and Irrelevance α = 0.76 (Brown, 2004a; Brown et al., 2015; Brown, Lake, Matters, 2011). The factors are interrelated, with the Improvement-Irrelevance, Improvement- School Accountability, School Accountability-Student Accountability relationships demonstrating moderate correlations (r[491] = 0.68, 0.56, and 0.58, respectively; Brown, 2004a). All other correlations between factors were reported as Low to Negligible.

Analyses of internal consistency using the pretest data from the current study demonstrated two of the four previously established factors had adequate internal consistency: Improvement α = 0.823 and Irrelevance α = 0.747. These two composites were created as described in Brown (2004a) by aggregating the relevant items. The original three items of the School Accountability factor demonstrated an alpha of 0.591, but this was increased to 0.792 when the item “Assessment provides information on how well schools are doing” was dropped. Consequently, a Revised School Accountability factor was created from the two remaining items and was used throughout the analyses. The Student Accountability factor demonstrated an alpha of 0.276, and this consistency was not improved with any item removals. As a result, the Student Accountability factor was dropped from further analyses. Correlations among the pretest factors trended towards those reported in Brown (2004a), but were generally weaker; the Improvement-School Accountability relationship demonstrated a nearly moderate relationship (r[67] = 0.46, p <.000), whereas the Improvement-Irrelevance and School Accountability-Irrelevance relationships were notably weaker (r[67] = -0.22, p =.08 and r[67] = 0.20, p =.09, respectively)

Posttest data demonstrated similar internal consistencies and correlations among factors. The Improvement (α = 0.847) and Irrelevance (α = 0.817) factors demonstrated good consistency, whereas the items pertaining to School Accountability and Student Accountability factors did not hang together. When the item "Assessment provides information on how well schools are doing" was again dropped from the School Accountability factor, however, the alpha rose to an acceptable level (α = 0.855). As such, the revised School Accountability factor was used throughout analyses.

Data-use self-efficacy and practices. Teachers were asked to report their agreement towards seven statements pertaining to their data-use practices. These items were created for the current study through a review of other evaluations' analysis of crucial teachers' practices in using data (e.g., Halverson et al., 2006; Stecker et al., 2005; Wayman & Jimerson, 2014). Items were placed on a six-point Likert scale (1- Strongly Disagree to 6- Strongly Agree). An exploratory factor analysis (EFA) with pretest data, which included analysis of the Eigenvalues and corresponding Scree plot in compliance with best practices in EFAs (Cudeck, 2000; Gorsuch, 1988), demonstrated good fit onto a two-factor model (χ^2 (8, N = 72) = 82.805, p < .05). In analyzing these two factors theoretically, the first factor included two items that pertain to teachers' report of their efficacy with data-use ("I feel confident in interpreting individual-level assessment data [e.g., a student's answers to a math test] to inform my instruction" and "I feel confident in interpreting classroom-level assessment data (e.g., scores across students' math tests) to inform my instruction"). Consequently, this factor will be referred to as the Data-Use Self-Efficacy from now on. The second factor contained items surrounding actionable practices in the classroom (e.g., "I use assessments to inform classroom practice"), and is thus referred to as Data-Use Practices. One item ("I use assessments to inform grouping children according to needs") loaded onto both factors, but demonstrated better fit statistically and theoretically with the second factor and thus was aggregated into the Data-Use Practices composite. Both composites demonstrated acceptable reliability at pretest (Data-use Self-Efficacy $\alpha = 0.834$; Data-Use Practices $\alpha = 0.844$) and posttest (Data-Use Self-Efficacy $\alpha = 0.811$; Data-Use Practices $\alpha = 0.835$). Factor loadings of these items can be seen in Table 5.

Fidelity measures. To address the sixth, seventh, and eighth research questions, data on fidelity were collected beginning at pretest and ending after posttest survey administration. Following Moncher and Prinz's (1991) model, data were collected on content and process fidelity. The manner through which these constructs were operationalized and relevant data were collected are described below.

Content fidelity: Between-group measures. To ensure the intervention group adhered to different intervention-related experiences compared to the BAU group, content fidelity was operationalized as participating in a consultation session around VKRP data, because the essence of the intervention was prescribed to occur during this meeting. As such, IFDDC adherence was categorized as a binary variable; teachers who did not receive consultation with a VKRP consultant were assigned 0, and teachers who received consultation with a VKRP consultant were assigned 1. Data on adherence were collected through two avenues: consultants' and teachers' report. After each consultation session, consultants were expected to record the occurrence under a teacher's ID on the teacher's consultation log, which can be seen in Appendix 4 (consultation session = 1). At the end of the intervention period, any teacher who was not recorded as receiving a consultation session with a VKRP consultant was specified as not receiving a session (consultation session = 0). Additionally, all teachers were asked to specify if the statement "I received one-on-one consultation around my VKRP data" was true (1) or false (1) on the posttest survey, thereby providing teachers' report of intervention adherence. Together, consultants' and teachers' reports of program adherence provide a multi-rater estimate of content fidelity of IFDDC.

Process fidelity: Within-group measures. To understand the process fidelity, or treatment occurrences (Moncher & Prinz, 1991), of the intervention, data were only collected on participants who received the intervention. The process of the content (consultation session) was operationalized as the number of instructional domains and subdomains covered (domains: math, literacy, social skills, self-regulation; subdomains: geometry, patterning, numeracy, computation, beginning sounds, blending, rhyme, sound-to-letter segmentation, alphabet recognition, comprehension, word concept, letter sounds, print knowledge, writing, spelling), instructional level covered (classroom, small-group, individual), and the specific strategy recommended (see Appendix 8 for strategy options). This information was ascertained through the consultation log, presented in Appendix 4.

Satisfaction measure. Information about treatment teachers' attitudes towards the consultation was collected with an 11-item survey created for this project. Items were rated on a 6-point Likert scale (continuous; 1- Strongly Disagree to 6- Strongly Agree) and indicate satisfaction with components of the intervention, such as "It was useful to cover the individual children's data during consultation" and "It was easy to work with the consultant when brainstorming strategies." A satisfaction scale was created using the average of the 11 items. This scale demonstrated good reliability ($\alpha = 0.995$). All items related to satisfaction with the intervention can be seen in Appendix 10.

Data Analytic Approach

Data analysis occurred in three phases. The first phase surrounded data preparation and determining appropriate models for later phases. The second phase included intent-to-treat analyses, and the third phase involved in-depth analysis of the intervention through between- and within-group fidelity analyses. All analyses were completed using SPSS Version 23 (IBM Corp, 2015) with the exception of the mediation analysis (described below), which was completed with Mplus Version 6.11 (Muthén & Muthén, 1998-2011).

Phase 1: Data preparation.

Missing data. Although efforts were made to collect all data, 14 percent of data was missing at the end of posttest data collection. To account for this, Restricted Maximum Likelihood (REML) imputation was used. This method was selected over similar methods (e.g., FIML), because it is most preferred for estimating variances in small sample sizes (Peugh, 2000; Raudenbush & Bryk, 2002). Twenty imputations across

200 iterations were used to create the imputed data set from which all analyses were conducted unless otherwise stated.

***Unconditional models*.** To determine if multilevel modeling was necessary, unconditional models were run to examine how much responsive variable variation existed at the school-level (Peugh, 2010). For this intervention, level 1 was the teacher and level 2 was the school. The intraclass correlation (ICC) was calculated using unconditional random effect ANOVAs with the factors of the five outcome measures. ICCs for Data Interpretation Ratio- Classroom, Data Interpretation Ratio-Student, Data-Use Self-Efficacy, Data-Use Practices, Improvement Beliefs variables were as follows: 0.004, 0.125, 0.000, 0.015, and 0.000, respectively. Following Peugh (2010)'s guidelines, multilevel analyses were not used as most of the ICCs were close to zero (ICC < .05), indicating limited response variation at the school-level. As a result, all analyses were conducted without nesting, but controlled for school as a fixed effect with multiple dummy codes.

Phase 2: Impact analyses. To address the first research aim concerning the impacts of the intervention, several analyses were conducted with intent-to-treat (ITT) models. ITT models include every subject who was randomized into a condition regardless of what happens after randomization (Fisher et al., 1990; Gupta, 2011). ITT models are considered essential in evaluating randomized trials, because they ignore variance due to dosage and/or noncompliance, a common occurrence in applied work (Fisher et al., 1990; Greenberg et al., 2003, 2005; Gupta, 2011). Because the data were imputed, pooled statistics are reported for all ITT analyses.

Direct effects of intervention on outcome variables. To address the first, second, third, and fourth research questions, three ITT analyses were calculated on the seven outcomes variables (Data Interpretation Skills- Classroom, Data Interpretation Skills- Individual Student, Improvement Belief, School Accountability Belief, Irrelevance Belief, Data-Use Self-Efficacy and Data-Use Practices). The first analysis calculated the variance of each outcome variable explained by the intervention condition using *t* statistics from a linear regression model. *t* statistics are most often interpreted using their associated *p* values, which express the strength of evidence against the null hypothesis by accounting for the sample size (Moore & McCabe, 1993; Cohen, 1995). Evaluations with small samples have a lowered positive predictive value, or "probability of finding true effects" (Button et al., 2013, p. 366; Cohen, 1995) using *p* values, however. As such, best practices in statistical interpretation of underpowered samples denote calculating the local effect sizes of outcomes as essential (Button et al., 2013, Cohen, 1995; Kraemer & Kupfer, 2006). Local effect sizes describe the magnitude of prediction for outcome variables, and the Cohen's *d* statistic was selected for the current study's local effect calculation given that the outcome comparisons were between two group means (Cohen, 1977, 1995). Finally, the actual power was calculated to provide the probability of correctly rejecting the hypothesis test for each of the outcome variables (Cohen, 1995; Lenth, 2007) as recommended for underpowered studies (Lenth, 2007).

Linear regression models. In the linear regression models, intervention condition was the independent variable (IV) and the outcome was the dependent variable (DV). School, teacher demographics (age, race, education, years of teaching experience in Kindergarten, total years of teaching experience), teachers' perception of their school's climate towards data-use, and DV pretest responses were included as covariates across all seven models. The amount of variance in the outcome explained by the intervention condition is reported using the unstandardized pooled beta values. Interpretation statistics of the beta values are reported with *t* statistics and associated *p* values.

Local effect size calculation. The Cohen's *d* statistic was calculated for each of the seven outcome variables to provide an interpretation of the magnitude of the effects. The Cohen's *d* statistic for each outcome variable was calculated using the formula $d = (\bar{x}_{TRE} - \bar{x}_{BAU})/s$, where $\bar{x}_{TRE}$ is the mean for the Treatment group, $\bar{x}_{BAU}$ is the mean for the BAU group, and s is the standard deviation of $(\bar{x}_{TRE} + \bar{x}_{BAU})$. Results are interpreted using the guidelines put forth by Cohen (1977, 1995) specifying that $d \geq 0.8$ should be considered a large effect, $0.5 \geq d > 0.8$ should be considered a moderate effect, and $0.2 \geq d > 0.5$ should be considered a small effect.

Power calculation. The actual power of each outcome variable was calculated posthoc with an alpha of 0.05 as recommended (Lenth, 2007). Power was calculated using the derivation described by Lenth (2007) where the *t* statistic (and consequently the *p* value) is compared to the degrees of freedom in the model. In social sciences research 0.8 is the accepted power.

Indirect effects of intervention on data-use practices: Mediation analysis. To address the fifth research question, a mediation model assessing the effects of skills, beliefs, and self- efficacy on the relationship between the intervention and data-use practices was created. The IV was identified as the Treatment Condition variable, the DV was determined to be the Data-Use Practices variable, and the four mediating variables (MVs) were established as Data Interpretation Skills- Classroom, Data Interpretation

Skills- Student, Improvement Belief, and Data-Use Self-Efficacy variables. Teacher age, race, education, years of teaching experience in Kindergarten, total years of teaching experience, if they taught in their current school the previous academic year and school were included as covariates in all path calculations. Additionally, pretest responses were included in models for the associated MV: for example, the estimate of the slope between the treatment condition and Data-Use Self-Efficacy was calculated controlling for pretest responses for Data-Use Self-Efficacy. This model was built in Mplus (Muthén & Muthén, 1998-2011), which provided estimates for the a-path (slope between Treatment Condition and MV), b-path (slope between MV and Data-Use Practices variables), c'-path (slope between Treatment Condition and Data-Use Practices variables), and *ab* path (indirect effect of MV on c' path).

Evaluation of bootstrapping confidence intervals was selected as the method to estimate the significance of the indirect effect because bootstrapping does not require an assumption of normal distribution, an assumption which can lead to positively biased results in smaller samples (Hesterberg, Moore, Monaghan, Clipson, & Epstein, 2005; Preacher & Hayes, 2008). Bootstrapping programs use re-sampling, or randomly replacing participants from the samples with other participants, to run the indirect effect across thousands of iterations, thereby creating a normal distribution from which to analyze the effect (Preacher & Hayes, 2008; Wu, 1986). To provide an estimation of the indirect effect's magnitude, the percent mediation (P_m) of each model demonstrating a significant indirect effect was calculated by dividing the indirect effect (*ab*) by the total effect (c') to determine the percent of the total effect accounted for by the indirect effect.

Phase 3: Fidelity and satisfaction analyses. As previously mentioned fidelity examination provides valuable information in understanding the implementation of applied interventions and is considered a best practice in RCT implementation (Chalmers et al., 1981; Downer, Locasale-Crouch, Hamre & Pianta, 2009; Dusenbury, Brannigan, Falco, & Hansen, 2003; Greenburg et al., 2003). To do this, the crucial components of the intervention must be determined. For the current study, two components were deemed necessary to successful implementation: the presence of a consultation meeting and the nature of cycles within a consultation meeting. The first component was assessed using between-group fidelity analyses, and the latter was explored within the treatment group only. Finally, teachers' satisfaction with the intervention was analyzed to address the ninth research question.

Content fidelity: Between-group fidelity analyses. Content fidelity, or adherence to the intervention's protocol, was operationalized as participation in a consultation session as described in the Measures section (page 93). Given the binary nature of the content fidelity variable assessing whether teachers received consultation or not, typical methods of evaluating fidelity using the average of quality indicators (Moncher & Prinz, 1991; Pett, 2015) was not applicable for the current study. Instead, between-group fidelity was calculated with the binary complier index method, which uses a binary variable to represent whether the participants received content of the intervention ($T = 1$) or not ($T = 0$). The binary index complier was calculated across consultants' and teachers' report for both intervention conditions using the procedure set forth in Dhillon, Darrow and Meyers (2014): sum T for each group (t^{TX} and t^{BAU}) then divide the sum by the number of participants in each group (n^{TX} and n^{BAU}) across reporters. These calculations provide interpretations of intervention adherence for each group across reporters thereby allowing for evaluation of content fidelity (t) in the treatment group and null fidelity ($1 - t$) in the BAU group.

Using the more conservative report of content fidelity for each group established above, the achieved relative strength (ARS) was calculated using the selected binary index complier. The ARS statistic allows for comparison of process fidelity among groups, thereby providing an interpretation of effect size while accounting for fidelity. According to Hulleman and Cordray (2009), the ARS index is calculated by taking the difference between achieved fidelity in the two groups and dividing this difference by their pooled standard deviation. This produces a Hedges' *g* statistic, which is then inferred using traditional effect size interpretations. Because the traditional Hedges' *g* formula uses means and the binary complier index provides a proportion, the indicated formula accounts for the proportionality as displayed below (Hulleman & Cordray, 2009):

$$g = [2 * arcsin\theta\left(\sqrt{(TTx)}\right) - 2 * arcsin\theta\left(\sqrt{(TBAU)}\right)] * \ (1 - \left(\frac{1}{4(n1 + n2) - 9}\right)$$

where *TTx* represents the treatment groups' fidelity, *TBAU* = the fidelity of the BAU group, *n1* = the sample size of the treatment group, and *n2* = the sample size of the BAU group.

Process fidelity: Within-group fidelity and satisfaction analyses. Because evaluation of process fidelity requires the intervention to have been delivered, only teachers who were established as receiving content fidelity in the previous analyses were included for the following analyses. Using this subsample, process fidelity was assessed through intervention differentiation across consultation sessions by three components:

instructional cycles, teachers, and consultations. Differentiation of consultation sessions by cycles was analyzed with frequencies and percentages of the number of cycles completed, the instructional content of cycles, the instructional level of cycles, and instructional strategies utilized during cycles per consultation session. Additionally, the instructional domains and levels utilized during cycles per session were cross-tabulated to provide information about the intersection of these two aspects of process fidelity. Differentiation of consultation sessions by teacher was calculated through frequencies and percentages of the variation in number of cycles, instructional domains, and instructional level per teacher. Differentiation of consultation sessions by consultant was calculated through frequencies and percentages of the variation in number of cycles, instructional domains, and instructional level per consultant among sessions.

In addition to intervention differentiation, the experiences of the treatment group were also evaluated through their satisfaction with the intervention. The average, standard deviation, and range of the satisfaction rating scale for teachers who received a consultation session was calculated to attain a descriptive understanding of teachers' perceptions towards the consultation. Furthermore, the number of negative and positive responses was summed across items. These totals were divided by the total number of responses. Negative responses were operationalized as ratings of 1 (Strongly Disagree), 2 (Mostly Disagree), or 3 (Slightly Disagree), while positive responses were operationalized as ratings of 4 (Slightly Agree), 5 (Mostly Agree), or 6 (Strongly Agree). These sums were divided by the overall number of responses on the satisfaction-related items (33 participants * 11 items, $N = 407$) to create ratios of negativity and positivity towards the intervention. Positive and negative responses were evaluated by the number of teachers who contributed to each ratio and the items which comprised each ratio to provide a deeper understanding of specific components of the intervention that led to negative and positive feelings, respectively.

CHAPTER IV

Results

Impact Analyses

The impact analyses address the first research aim concerning the efficacy of the intervention.

Intervention impacts on outcome variables. The pooled unstandardized betas of all variables in the linear regression models are displayed in Table 6. Across models, pretest responses significantly predicted their associated DV. Few other covariates produced significant effects on any outcome. Regression coefficients for covariate variables can be seen in rows three through 11 of Table 6.

Group mean, *t* values, and local effect sizes of the seven outcome variables can be seen in Table 7. Results are described below by outcome by construct as represented in the research questions.

Intervention impact on teachers' data interpretation skills. After controlling for covariates and teachers' pretest responses, the treatment condition had a statistically significant impact on the Data Interpretation Skills- Classroom variable ($B = 0.37$; $t = 4.43$, $p < 0.00$). The direction of the *B* value signifies that teachers in the treatment group demonstrated more accurate data interpretation skills as measured by their ability to select data-driven instructional strategies at the classroom level than BAU teachers. The intervention had a large effect on this variable ($d = 1.06$). The power of this outcome was calculated to be 1.00, signifying a 100 percent probability of correctly rejecting the null hypothesis that there would be no group differences in data interpretation skills at the classroom-level after the intervention.

Although the treatment condition did not have a statistically significant impact on teachers' data interpretation skills at the student-level ($B = 0.09$; $t = 1.42$, $p = 0.16$), the intervention demonstrated a small effect on this outcome ($d = 0.34$). The power of the Data Interpretation Skills- Student variable was calculated as 0.41, indicating a 41 percent probability of correctly rejecting the null hypothesis that there would be no group differences in data interpretation skills at the student-level after the intervention.

Intervention impact on teachers' beliefs about assessment. After controlling for covariates and teachers' pretest responses, the treatment condition did not have a statistically significant impact on teachers' beliefs about improvement, school accountability, or irrelevance with regard to assessment (Improvement Belief $B = 2.06$; $t = 1.59$, $p = 0.11$; School Accountability Belief $B = 0.49$; $t = 1.08$, $p = 0.28$; and Irrelevance Belief $B = 1.05$; $t = 0.75$, $p = 0.45$, respectively). Small effects were produced on the Improvement and School Accountability Belief variables by the intervention ($d = 0.38$ and $d = 0.26$, respectively). There was no discernable effect on the Irrelevance belief. The power for these three outcome variables were calculated as follows: 0.24, 0.03, and 0.07. These values suggest a 24, three, and seven percent probability of accurately rejecting the null hypothesis that there would be no group differences in Improvement, School Accountability, or Irrelevance beliefs, respectively, after the intervention.

Intervention impact on teachers' self-efficacy with data-use. After controlling for covariates and teachers' pretest response, the treatment condition produced a statistically significant impact on Data-Use Self-Efficacy ($B = 0.58$, $t = 2.04$, $p = 0.04$). The direction of the *B* value reveals teachers in the treatment group reported higher self-efficacy with regards to data-use at posttest as compared to the BAU teachers. The effect size of the intervention on this variable was small-to-moderate ($d = 0.49$). The power of the Data-Use Self-Efficacy variable was calculated as 0.69, indicating a 69 percent probability of correctly rejecting the null hypothesis that there would be no group differences in feelings of efficacy surrounding data-use after the intervention.

Intervention impact on teachers' data-use practices. After controlling for covariates and teachers' pretest response, the treatment condition did not produce a statistically significant impact on data-use practices ($B = 0.25$; $t = 0.51$, $p = 0.61$). There was a small effect of the intervention on this outcome, however ($d = 0.24$). Power for the Data-Use Practices was calculated to be 0.31, suggesting a 31 percent probability of correctly rejecting the null hypothesis that there would be no group differences in data-use practices after the intervention.

Indirect effect of intervention on data-use practices through changes in Data Interpretation Skills, Improvement Belief, and Self-Efficacy. The fifth research question concerning the influence of skills, beliefs, and self-efficacy on the relationship between the treatment condition and practices was addressed with a mediation analysis. The IV was set as the treatment condition, the DV was set as Data-Use Practices, and the MVs were Data Interpretation Skills- Classroom, Improvement Belief, and Data-Use Self-Efficacy. All four MVs were built into the model simultaneously, so each effect could be evaluated while controlling for the others. The first criterion for a test of mediation assumes the variables that constitute the c' path (IV-DV) will be significantly related, which was not the case as described by the lack of a significant impact of the treatment condition on the Data-Use Practices variable ($B = 0.25$; $t = 0.51$, $p = 0.61$). Consequently, the criteria for a mediation analysis were not met for the current study. Literature demonstrates, however, that the indirect effect of the MV on the relationship can still provide valuable information about the process occurring between the IV, DV, and MV (Hayes, 2009; Shrout & Boldger, 2002), especially in underpowered studies (Preacher & Selig, 2008). Because there was a small effect for this relationship ($d = 0.24$) despite the lack of power of the Data-Use Practices variable ($\beta = 0.31$), it is reasonable to assume the inferential worth of the p-value is low (Button et al., 2013; Cohen, 1995). In these cases, the indirect effect can be calculated, but should be interpreted as the process through which the MV is affecting the DV-IV relationship as opposed to a true mediation effect (Hayes, 2009; Shrout & Boldger, 2002). As such, the results of the current model (displayed in Table 8 and Figure 2) are described within this frame and should be interpreted accordingly.

To reduce the number of paths modeled (thereby increasing the number of free parameters), the Data Interpretation Skills- Student was dropped from the model as a MV. This variable was excluded, because the construct of ‘data interpretation skills’ was already accounted for by the Data Interpretation Skills- Classroom variable, which demonstrated better empirical alignment with the theoretical model in the impact analyses. As such, the final analytical model included three MVs (Data Interpretation Skills- Classroom, Improvement Belief, and Data- Use Efficacy), pretest variables associated with MVs and DV, correlations among the MVs, demographic covariates, the IV (Treatment Condition), and the DV (Data-Use Practices). This model demonstrated relatively good fit (RMSEA probability = 0.866, CFI = 0.994; TLI = 0.969). Although the RMSEA probability is higher than acceptable (Bollen, 1990; Raykov, 2000), research demonstrates the Tucker Lewis Index (TLI) is the best estimate of fit in smaller sample sizes because it compares the chi-square statistic to the null model, producing a relative fit index (Bollen, 1990; Marsh, Hau, & Wen, 2004). Given the high TLI, this model can be considered relatively stable and interpretable.

Results are reported as the standardized estimates of the a-path, b-path, and ab-path of each MV. The c' path (Treatment Group to Data-Use Practices) remained constant throughout the model: Estimate = 0.02, *SD* = 0.02, $p = 0.73$. Results described below and illustrated in Figure 2 are simplified for ease of interpretation but results of the full analytical model are listed in Table 8.

Indirect effect of intervention on data-use practices through changes in data interpretation skills. After controlling for the other paths of the model, the a-path describing the relationship between the treatment group and data interpretation skills at the classroom-level was significant (Estimate = 0.06, *SD* = 0.04, $p = 0.03$). The b-path describing the relationship between data interpretation skills at the classroom-level and practices was also significant (Estimate = 0.04, *SD* = 0.03, p =0.04). Bootstrapping procedures revealed an insignificant indirect effect of Treatment Condition on Data-Use Practices through the Data Interpretation Skills-Classroom variable, *ab* = 0.01 [-0.02, 0.03], therefore the ratio of the total effect accounted for by the indirect effect (P_m) was not calculated.

Indirect effect of intervention on data-use practices through changes in the improvement belief. After controlling for the other paths of the model, the a-path describing the relationship between the treatment group and the improvement belief was insignificant (Estimate = 0.01, *SD* = 0.03, *p* =0.72). The b-path describing the relationship between the Improvement Belief and practices was significant (Estimate = 0.03, *SD* = 0.03, *p* = 0.18). Bootstrapping procedures revealed an insignificant indirect effect for this relationship, *ab* = 0.00 [-0.05, 0.04], therefore the ratio of the total effect accounted for by the indirect effect (P_m) was not calculated.

Indirect effect of intervention on data-use practices through changes in self-efficacy. After controlling for the other paths of the model, the a-path describing the relationship between the treatment group and self-efficacy at the classroom-level was significant (Estimate = 0.13, *SD* = .03, *p* =.03). The b-path describing the relationship between self-efficacy and practices was significant (Estimate = 0.16, *SD* = .08, *p* =.04). Bootstrapping procedures revealed an indirect effect approaching significance of Treatment Condition on Data-Use Practices through the Data-Use Self-Efficacy variable, *ab* = 0.03 [0.01, 0.06], *p* = 0.04, with the intermediary variable accounting for 12 percent of the total effect (P_m = .12).

Fidelity and Satisfaction Analyses

The second research aim was addressed with the following fidelity and satisfaction analyses. Because these questions are descriptive in nature, results were analyzed using non-imputed data throughout this section. Content fidelity was analyzed to answer the sixth and seventh research questions concerning between-group experiences. Process fidelity was analyzed to explore the eighth research question surrounding within-group intervention differentiation. The final research question about teachers' attitudes towards the intervention was addressed through examining results of the satisfaction survey.

Content fidelity: Between-group analyses. Fidelity to the intervention was assessed across treatment and BAU teachers to evaluate between-group differences in intervention-related experiences.

Binary complier index of consultation using consultants' report. The binary complier index reflecting if a teacher received consultation or not based on the consultants' report was calculated for the treatment (t^{TX}) and BAU (t^{BAU}) groups. In the treatment group, 86 percent of teachers received a consultation session ($t^{TX}= 0.86$), indicating 14 percent infidelity with the model ($1- t^{TX} = 0.14$). No teachers in the BAU group received a consultation session ($t^{BAU}= 0.00$) signifying no fidelity for this group, as hypothesized.

Binary complier index of consultation using teachers' report. To check if teachers received consultation around their VKRP data with someone other than a VKRP consultant, the binary complier index of consultation was calculated using teacher report. In the treatment group, 86 percent of teachers reported receiving a consultation session ($t^{TX}= 0.86$), indicating 14 percent infidelity. Evaluation of the treatment teachers who responded positively to this item (n = 33) revealed that all did receive a consultation with a VKRP consultant. Similarly, teachers assigned to the treatment group who indicated they did not receive consultation (n = 5) matched those recorded as not receiving consultations by the VKRP consultants. In the BAU group, one teacher reported receiving consultation around her VKRP data ($t^{BAU}= 0.03$) signifying minimal fidelity, as expected for this group.

Achieved relative strength of the intervention. The achieved relative strength (ARS) of the intervention calculates a modified Hedges' *g* statistic that accounts for fidelity among both groups. To do this, the previously created fidelity constructs are evaluated within the context of the sample size. For the treatment group, the binary complier index of 0.86 was used since this was consistent across consultant and teacher report. In the BAU group, however, one teacher reported receiving consultation around her VKRP data, and although the specifics of this consultation are unknown, this occurrence represents possible fidelity to the IFDDC model. As such, the binary complier index of 0.03 was used to represent the fidelity proportion of the BAU group. After establishing estimations of fidelity, the ARS *g* was calculated for the seven outcomes (see Table 9). Given literature documenting that Hedges' *g* can be interpreted using the previously cited Cohen's (1977) guidelines (Cohen, 1992; Lenth, 2007), the treatment condition produced several effects. There was a large effect on the Data Interpretation Skills- Classroom ($g = 1.41$) and Data-Use Efficacy ($g = 0.82$) variables. A moderate effect was produced for the Data Interpretation Skills- Student ($g = 0.56$), Improvement Belief ($g = 0.44$), Data-Use Practices ($g = 0.39$), and School Accountability Belief ($g = 0.44$) variables. The effect of the Irrelevance belief continued to be below a practically meaningful magnitude.

Process fidelity: Within-treatment-group analyses. The within-group intervention fidelity analyses aim to explore the differentiation that occurred during the core component of the intervention, the consultation session. Therefore, these analyses only use the subsample of treatment teachers who a received consultation session as established in the between-group analyses (n = 33). Within-group fidelity is discussed as differentiation of consultation sessions by cycles, teacher, and consultant.

Differentiation of consultation sessions by cycles. Consultation session differentiation was explored by the quantity of, instruction content of, instructional level of, and instructional strategies utilized during cycles. Results are displayed in Table 10 and described below.

Quantity of cycles. On average, consultant-teacher pairs completed 1.96 (SD = 0.87) cycles. More specifically, three (10%) consultation sessions consisted of only one cycle, 17 (54%) consisted of two cycles, and eight consisted of three cycles (26%). This totals 61 cycles completed across all consultations. The remaining three consultation sessions (9%) did not complete any cycles. On two of these occasions, consultants were not able to guide teachers towards any data points of interest, and in the third the teacher reported that the data were not accurate for her students because of the bilingual-nature of her classroom and thus did not want to base any instructional decisions on the data.

Instructional domain of cycles. Cycles focused on one of the four instructional domains measured during VKRP- math, literacy, social skills, or self-regulation- and associated subdomains which are listed below but described in greater depth in Appendix 8.

Twenty-five cycles (41%) focused on math. Of those cycles with an instructional domain focus of math, the subdomains were represented as follows: Geometry n = 7 (28%), Numeracy, n = 11 (44%), Computation n = 6 (24%), and Patterning n = 1 (4%). It is important to note that the selected division's math curriculum lists 'patterns' as the topic for the first quarter, which had passed by the time of the consultation sessions. In accordance with this, several treatment teachers reported not feeling as though the Patterning subscale of the *Birthday Party: EMAS* was a valid representation of their students' skills by the time consultations occurred. One cycle (2%) focused on literacy, and more specifically on Rhyme. Teachers reported feeling more comfortable interpreting literacy data on their own, and thus little consultation time was dedicated to literacy data. Twelve cycles (19%) focused on social skills, whereas 23 cycles (38%) focused on self-regulation.

Instructional level of cycles. In addition to instructional content covered, the focus of cycles varied by instructional level, or the level at which the data were being interpreted. The instructional level of cycles occurred at the individual student-, small group-, or whole group-level. Twenty-one cycles (34%) focused on individual students, 20 cycles (33%) focused on small groups, and 20 cycles (33%) focused on whole group instruction.

Cross-tabulation of instructional domain and levels. Cycles varied in the level of instructional focus within instructional content. Of the 25 cycles that focused on math data, 5 (20%) were at the individual-level, 13 (52%) at the small group-level, and seven (28%) at the whole group-level. The one cycle that focused on literacy was at the individual-level (100%). Of the twelve cycles that focused on social-skills data, three (25%) were at the individual-level, three (25%) were at the small group-level, and six (50%) were at the whole group-level. Of the 23 cycles that focused on self-regulation data, 12 (52%) were at the individual-level, four (17%) at the small group-level, and seven (31%) at the whole group-level.

Instructional strategies utilized. After identifying the data point on which to intervene, consultants helped teachers select relevant instructional strategies. To do this, consultants used the VKRP strategies when possible. When no relevant strategy existed, consultants worked with teachers to create a new strategy. New strategies were designed for four cycles (7%), suggesting that the vast majority of teachers' identified points in the data for which VKRP provided relevant instructional strategies.

There were five different strategies recommended across the 25 cycles focusing on math. The most commonly referenced activity was one in which teachers scaffold students through learning early number sense by using gummy bears. This strategy was utilized for nine cycles.

Only two strategies were recommended for social skills-focused cycles. Of these 12 cycles, nine focused on supporting problem-solving skills by providing students with problem-solving options like "ask for help" and "find another activity." The other three cycles aimed to support friendship skills by pairing a student with low social skills with a student with higher social skills to complete an activity.

Four strategies were recommended across the 23 cycles focused on self-regulation. The activity utilized most often surrounds supporting children's emotion regulation through a social story. This strategy was provided during 15 cycles.

Differentiation of consultation sessions by teacher. Because sessions were individualized to teachers, there are not specific fidelity criteria for how sessions should have proceeded. Instead, the only component specified in the consultation manual was that at least one cycle be completed. Thirty sessions (n = 30, 91%) included at least one cycle as aforementioned. Of those teachers who completed more than one cycle (n = 25, 76%), six meetings focused on only one instructional domain; that is three meetings that included more than one cycle focused on only self-regulation, and three meetings that included more than one cycle focused on math only. Most teachers who completed more than one cycle, therefore, varied in the emphasis of their instructional domain. Meetings were fairly consistent in their instructional level when more than one cycle was completed; 12 cycles focused on one instructional level only. Three of these meetings emphasized individual students, four emphasized small groups, and five emphasized whole group instruction.

Differentiation of consultation sessions by consultant. Consultant 1 facilitated 26 meetings covering 46 cycles and Consultant 2 facilitated seven meetings with 17 cycles. This corresponds to an average of 1.77 cycles per meeting for Consultant 1 and 2.43 for Consultant 2. The majority of Consultant 1's cycles focused on the instructional domain of self-regulation (*n* = 19, 41%) or social skills (*n* =12, 26%). The instructional domain of cycles led by Consultant 2 mostly emphasized math (*n* = 11, 65%), followed by self-regulation (*n* = 6, 35%). Consultant 1 facilitated 26 cycles (43%) focused on small-group and 15 cycles (33%) on whole group instructional levels, with the remaining 11 cycles (24%) focused on individual students. Consultant 2's cycles emphasized the individual student (*n* = 12, 71%) or whole group (*n* = 29%) instructional levels. A complete breakdown of instructional domain and level by consultant can be seen in Table11.

Satisfaction with consultation. Teachers reported generally positive feelings towards the consultation; on a scale of 1 (Strongly Disagree) to 6 (Strongly Agree), teachers who received consultation reported an average satisfaction rating of 4.26 (SD =

0.59). This mean falls between the ratings ‘Slightly Agree’ and ‘Mostly Agree.’ Teachers provided mostly positive feedback towards the intervention; for example, on the item “Overall I found the consultation useful in informing my teaching practices,” 82 percent responded with a statement of positive agreement. In sum, there were 355 positive responses recorded across the 407 satisfaction-related responses, which calculates to a positive response ratio of 87 percent across these items. The majority of teachers responded with all positive feedback (n = 26). Exploration of the 13% (n = 52) negative responses revealed that three teachers responded to every item negatively (n = 33), which accounted for 63 percent of the negativity reported. Teachers who responded to every item with a negative statement were the three teachers who participated in consultation sessions in which no cycles were completed. The remaining 37 percent of negative responses were spread across four additional teachers, none of whom provided negative responses to more than half of items.

CHAPTER V

Discussion

In recent years, there has been a particularly intense focus on measuring students' abilities as they enter kindergarten, which is well-aligned with the standards-based movement (SBM). This attention originates from compelling research demonstrating strong associations between students' incoming skills at kindergarten entry, or their readiness for kindergarten, and later academic and social success (Goldstein et al., 2014; Meisels, 1996, 1999). In addition to predicting achievement, data obtained from kindergarten entry assessments (KEAs) can be used to identify students for and create early interventions, which has been deemed essential to eliminating national education gaps (Fryer & Levitt, 2004; Ramey & Ramey, 2004). Research supports the use of assessments to identify and inform interventions for these students, who demonstrate improved learning outcomes when teachers use assessment data to select instructional strategies (Ball & Gettinger, 2009; Stecker et al., 2005). Exploratory evaluations of this process, however, have found several barriers to implementation, mostly surrounding teachers' practices, skills, and beliefs related to the use of data in the classroom (Brown, 2004, 2009; Stecker et al., 2005).

The purpose of the current study was to extend the growing body of literature evaluating how to assist teachers in using data to inform their classroom practices through the creation and implementation of a new instructionally focused, data-driven consultation (IFDDC) model. Within this larger goal, two specific aims were formed: (1) to evaluate the impact of IFDDC on teachers' skills, beliefs, self-efficacy, and practices, and (2) to explore the fidelity of the intervention in greater depth to understand variation in implementation.

Examination and Interpretation of Study Results

Understanding the effects of the intervention on teachers' skills, beliefs, self-efficacy, and practices surrounding data-use. The first aim of the current study was to establish the efficacy of the IFDDC intervention on influencing teachers' skills, beliefs, self-efficacy, and practices. Overall, results indicate that the intervention had several positive effects on these teacher outcomes, as elaborated below. Given the underpowered nature of the study, impacts are interpreted across statistically significant findings and effect sizes that reach a statistically relevant magnitude (Cohen, 1995) to provide a complete picture of the outcomes demonstrated.

Intervention effects on teachers' data interpretation skills. The strongest effect produced by the intervention was on teachers' data interpretation skills at the classroom-level. Teachers who received IFDDC demonstrated significantly better skills in interpreting classroom data than teachers who did not receive IFDDC, as hypothesized. This finding suggests that through our model, which included only one in-person consultation session, teachers were able to acquire data interpretation skills for this particular set of classroom-level assessment reports. Their quick uptake of these skills is promising as it suggests relatively low dosage, cost effective consultation models that explicitly focus on understanding data may produce impacts on teachers' interpretation skills. Interestingly, while the intervention had a large impact on teachers' interpretation skills with respect to classroom data, only a small effect was detected on their ability to interpret student-level data. There are three possible explanations for this discrepancy: measurement was inconsistent across levels, the consultation did not focus on student-level data interpretation and thus teachers could not be expected to improve in this skill, or classroom-level data are easier for teachers to interpret. The first justification relating to how the variables were measured calls for evaluation of two components, differences in measurement and fatigue across levels of skills. Classroom- and student-level data interpretation skills were assessed in the same manner, and the two variables demonstrated similarly high concordance among expert reviewers, suggesting there were few differences in measurement between the levels. It is possible, however, that participants were fatigued by the time they were asked to interpret student-level data, as this was the last component of the teacher survey. Expert reviewers only answered the two data-interpretation questions instead of the full survey, so this fatigue likely did not affect their ratings. To clarify this, a larger study could randomize the order by which teachers are asked to interpret data, which would provide information on the potential effects of fatigue.

Another possible explanation is that the consultation focused more heavily on interpreting classroom-level data, rather than student-level data. Exploration of the consultation processes demonstrated forty-one percent of all sessions included at least one cycle focused on interpreting individual student data. Moreover, two-thirds of all teachers participated in consultation sessions that included student-level cycles. This suggests that the some teachers did no engage in session that included student-level data interpretation, and thus would not be expected to improve in this domain. In this way, the focus of consultation sessions may have contributed to the disparity between classroom- and student-level data interpretation skills.

Additionally, there may be something fundamentally different about the way teachers interpret classroom- and student-level data that makes the former easier. It is possible that teachers are asked to interpret classroom data more often and thus could build off of their prior experiences. In support of this hypothesis, many data-driven interventions provide classroom-level interpretation such as whole-group instruction strategies or suggested groupings based on data (Grow Network, 2000; Invernizzi et al., 2003). For example, Grow Reports© (Grow Network, 2004), which were employed across school districts in New York and Chicago public schools from 2001 to 2003, provide teachers with their classroom's data, areas for growth, and web-based tools to improve the areas for growth (Brunner et al., 2005; Easton & Luppescu, 2004). Descriptions of the agency's goal and how the reports were designed highlighted "class-wide priorities" as the targeted goal for the intervention (Brunner et al., 2005, p.2). The emphasis on classroom-level instruction evidenced by Grow Reports© and other data-driven interventions suggests that teachers receive aid in interpreting classroom data more often than individual students' data. It is conceivable that teachers may have been asked to interpret classroom-level data previously and therefore had a skill base from which to draw during the current intervention. In support of this hypothesis, the mean-level of teachers' perceptions of administrators' support for data-use was relatively high, suggesting the selected division fostered a culture supportive of data-use. To fully understand how to best aid teachers in data interpretation skills at all levels, therefore, it will be essential to not only diversify the participant base but also to consistently embed interpretation of data across levels.

Intervention effects on teachers' beliefs about assessment and the uses of data. As hypothesized, the intervention had a statistically insignificant but small effect on teachers' 'improvement' beliefs; teachers in the treatment group reported ascribing to the belief that assessments could be used for informing practice (Improvement) at a higher rate than BAU teachers at posttest but the difference in this rate was not statistically significant. This small effect was not limited to beliefs surrounding the use of assessment for informing practice, however, as predicted; the intervention also produced a statistically insignificant but small effect on teachers' endorsement of beliefs pertaining to assessment for keeping schools accountable. One argument for this unexpected impact is that these constructs are more closely related for teachers of young children. The majority of work on the measure of beliefs used in the current study has been conducted with teachers of older students who are required to participate in state accountability tests (Brown, 2004a; Brown, 2006; Brown et al., 2015). It may be that the differentiation between the Improvement and School Accountability belief is not as salient for teachers of younger students who are often not required to take these same assessments. In support of this assertion, the Improvement and School Accountability factors were moderately correlated in our pretest data. Furthermore, an evaluation which employed a different method of assessing teachers' beliefs found similar correlations between these constructs in teachers of young children (Ebbeler et al., 2016). This preliminary evidence suggesting association between beliefs about the utility of assessment for improving instruction and school accountability is consistent with theories that less stringent accountability standards affect teachers less (Hamilton et al., 2007; Ingram et al., 2004). As such, it is possible that the teachers in our study did not discern between the impacts of assessment for improving learning and school accountability, but rather associated these two constructs as a more general positive belief about assessments.

Intervention effects on teachers' self-efficacy surrounding data-use. The intervention had a statistically significant impact on teachers' reported efficacy with data; as hypothesized, teachers who received IFDDC were more likely to report feeling efficacious with data-use than the BAU teachers after controlling for their pretest responses. This effect was noted to be small-to-moderate in magnitude (d = 0.49), which is consistent with effect sizes noted in evaluations of similarly-focused PD (Ross, 1992; Rowe & Sykes, 1989; Tschannen-Moran & McMaster, 2009). Building self-efficacy in a domain is important when expecting associated changes in behaviors (Bandura, 1993; Tschannen-Moran & McMaster, 2009), so the direct effects of IFDDC on this construct represent a promising step for reaching the ultimate goal of affecting teachers' data-use practices.

The effects of IFDDC on teachers' feelings of efficacy pertaining to data-use are especially encouraging given surveys demonstrating teachers report experiencing more difficulty in using data as compared to many other classroom-related practices (Tschannen-Moran et al., 1998). Through a data-focused PD model such as that presented in the current study, it may be possible to increase teachers' self-efficacy around data-use and their general willingness to participate in future data-related interventions.

Intervention effects on teachers' classroom practices surrounding data-use and preliminary evidence for mechanisms behind these effects. The intervention effects on the ultimate goal IFDDC, increasing teachers' classroom practices relating to data-use, was assessed through direct and indirect methods to provide a full interpretation of the impacts of IFDDC on this essential outcome.

Direct effects of the intervention on teachers' self-reported practices. The intervention produced a small effect on teachers' reported classroom practices in that teachers in the treatment group reported implementing more practices surrounding data-use at posttest. This impact, albeit statistically insignificant, suggests that PD can impact teachers' self-reported instructional with relatively low dosage if that dosage is focused in scope. It is important to qualify these results as reported changes in practice, and as such it is unknown the extent to which classroom instruction was actually affected. Research demonstrates variability in the accuracy of teachers' descriptions of their instructional practices (Shoval, Erlich, & Fejgin, 2010; Spear-Swerling, Brucker, & Alfano, 2005), although teachers tend to achieve higher agreement with observers when they are asked to report on practices in a specific area (Koziol & Burns, 1986; Spear-Swerling et al., 2005) as occurred in the current study. Little is known about teachers' perceptions of their data-use practices specifically, and other data-driven interventions have cited similar challenges with discerning practices from teachers' reports (Chick & Pierce, 2011; Ebbeler et al., 2016). Given the necessity of influencing teachers' practices for affecting student achievement, future iterations of IFDDC or comparable models may consider adding an observation measure to allow for comparison of teachers' perceptions and independently observed practices that arise from data-use in the classroom.

Understanding the mechanisms through which the intervention affected teachers' classroom practices. The final goal of the impact analyses was to explore the validity of the theory of change (TOC) model presented in Chapter II, which specified that the intervention would have indirect effects on teachers' reported data-use practices through changes in skills, beliefs, and self-efficacy. As discussed previously, no statistically significant impact was observed between the treatment condition and reported classroom practices, violating the first assumption of mediation that there must be a direct effect of the IV on the DV (Button et al., 2013; Cohen, 1995; Hayes, 2009; Shrout & Boldger, 2002). The intervention produced a small effect on reported practices, however, providing evidence that a more significant direct effect may exist in a fully powered study. Consequently, the indirect effects of the intervention on practices through the skills, beliefs, and self-efficacy variables were modeled, though findings should be interpreted within an exploratory frame. That is, these results are relevant in providing preliminary evidence about the TOC model, but will require further evaluation before any strong conclusions can be drawn. As such, interpretations are discussed below within a preliminary schema.

Of the variables modeled, the intervention only produced a significant indirect effect on reported practice through teachers' self-efficacy. This finding is consistent with Bandura's (1993) theory that self-efficacy leads to action. Unfortunately, data on these constructs were collected simultaneously, making the direction of relationship between the two impossible to disentangle; that is, although it was hypothesized that intervention built teachers' self-efficacy which in turn contributed to changes in their practices, it is possible that the intervention changed their practices, which then augmented their self-efficacy. In support of the TOC model presented, studies of this mechanism have demonstrated the relationship to be in the direction hypothesized in that teachers' sense of self-efficacy results in behavioral changes (Tschannen-Moran et al, 1998). Furthermore, the direct effect of the intervention on self-efficacy was much stronger than that produced by the intervention on practices, lending further support for the hypothesized direction. In this way, it is possible that IFDDC set the foundation for teachers to change their data-use practices by first building self-efficacy in this domain. Measuring these constructs at multiple time points will aid future endeavors in supporting or refuting this theory.

In contrast to hypotheses based on the TOC model, neither the improvement belief nor data interpretation skills at the classroom-level were pathways through which the intervention had an indirect effect on teachers' data-use practices. There are two potential reasons for this dearth of effects: the hypotheses were correct and the study was too underpowered to detect effects, or the hypotheses were incorrect and no indirect effects are present. In support of the former explanation is the overall underpowered nature of the study; for example, post-hoc calculations revealed power for the data-use practices outcome was 0.31. This suggests that there was a relatively high probability of incorrectly supporting the null hypothesis in evaluation of this construct, so it is possible that the indirect effects were not detected given the low power.

Alternatively, it is possible that neither the improvement belief nor data-interpretation skills functioned as mediators. Some evidence suggests that for teachers, changes in beliefs can be preempted by changes in practices (Fang, 1996; Mansour, 2009), and the relatively small effect of IFDDC on practices may not have been strong enough to influence beliefs. Furthermore, the intervention focused more explicitly on improving practices by recommending specific strategies than on changing beliefs. It is possible that changes in practices were facilitated by these direct recommendations, thereby skipping the hypothesized indirect effect through changes in beliefs.

Literature detailing the necessity of data literacy skills for changes in data-use is well-established (Chick & Pierce, 2013; Pierce et al., 2011), however, requiring a different explanation. Instead, it is possible the skills measured as a part of IFDDC were specific as previously discussed, indicating less generalizability of these skills to teachers' other data-use practices. In other words, the measured data interpretation skills, while they were enhanced by IFDDC, did not contribute to the relationship between the intervention and changes in practices. Similar to that needed to untangle the relationship between IFDDC, self-efficacy, and classroom practices, it will be crucial for future evaluations to include multiple data collection periods to allow for more certain interpretation of the direction and strength of these relationships.

Understanding between- and within-group variability that occurred during IFDDC implementation. After evaluating the efficacy of the intervention, the second aim of the current study was to explore the effects of IFDDC implementation by analyzing between- and within-treatment group data.

Achieved content fidelity and implications for intervention effects. Results of the content fidelity analyses indicated differences in the data-related experiences of teachers by treatment condition as hypothesized. Specifically, teachers assigned to the intervention group were recorded as and reported experiencing consultation around their assessment data at a higher rate than their BAU counterparts. The achieved fidelity of the intervention (86%) is relatively high compared to similar models where participation is not mandated (Dimmitt et al., 2007; Ebbeler et al., 2016), suggesting the IFDDC protocol set achievable expectations for implementation in schools. The five teachers in the treatment group who did not receive the intervention were limited by scheduling difficulties; specifically, the time slots offered by the consultants did not match their availability. This suggests relatively simple supports, such as greater availability of consultants or classroom coverage, could augment implementation. In addition to the high fidelity observed in the treatment group, the BAU group was recorded as and reported experiencing low fidelity to the model as expected. Only one teacher (3%) assigned to the BAU condition received one-on-one consultation around using assessment data, which is consistent with evaluations demonstrating little support is typically provided to teachers around using data (Dimmitt et al., 2007).

Even with the relatively high fidelity of the intervention, infidelity to the model had an impact on the achieved effects as demonstrated by the consistent increases in effect magnitude in the ARS analyses as compared to the typical procedure for evaluating effects used to produce the Cohen's d statistics. The ARS procedure allowed for the intervention effects to be evaluated within the scope of the achieved content fidelity (Hulleman & Cordray, 2009), providing an estimation of the outcomes considering infidelity. When the ARS is higher than the observed effect, as occurred for all outcomes in the current study, we can conclude infidelity to the model reduced the efficacy of the intervention following literature indicating the comparability of d and g statistics (Cohen, 1977; 1995; Lenth, 2007). Specifically the intervention's effect on data interpretation skills at the classroom-level increased by 33% (d = 1.06; ARS g = 1.41), at the student-level by 64% (d = 0.34; ARS g = 0.56), Improvement belief by 16% (d = 0.38; ARS g =

0.56), School Accountability belief by 27% (d = 0.26; ARS g = 0.33), self-efficacy by 67% (d = 0.49; ARS g = 0.82), and data-use practices by 62% (d = 0.24; ARS g = 0.39). It is possible, therefore, that if IFDDC was implemented with greater fidelity through the aforementioned supports, larger effects could be achieved.

Differentiation of process fidelity throughout consultation sessions. Process fidelity analyzed within the treatment group provided useful information on the proceedings of the intervention. These results coalesced into interpretations across the individualization of the model, the necessity of SEL supports, and the availability of relevant, evidence-based instructional strategies.

The utility of an individualized intervention model. As expected given the individualization built into the protocol, there was variation in the intervention received by teachers in both quantity and instructional foci of cycles. During some sessions, consultants and teachers focused on one data point and associated instructional strategy, whereas in other sessions attention was diverted across multiple data points and/or strategies. Even allowing for this variation in the content of the intervention, effects were evident as described above. Individualization of PD is considered essential, but data-driven models are often rigid in their content delivery (Halverson et al., 2006; Pierce & Chick, 2011). Findings from the current study offer preliminary evidence that a data-focused consultation protocol that is individualized can still be effective. Future work might evaluate the extent to which individualization is possible within sessions while continuing to produce effects, and if effects are reduced with a more prescribed focus. Furthermore, future evaluations should consider identifying the components of the intervention that contributed most to changes in skills, beliefs, self-efficacy and practices.

The small sample size of the present study did not allow for evaluation of this, but examinations of consultation and coaching consistently indicate the specific proceedings of sessions have differential effects on measured outcomes (Kretlow & Bartholomew, 2010; Noell & Witt, 1999). As such, it is conceivable that there are aspects of the intervention that are more effective, and these components could be elevated while still maintaining a level of individualization.

The need for SEL instructional support. Despite offering consultative services across math, literacy, and social-emotional data, 57 percent of cycles were allocated to the latter. These cycles were implemented across all instructional levels, suggesting teachers more often requested assistance in interpreting social-emotional rather than math or literacy data. The consistency with which social-emotional data was emphasized supports Cook's (2002) results demonstrating teachers' feelings of unpreparedness with this instruction. Our results suggest teachers are seeking help in facilitating classrooms which are supportive social-emotional development for their students, which is promising given the connection between this domain and students' success (Raver & Zigler, 1997; Webester-Stratton & Reid, 2004) and teachers' wellbeing (Jennings & Greenberg, 2009; Marzano, Marzano, & Pickering, 2003).

Although there are many social-emotional interventions, like those noted in Durlak and colleague's (2011) review, that aim to help teachers develop these skills in their students, our results demonstrate teachers continue to need support in this domain. This could be the result of the relatively recent push for social-emotional emphasis in classrooms (U.S. DOE, 2013) which would suggest resources supportive of social-emotional development have yet to be integrated into teachers' practices. There could also be something inherently challenging about this type of instruction for teachers to implement. For example, Jennings and Greenberg (2009) highlight the importance of teachers' own social-emotional competence in facilitating prosocial classrooms. Unlike models for improving teachers' academic skill instruction, PD aimed at augmenting teachers' social-emotional teaching practices will likely need to build teachers' self-awareness and -regulation (Brown, Ryan, & Creswell, 2007; Jennings & Greenberg, 2009). Another possible cause could relate to the reluctance of schools and educators to implement SEL for fear of losing instruction time to non-academic domains (Kramer, Caldarella, Christensen, & Shatzer, 2010 Seifer et al., 2004). Likely, the explanation lies somewhere between these possibilities, suggesting the path to improving social emotional instruction in classrooms will necessitate intervention at the teacher- and school-level.

The availability and relevance of evidence-based instructional strategies for teachers. Examination of the strategies utilized during consultation sessions revealed almost all recommended strategies existed prior to the intervention; that is, few cycles required creating new instructional strategies. This finding highlights that useful instructional practices exist for teachers, but they need help locating and identifying when to implement them. Planning for instruction is consistently relayed as a necessary component for effective teaching (Middendorf, 2007; Miles & Darling-Hammond, 1998), but teachers frequently indicate not having sufficient time to plan for instruction (Collison & Cook, 2001; Melnick & Meister, 2008). Teachers report allocating 45 to 220 minutes per week on designing lessons (Miles, 1995; Miles & Darling-Hammond, 1998; Spear-Swerling & Zibulsky, 2014), time which could be streamlined by supports that identify effective practices for them such as occurred in the current intervention.

Exploring teachers' satisfaction with IFDDC. Results indicate teachers were generally satisfied with IFDDC; seventy-nine percent of teachers who participated in consultation sessions responded with exclusively positive statements across all aspects of the intervention. Explorations of those who did not report entirely affirmative sentiments revealed that the majority of negative responses originated from three teachers who participated in a consultation session but completed zero instructional cycles. This signifies although consultants met with teachers, the dyad was not able to identify data points on which to intervene and/or strategies related to those data points. It is logical that these teachers had negative responses to IFDDC in these cases, because the consultation session did not achieve the intervention goal: to help teachers implement instructional strategies based on their classroom- and student-level data. It may have been that the assessment data were not relevant to their classroom and/or that the teachers did not think the data were relevant to their classrooms. In either case, it will be important to gain a deeper understanding of what led to the inability to help these teachers so that either the assessment or framing around the assessment can be altered to be globally useful.

Despite these few instances of negative feedback, the overall responses to the intervention were pro-IFDDC. This positive sentiment towards the intervention is promising, because there were several potential barriers to teachers' satisfaction with participation. To begin, teachers were not provided with classroom coverage or additional paid time to meet with consultants. Instead, they scheduled meetings with consultants during their planning time or before or after school hours. School-based PD models almost always account for teachers' time by providing classroom coverage or small stipends for participation (Belzer, 2005; Borko, 2004; Desimone, 2009). Furthermore, teachers in the current study were not provided with professional learning credits, which teachers are required to obtain every year in most states (Sarama, 2002). Teachers' satisfaction with IFDDC despite a lack of logistical or tangible compensation indicates teachers perceived the intervention to be valuable over and above the time spent participating.

In addition to the logistics of participation, IFDDC also had the potential to be less than satisfying given some teachers' noted disliking and subsequently avoiding data-related experiences (Kelly, Downey, & Rietdijk, 2010; Marsh, 2012). This was not true of the current intervention, however, as the protocol emphasized data-driven instruction to which teachers responded positively. High levels of satisfaction with IFDDC may relate to the individualization of the intervention content; results from other PD models demonstrate teachers appreciate tailored attention to their classroom practices (Garet et al., 2001; Nir & Bogler, 2008; Tschannen-Moran & McMaster, 2009). In fact, teachers respond so positively to tailored content that Garet and colleagues (2001) included individualization as an essential feature of PD in their review of effective school-based PD models. It is possible that providing highly individualized content in a data-driven model may supersede the negative feelings teachers' harbor towards data-use in general. It should be noted that the belief schemas of teachers' towards assessment as measured at pretest were generally positive as were their perceptions of their schools' climate towards data-use, suggesting the participant groups' conceptions were more positive than that of other samples (Brown, 2004a). As such, evaluating the effects of IFDDC in divisions with more diversity in teachers' conceptions about and school climates towards data-use will be important in evaluating the attitude of teachers towards this data-focused model.

In sum, despite the potential drawbacks to participation, teachers reported feeling satisfied with the intervention. As such, future evaluations may consider using components found successful in the current study when designing future PD models.

Limitations and Recommendations for Future Research

The findings of the current study should be taken in context of several limitations. All limitations are described with recommendations aimed at accounting for these drawbacks in future studies.

Measurement. There were several measurement-related limitations in this study. Foremost, four of the outcomes self-efficacy, practices, classroom data-interpretation skills, and student data-interpretation skills were assessed through measures created for this evaluation. Although all four measures demonstrated good reliability, the validity cannot be ensured without additional research. In addition to validating these measures, each was limited in scope. For example, the Data-Use Practices measure was gathered through teacher-report, meaning we relied on teachers to accurately describe their classroom practices. As previously described, there is mixed evidence on the accuracy of teachers' perceptions of their practice (Shoval, Erlich, & Fejgin, 2010; Koziol & Burns, 1986; Spear-Swerling, Brucker, & Alfano, 2005), suggesting only limited inferences can be drawn regarding actual changes in practice as a result of IFDDC. Adding an observational measure of data-use practices to the procedure would allow for more definitive conclusions about teachers' changes in practice, an essential component to improving students' learning (Richardson, 1990; Stecker et al., 2005).

The data interpretation measure was also limited in scope. The measure used quantified proficiency in interpretation as teachers' ability to select relevant instructional strategies. Although the ability to select and enact relevant instruction is the ultimate outcome when interpreting data (Stecker et al., 2006), there are many more discrete steps to understanding data that were not captured. For instance, the Wayman and Jimerson (2013) competencies used as a partial foundation for the current intervention cite triangulation, or using multiple data points, as a crucial sub-skill of interpreting data. Although participants needed to do this in forming their answer, their ability to do so was not measured. It is unknown, therefore, if teachers who did not choose related instructional strategies were unable to triangulate the data or unable to select indicated strategies after correctly triangulating. A measure which requires teachers to walk through each of the data interpretation steps separately could isolate specific deficiencies which could then be the target for intervention. Furthermore, the data reports on which teachers' data interpretation skills were measured was specific to VKRP, reducing the generalizability of these findings. That is, teachers who participated in IFDDC demonstrated an increase in their ability to interpret the particular reports used in IFDDC, which may not be representative of all data formats. In addition to broadening the interpretive steps of the measure, it may also be useful to include different presentations of data to ensure teachers can generalize these skills across reports.

Lack of power and teacher variation. A major limitation of the current study is the lack of power as demonstrated by the a priori and ad hoc power analyses. Although all efforts were made to recruit 116 teachers in keeping with the a priori target, time and resource constraints only allowed for 72 teachers to be consented and randomized. An RCT with a larger sample may be able to detect effects for outcomes that were underpowered in the current study.

Not only should future studies be more expansive, but also include participants from different school divisions. The present study only sampled from one division due to consultant availability, which has two potential implications for predicting outcomes: a lack of variability in participant demographics and district policy coherence. The current sample included mostly white females with masters' degrees, which are demographics consistent with the division sampled. Teacher characteristics across the U.S. are more varied, however; for instance, one-sixth of public school teachers are male and 45 percent have no graduate-level training (Feistritzer, 2011). Given evidence demonstrating differences in intervention uptake and potential effects given teachers' characteristics (Klassen & Chiu, 2010; Witt, Noeel, LaFleur, & Mortenson, 1997), a more diverse sample would be beneficial for understanding bidirectional connections between teacher demographics and the intervention in future evaluations.

In addition to diversity in participants, including more school divisions in subsequent studies would allow for evaluation of the intervention in the context of differing political climates and educational requirements. As described in Bronfenbrenner's ecological systems theory (1979), exo-systems such as state agencies and legal mandates influence more proximal entities to individual's functioning. Research substantiates this model in educational systems and is directly related to school-based interventions through policy coherence (Cohen, 1995; Newmann, Smith, Allensworth, & Bryk, 2001). The construct of intervention coherence with district policies describes the alignment between a particular intervention and the educational requirements of the district in which the intervention is occurring (Cohen, 1995; Newmann et al., 2001). Although it is often discussed within the confines of curriculum interventions, PD models are also likely to be influenced by school- or district-level requirements. For example, the division sampled in the current study required teachers to use data during parent-teacher conferences, which necessitates looking at a student's data at least once. Furthermore, this division has been involved with PALS, an ongoing readiness assessment since 2001, suggesting teachers participating in the intervention had some base familiarity with classroom-level data. The emphasis of data-use in the division was reiterated by teachers on the survey beckoning their perceptions of data-use climate, in which 96 percent of teachers agreed to some degree with the statement "My school administrator clearly defines standards for using assessment to inform instructional practices." As such, the current intervention focusing on using readiness data to inform classroom practices conceptually and logistically fit within the district-level requirements of data-use. It is conceivable to envision a less data-centric division; for instance, one in which teachers were not primed with how to use data from previous years and/or were not allocated adequate time to meet and review data with a consultant. Indeed, Slavin and colleagues (2011) found that many divisions in their study did not emphasize data-use before the intervention. In sum, the current division was a prime candidate for IFDDC, which may have affected the uptake of the intervention, and thus future research should evaluate this intervention in divisions with different climates to understand the exo-system effect on outcomes.

Low intervention dosage. Although not necessarily a limitation, the dosage of the current intervention should be considered when interpreting effects. IFDDC included one consultation meeting and one required follow-up encounter which can be estimated as 1.5 hours of contact between consultants and teachers. This is relatively little interaction for a PD model, as meta-reviews of PD models cite ten to fifteen hours as the base level of contact with teachers to find medium-to-large effects on instruction (Hill, Beisiegel, & Jacob, 2013; Wayne, Yoon, Zhu, Cronen, & Garet, 2008). The large effect detected on teachers' ability to interpret classroom-level data with this low dosage indicates the current model of IFDDC may be a cost-effective intervention for affecting teachers' skills.

With the current intervention dosage considered, a similar model with greater frequency and/or intensity may be able to produce more robust effects. For example, literature on effective consultee-centered consultation often cites two to five sessions as the standard for these models in early childhood educational settings (Hylander, 2012; Sandoval, 2013). Additional sessions would open many opportunities for the expansion of the model, such as providing more vigorous feedback to teachers or allowing for data collection after strategy implementation. The latter is crucial in building teachers' understanding of the cyclical nature of data-use (Boudett, City & Murnane, 2005). Supplying additional feedback to teachers, either around their data interpretation or strategy implementation, has the potential to further cultivate their self-efficacy around data-use and subsequent motivation to change (Bandura, 1993; Rosenfield, 2013; Sandoval, 2013). Given the benefits of providing more sessions, future evaluations could consider utilizing a remote consultation approach to raise dosage but maintain cost-effectiveness. Similar PD models delivered via telecommunication have been demonstrated to produce positive effects at both the teacher- (Artman-Meeker, Hemmeter, & Snyder, 2014; Downer, Kraft-Sayre, & Pianta, 2009) and student-level (Allen, Hafen, Gregory, Mikami, & Pianta, 2015; Cabell & Downer, 2011). These results indicate that the dosage of IFDDC could be increased in a relatively low-cost manner without sacrificing effects.

In addition to increasing the frequency and/or intensity of consultation, IFDDC could be supplemented with coaching surrounding the implementation of strategies, rather than continued data-work. Coaching models have demonstrated success in increasing teachers' use of effective instruction (Artman-Meeker et al., 2014; Downer et al., 2009); therefore, follow-up sessions that focus exclusively on teachers' practices has the potential to ensure actual rather than self-perceived changes in practice.

The model could also be augmented by integrating school administrators into the intervention. Models of data-driven reform emphasize district-level interventions, frequently by expecting administrators to facilitate the consultative process with teachers (Slavin et al., 2011; Wayman, Jimerson, & Cho, 2012). This approach has multiple benefits including building collective efficacy (Bandura, 1993), affecting school-wide culture (Hamilton et al., 2009; Wahlstom, 2008), and increased implementation fidelity (Slavin et al., 2011).

Practical Implications

Several recurring concepts emerged in the evaluation of IFDDC that were not directly related to the current study but have implications for teachers' practice.

Automatically linking data to strategies. Given data that suggests the flexibility of PD models to cater to the specific needs of teachers is an essential component of effective interventions (Desimone, 2009; Hill et al., 2013), the demonstrated range in process fidelity of IFDDC was purposeful. When contemplating the potential of bringing this or similar interventions to scale, however, a more systematic approach to this individualization of sessions may be considered. For instance, a teacher's data could be automatically linked to strategies using an algorithm, allowing consultant-teacher dyads to spend more time discussing implementation. Many data-driven interventions utilize computer-generated instructional recommendations based on data and have demonstrated medium impacts on practice (e.g., Ebbeler et al., 2016; Fuchs et al., 1992). Furthermore, creating a system that automatically links strategies to teachers' data would diminish consultant-level variation in recommendations. In the current study, it was impossible to tease apart these differences given the small sample and the aforementioned individualization of the intervention to teachers. However, it is reasonable to expect some differences in the types of recommendations consultants provide given their background, prior experiences, and/or education. This may not be a limitation, per se, but adds further noise into an already flexible intervention. Linking strategies to data automatically eliminates this variation and would allow the intervention's efficacy to be evaluated more clearly.

Teacher education focusing on data-interpretation. As was expected given literature revealing little-to-no emphasis in pre-service work on data-interpretation (Bowen & Roth, 2005), teachers in the current sample demonstrated low levels of data-interpretation skills at pretest. Although this offers a ripe area for in-service PD, it also illuminates the potential for augmenting teachers' knowledge and skills related to data-interpretation during pre-service education and student teaching. This direction fits into new emphases of teacher education, which often support the importance of reflective thinking on instructional practices (Lee, 2005). Within the frame of collecting and analyzing data for improving instruction during pre-service work, teaching data-interpretation skills would provide a basis for data-driven practices, which is an area currently lacking attention in teacher preparation programs (Dimmitt et al., 2007).

In sum, there are many ways in which the supports surrounding IFDDC could be enhanced to potentially increase gains in teachers' ability to use assessment to inform their instruction.

Conclusion

In reconsidering the aims of this study, IFDDC produced positive results through evaluating the model's effects on teachers' skills, beliefs, self-efficacy and practices, lending support to portions of the TOC model. Furthermore, these effects were evident despite individualization of the consultation sessions, suggesting a flexible, data-driven model has the potential to guide teachers. These findings are particularly promising given the underpowered sample and low dosage included in the model. Although limitations were present, specifically surrounding measurement and sample variation, the effects of this study provide preliminary evidence for a low-dosage, easily implemented consultation model in refining teachers' skills, beliefs, self-efficacy, and practices surrounding data-use, with the ultimate goal of improving students' learning.

REFERENCES

Ackerman, D.J. & Barnett, W.S. (2005) *Prepared for kindergarten: What does "readiness" mean? Preschool policy brief.* New Brunswick, NJ: National Institute for Early Education Research.

Administration for Children & Families. (2015) *Head Start early learning outcomes framework [HSELOF]*. Washington, DC: Author. Retrieved from: http://eclkc.ohs.acf.hhs.gov/hslc/hs/sr/approach/pdf/ohs-framework.pdf.

Ajzen, I. (2005) *Attitudes, personality, and behavior* (2nd ed.) New York: Open University Press.

Al Otaiba, S., Connor, C. M., Folsom, J. S., Greulich, L., Meadows, J., & Li, Z. (2011). Assessment data-informed guidance to individualize kindergarten reading instruction: Findings from a cluster-randomized control field trial. *Elementary School Journal, 111* (34), 535-560. doi: 10.1086/659031.

Allen, J. P., Hafen, C. A., Gregory, A. C., Mikami, A. Y., & Pianta, R. (2015). Enhancing secondary school instruction and student achievement: Replication and extension of the My Teaching Partner–Secondary intervention. *Journal of Research on Educational Effectiveness*, *8*(4). doi: 10.1080/19345747.2015.1017680.

Allinder, R. M. (1994). The relationship between efficacy and the instructional practices of special education teachers and consultants. *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children, 17*(2), 86-99. doi: 10.1177/088840649401700203.

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of experimental child psychology, 106*(1), 20-29. doi:10.1016/j.jecp.2009.11.003.

American Federation of Teachers (1990). *Standards for teacher competence in educational assessment of students*. Washington, DC: National Education Association.

Anderson, C. S. (1982). The search for school climate: A review of the research. *Review of Educational Research, 52*(3), 368-420. doi: 10.3102/00346543052003368.

Artman-Meeker, K., Hemmeter, M. L., & Snyder, P. (2014). Effects of distance coaching on teachers' use of pyramid model practices: A pilot study. *Infants & Young Children, 27*(4), 325-344. doi: 10.1097/IYC.0000000000000016.

Ashton, P.T., & Webb, R.B. (1986). *Making a difference: Teachers' sense of efficacy and student achievement*. White Plains, NY: Longman.

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher, 36*(5), 258-267. doi: 10.3102/0013189X07306523.

Badian, N.A. (1988). The prediction of good and poor reading before kindergarten entry: A nine-year follow up. *Journal of Learning Disabilities, 21*(2). doi: 10.1177/002221948802100207.

Baker, S., & Smith, S. (1999). Starting off on the right foot: The influence of four principles of professional development in improving literacy instruction in two kindergarten programs. *Learning Disabilities Research & Practice, 14*(4), 239-253. doi: 10.1207/sldrp1404_5.

Ball, C., & Gettinger, M. (2009). Monitoring children's growth in early literacy skills: Effects of feedback on performance and classroom environments. *Education and Treatment of Children, 32*, 189-212. doi: 10.1353/etc.0.0055.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational psychologist, 28*(2), 117-148. doi: 10.1207/s15326985ep2802_3.

Barkley, R. A. (2006). Primary symptoms, diagnostic criteria, prevalence, and gender differences. *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment, 3*, 76-121. New York: Guilford Press.

Baroody, A.J., Eiland, M., & Thompson, B. (2009). Fostering at-risk preschoolers' number sense. *Early Education & Development. 20*(1), 49. doi: 10.1080/10409280802206619.

Batten, M. (1991). Identifications and development of teachers' professional craft knowledge. In H.W. Kam & R.Y.L. Wong (Eds.), *Improving the quality of the teaching profession* (pp. 295-305). Singapore: The Institute of Education.

Bear, D.R., Invernizzi, M., Templeton, S., & Johnston, F. (2011). *Words their way: Word study for phonics, vocabulary, and spelling instruction* (5th ed.). New York: Pearson.

Belzer, A. (2005). Improving professional development systems: Recommendations from the Pennsylvania adult basic and literacy education professional development system evaluation. *Adult Basic Education, 15*(1), 33.

Bergan, J.R. (2008). *Teaching writing in kindergarten; A structured approach to daily writing that helps every child becomes a confidence, capable writer*. New York, NY: Scholastic Teaching Resources.

Bergan, J.R., & Schnaps, A. (1983). A model for instructional consultation. In J. Alpert & J. Meyers (Eds.), *Training in consultation* (pp.104-119). Springfield, IL: Thomas.

Bernhardt, V.L. (2003). No schools left behind. *Educational Leadership, 60* (5), 26-30.

Bernhardt, V.L. (2013). *Data analysis for continuous school improvement*. New York, NY: Routledge.

Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J. A., Greenberg, M. T., ... & Gill, S. (2008). Promoting academic and social-emotional school readiness: The Head Start REDI program. *Child Development, 79*(6), 1802-1817. doi: 10.1111/j.1467-8624.2008.01227.x.

Bingimlas, K. & Hanrahan, M. (2010). The relationship between teachers' beliefs and their practice: How literature can inform science education reformers and researchers. In M.F. Taspar & G. Cakmakci (Eds.), *Contemporary Science Education Research: International perspectives* (pp. 415-422). Ankara, Turkey: Pegem Akademi.

Black, P. & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 1(*1). doi: 10.1007/s11092-008-9068-5.

Blackwell-Bullock, R., Invernizzi, M., Drake, E. A., & Howell, J. L. (2009). Concept of word in text: An integral literacy skill. *Reading In Virginia, 31*, 30-35.

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist, 57*(2), 111-135. doi: 10.1037/0003-066X.57.2.111.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78*(2), 647-663. doi: 10.1037/0003-066X.57.2.111.

Bodovski, K., & Farkas, G. (2007). Do instructional practices contribute to inequality in achievement? The case of mathematics instruction in kindergarten. *Journal of Early Childhood Research, 5*(3), 301-322. doi: 10.1177/1476718X07080476.

Bollen, K.A. (1990). Overall fit in covariance structure models: Two types of sample size effects. *Psychological Bulletin, 107*, 256-259. doi: 10.1037/0033-2909.107.2.256.

Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. *American Educational Research Journal, 42*(2), 231–268. doi: 10.3102/00028312042002231.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher, 33*(8), 3–15. doi: 10.3102/0013189X033008003.

Boudett, K.P., City, E.A., & Murnane, R.J. (2005). *Data wise: A step-by-step guide to using assessment results to improve teaching and learning*. Cambridge, Massachusetts: Harvard Education Press.

Bowen, G. M., & Roth, W. M. (2005). Data and graph interpretation practices among preservice science teachers*. Journal of research in science teaching, 42*(10), 1063-1088. doi: 10.1002/tea.20086.

Bowles, R. P., Pentimonti, J. M., Gerde, H. K., & Montroy, J. J. (2013). Item response analysis of uppercase and lowercase letter name knowledge. *Journal of Psychoeducational Assessment, 32*(2), 146-156. doi: 10.1177/0734282913490266.

Bowman, B., Donovan, S., & Burns, M. S. (2000). *Eager to learn: Educating our preschoolers*. Washington, DC: National Academy Press.

Boyer, E.L. (1991) *Ready to learn: A mandate for the nation*, Carnegie Foundation for the Advancement of Teaching, Princeton, NJ.

Bredekamp, S., & Copple, S. (1997). *Developmentally appropriate practices in early childhood programs* (revised ed.). Washington, DC: National Association for the Education of Young Children.

Brighton, C. M. (2001). *Internal factors that influence teacher change: teachers' beliefs and conceptions* (Doctoral Dissertation). Retrieved from University of Virginia, ProQuest, UMI Microform. (UMI number: 3017282).

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Bronfenbrenner, U. (1990). Who cares for children? *Research and Clinical Center for Child Development*, 12, 27-40.

Bronfenbrenner, U., & Morris, P.A. (1998). The ecology of developmental processes. In W. Damon & R.M. Learner (Eds.) *Handbook of child psychology: Vol. Q. Theoretical models of human development* (5th ed., pp. 993-1029). New York: John Wiley & Sons.

Bronson, D.S. (2006). Developing social emotional competence. In D.F. Gullo (Ed.) *K Today: Teaching and learning in the kindergarten year* (pp. 47-58). Washington, D.C.: National Association for the Education of Young Children.

Bronson, M. B., Goodson, B. D., Layzer, J. I., & Love, J. M. (1990). *Child Behavior Rating Scale*. Cambridge, MA: Abt Associates.

Brown, G. T. L. (2004a). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Policy, Principles and Practice, 11*(3), 301-318. doi: 10.1080/0969594042000304609.

Brown, G. T. L. (2006). Teachers’ conceptions of assessment: Validation of an abridged instrument. *Psychological Reports, 99,* 166-190. doi: 10.2466/PRO.99.1.166-170.

Brown, G. T. L. (2009). Teachers’ self-reported assessment practices and conceptions: Using structural equation modelling to examine measurement and structural models. In T. Teo & M. S. Khine (Eds.), *Structural equation modelling in educational research: Concepts and applications* (pp. 243–266). Rotterdam, NL: Sense Publishers.

Brown, G. T. L. (2015). The impact of an assessment policy upon teachers’ self-reported assessment beliefs and practices: A quasi-experimental study of Indian teachers in private schools. *International Journal of Education Research, 71*, 50-64. doi: 10.1016/j.ijer.2015.03.001.

Brown, G. T. L., Kennedy, K.J., Fok, P.K., Chan, J.K.S., & Yu, W.M. (2009). Assessment for improvement: Understanding Hong Kong teachers’ conceptions and practices of assessment. *Assessment in Education: Principles, Policy, and Practice, 16*(3), 347-363. doi: 10.1080/09695940903319737.

Brown, G. T. L., Lake, R., & Matters, G. (2011). Queensland teachers’ conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching and Teacher Education, 27*(1), 210–220. doi: 10.1016/j.tate.2010.08.003.

Brown, G.T.L., Chaundry, H., & Dhamija, R. (2015). The impact of an assessment policy upon teachers’ self-reported assessment beliefs and practices: A quasi-experimental study of Indian teachers in private schools. *International Journal of Education Research, 71, 50*-64. doi: 10.1016/j.ijer.2015.03.001.

Brown, K. L. (2003). From teacher-centered to learner-centered curriculum: Improving learning in diverse classrooms. *Education, 124*(1), 49.

Brown, K. W., Ryan, R. M., & Creswell, J. D. (2007). Mindfulness: Theoretical foundations and evidence for its salutary effects. *Psychological Inquiry, 18*(4), 211-237. doi: 10.1080/10478400701598298.

Brunner, C., Fasca, C., Heinze, J., Honey, M., Light, D., Mardinach, E., & Wexler, D. (2005). Linking data and learning: The Grow Network study. *Journal of*

Education for Students Placed at Risk, 10(3), 241-267. doi: 10.1207/s15327671espr1003_2.

Bryant, D. M., Clifford, R. M., & Peisner, E. S. (1991). Best practices for beginners: Developmental appropriateness in kindergarten. *American Educational Research Journal, 28*(4), 783-803. doi: 10.3102/00028312028004783.

Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly, 25*(2), 166-176. doi: 10.1016/j.ecresq.2009.10.004.

Bus, A. G., & van IJzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology, 91*(3), 403.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365-376. doi:10.1038/nrn3475.

Cabell, S. Q., & Downer, J. T. (2011). Improving preschoolers' language and literacy skills through web-mediated professional development. *NHSA dialog, 14*(4), 316-322. doi: 10.1080/15240754.2011.613130.

Calderhead, J. (1996). Teachers: beliefs and knowledge. In D.C. Berliner & R.C. Calfee (Eds.), *Handbook of educational psychology* (pp.709-725). New York: Simon & Schuster Macmillan.

Campbell, M. K., Thomson, S., Ramsay, C. R., MacLennan, G. S., & Grimshaw, J. M. (2004). Sample size calculator for cluster randomized trials. *Computers in Biology and Medicine*, 34(2), 113-125. doi:10.1016/S0010-4825(03)00039-8.

Caplan, G. (1970). *The theory and practice of mental health consultation*. New York: Basic Books.

Carlson, D., Borman, G. D., & Robinson, M. (2011). A multistate district-level cluster randomized trial of the impact of data-driven reform on reading and mathematics achievement. *Educational Evaluation and Policy Analysis, 33*(3), 378-398. doi: 10.3102/0162373711412765.

Center for Public Education. (2012). *The United States of education: The changing demographics of the United States and their schools*. Washington, DC: Author. Retrieved from: http://www.centerforpubliceducation.org/You-May-Also-Be-

Interested-In-landing-page-level/Organizing-a-School-YMABI/The-United-States-of-education-The-changing-demographics-of-the-United-States-and-their-schools.html.

Chalmers, T. C., Smith, H., Blackburn, B., Silverman, B., Schroeder, B., Reitman, D., & Ambroz, A. (1981). A method for assessing the quality of a randomized control trial. *Controlled Clinical Trials, 2*(1), 31-49.

Chard, D. J., Baker, S. K., Clarke, B., Jungjohann, K., Davis, K., & Smolkowski, K. (2008). Preventing early mathematics difficulties: The feasibility of a rigorous kindergarten mathematics curriculum. *Learning Disability Quarterly, 31*(1), 11-20. doi: 10.2307/30035522.

Chick, H., & Pierce, R. (2013). The statistical literacy needed to interpret school assessment data. *Mathematics Teacher Education and Development, 15*(2), 1-20.

Choi, C. C. (1999). Public examinations in Hong Kong. *Assessment in Education: Policy, Principles and Practice, 6*(3), 405-417.

Cizek, G. J., Fitzgerald, S. M., & Rachor, R. A. (1995). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment, 3*(2), 159-179.

Clark, C. M., & Peterson, P. L. (1986). Teachers' thought processes. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 255-296). New York: Macmillan.

Clements, D. (2001). Mathematics in the preschool. *Teaching Children Mathematics, 7*, 270–275.

Clements, D., Sarama, J., & DiBiase, A. M. (Eds.). (2004). *Engaging young children in mathematics: Findings of the 2000 national conference on standards for preschool and kindergarten mathematics education*. Mahwah, NJ: Erlbaum.

Clements, D.H., & Sarama, J. (2003). *DLM Express math resource package*. Columbis, OH: SRA/McGraw-Hill.

Coates, B., & Hartup, W.W. (1969). Age and verbalization in observational learning. *Developmental Psychology, 1*(5), 556-562.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). New York, NY: Academic Press.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155.

Cohen, J. (1995). *The earth is round* (p<. 05). New York, NY: Rejoinder.

Conroy, M. A., Sutherland, K. S., Algina, J. J., Wilson, R. E., Martinez, J. R., & Whalon, K. J. (2015). Measuring teacher implementation of the BEST in CLASS intervention program and corollary child outcomes. *Journal of Emotional and Behavioral Disorders, 23*(3), 144-155. doi: 10.1177/1063426614532949.

Cook, B. G. (2002). Inclusive attitudes, strengths, and weaknesses of pre-service general educators enrolled in a curriculum infusion teacher preparation program. *Teacher Education and Special Education*, 25, 262–277. doi: 10.1177/088840640202500306.

Cooper, D.H. & Farran, D.C. (1988). Behavioral risk factors in kindergarten. *Early Childhood Research Quarterly, 3*, 1-19.

Cooper, P.M. (2009). *Hugging the middle: How teachers teach in an era of testing and accountability*. New York, NY: Teachers College Press.

Copley, J. V. (2004). The early childhood collaborative: A professional development model to communicate and implement the standards. In D. H. Clements & J. Sarama (Eds.), *Engaging young children in Mathematics: Standards for early childhood mathematics education* (pp. 401-414). Mahwah, NJ: Lawrence Erlbaum

Copple, C., Bredekamp, S., Koralek, & Charner, K. (2014) *Developmentally appropriate practice: Focus on kindergarten*. Washington, DC: National Association of the Education of Young Children.

Costenbader, V., Rohrer, A. M., & DiFonzo, N. (2000). Kindergarten screening: A survey of current practice. *Psychology in the Schools, 37*(4), 323-332. doi: 10.1002/1520-6807(200007).

Council of Chief State School Officers [CCSSO]. (2000a). *Annual survey of state student assessment programs*. Washington, D.C.: Author.

Council of Chief State School Officers [CCSSO]. (2000b). *Key state education policies on K–12 education: 2000*. Washington, DC. Retrieved from: http://www.ccsso.org/pdfs/KeyState2000.pdf.

Cozby, P.C. (2001). *Measurement concepts. Methods in behavioral research* (7th ed.). California: Mayfield Publishing Company.

Crockenberg, S., Jackson, S., & Langrock, A.M. (1996). Autonomy and goal attainment: parenting, gender and children's social competence. In M. Killen (Ed.) *Children's* *autonomy, social competence, and interactions with adults and other children: Exploring connection and consequences* (pp. 41-55). San Francisco; Jossey-Bass.

Cross, C.T., Woods, T.A., & Schweingruber, H. (2009). *Mathematics learning in early childhood: Paths towards excellence and equity.* Washington, DC: The National Academies Press.

Cryan, J.R., Sheehan, R., Wiechel, J., & Bandy-Hedden, I.G. (1992). Success outcomes of full-day kindergarten: More positive behavior and increased achievement in the years after. *Early Childhood Research Quarterly, 7(*2), 197-203. doi: 10.1016/0885-2006(92)9004-I.

Cuban, L. (1983). How did teachers teach, 1890–1980. *Theory into Practice, 22* (3), 160–165.

Cudeck, R. (2000). Exploratory factor analysis. In H.E.A. Tinsley & S.D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling,* (pp. 265-296). Cambridge: MA: Academic Press.

Curcio, F. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education, 18*, 382–393.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clinical Psychology Review*, *18*(1), 23-45. doi: 10.1016/S0272-7358(97)00043-3.

Datnow, A., Park, V., & Wohlstetter, P. (2007). *Achieving with data: How high-performing school systems use data to improve instruction for elementary students*. Los Angeles: University of Southern California, Rossier School of Education, Center on Educational Governance.

Davies, S., Janus, M., Duku, E., & Gaskin, A. (2016). Using the early development instrument to examine cognitive and non-cognitive school readiness and elementary student achievement. *Early Childhood Research Quarterly, 35*, 63-75. doi:10.1016/j.ecresq.2015.10.002.

Davis, J. & Bauman, K. (2013). *School enrollment in the United States: 2011*. United States Census Bureau, Department of Commerce, Economics and Statistics Administration. Retrieved from: https://www.census.gov/prod/2013pubs/p20-571.pdf.

Davis, W.W. (2001). Design and analysis of cluster randomization trials in health research. *Journal of the American Statistical Association, 96*(456), 1529-1529. doi: 10.1198/jasa.2001.s433.

Day, K. C., & Day, H. D. (1984). Kindergarten knowledge of print conventions and later school achievement: A five-year follow-up. *Psychology in the Schools*, *21*(3), 393-396.

De Feyter, J. J., & Winsler, A. (2009). The early developmental competencies and school readiness of low-income, immigrant children: Influences of generation, race/ethnicity, and national origins. *Early Childhood Research Quarterly, 24*(4), 411-431. doi:10.1016/j.ecresq.2009.07.004.

DeBacker, T. K., & Crowson, H. M. (2006). Influences on cognitive engagement: Epistemological beliefs and need for closure. *British Journal of Educational psychology, 76*, 535-551. doi: 10.1348/000709905X53138.

Deci, E.L., Koestner, R. & Ryan, R.M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*, 627-668. doi: 10.1037/0033-2909.125.6.627.

Dee, T., & Jacob, B. (2009). The impact of No Child Left Behind on student achievement. *National Bureau of Economic Research,* NBER Working Paper No. 15531.

Demaray, M. K., Ruffalo, S. L., Carlson, J., Busse, R. T., & Olson, A. E. (1995). Social skills assessment: A comparative evaluation of six published rating scales. *School Psychology Review, 24*, 648-671.

Dempster, F.N. (1991). Synthesis of research on reviews and tests. *Educational Leadership, 48*(7), 71-76.

Desimone, L. M. (2009). Improving impact studies of teacher's professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181–199. doi: 10.3102/0013189X08331140.

Dhillon, S., Darrow, C., & Meyers, C.V. (2014). Introduction to implementation fidelity. In C.V. Meyers & C. Brandt (Eds.), *Implementation fidelity in education research: Designer and evaluator considerations* (pp.8-24). New York: Routledge.

Dimmitt, C., Carey, J. C., & Hatch, T. (2007). Evidence-based school counseling; Making a difference with data-driven practices. Thousand Oaks, CA: Corwin Press.

Dougherty, A. M. (2009). *Psychological consultation and collaboration in school and community settings* (5th ed.). Belmont, CA: Brooks/Cole, Cengage Learning.

Downer, J. T., Kraft-Sayre, M. E., & Pianta, R. C. (2009). Ongoing, web-mediated professional development focused on teacher–child interactions: early childhood educators' usage rates and self-reported satisfaction. *Early Education and Development, 20*(2), 321-345. doi: 10.1080/10409280802595425.

Downer, J. T., Locasale-Crouch, J., Hamre, B., & Pianta, R. (2009). Teacher characteristics associated with responsiveness and exposure to consultation and online professional development resources. *Early Education and Development, 20*(3), 431-455. doi: 10.1080/10409280802595425.

Downing, S.M. (2003). Validity: On the meaningful interpretation of assessment data. *Medical Education, 3*7, 830-837. doi: 10.1046/j.1365-2923.2003.01594.x.

Dumas, X., Lynch, A.M., Laughlin, H.E. Smith E.P., & Prinz, R.J. (2001). Promoting intervention fidelity: Conceptual issues, methods, and preliminary results from the EARLY ALLIANCE prevention trial. *American Journal of Preventive Medecine, 20*(1S), 38-47. doi: 10.1016/S0749-3797(00)00272-5.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A., C., Klebanov, P., Pagani, L., …. Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*, 1428-1446. doi: 10.1037/0012-1649.43.6.1428.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82*(1), 405-432. doi: 10.1111/j.1467-8624.2010.01564.x.

Dusenbury, L., Brannigan, R., Falco, M. and Hansen, W. B. 2003. A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research, 18*, 237–256. doi: 10.1093/her/18.2.237.

Dusenbury, L., Weissberg, R.P., Green, P., & Domitrovich, C. (2014). *State standards to advance social and emotional learning: Findings from CASEL's state scan of social and emotional learning standards, preschool through high school*. Chicago, IL: Collaborative for Academic, Social, and Emotional Learning. Retrieved from: https://static1.squarespace.com/static/513f79f9e4b05ce7b70e9673/t/52df31e2e4b08544727c3d88/1390359010742/CASEL+Brief+on+State+Standards--January+2014.pdf.

Dyson, N. I., Jordan, N. C., & Glutting, J. (2013). A number sense intervention for low-income kindergartners at risk for mathematics difficulties. *Journal of Learning Disabilities, 46(*2), 166-181. doi: 10.1177/0022219411410233.

Early, D., Barbarian, O., Bryant, D., Burchinal, M., Chang, F., Clifford, R. et al., (2005). *Prekindergarten in eleven states: NCEDL's multistate study of pre-kindergarten and study of state wide early education programs*. Retrieved from http://fpg.unc.edu/sites/fpg.unc.edu/files/resources/reports-and-policy-briefs/NCEDL_PreK-in-Eleven-States_Working-Paper_2005.pdf.

Easton, J. Q., & Luppescu, S. (2004). *Teacher and principal responses to the grow network: A preliminary evaluation of use in Chicago public elementary schools*. Chicago, IL: Consortium on Chicago School Research.

Ebbeler, J., Poortman, C. L., Schildkamp, K., & Pieters, J. M. (2016). Effects of a data use intervention on educators' use of knowledge and skills. *Studies in Educational Evaluation, 48,* 19-31. doi: 10.1016/j.stueduc.2015.11.002.

Ehri, L. C. (1979). Linguistic insight: Threshold of reading acquisition. In T. G. Waller & G. F. MacKinnon (Eds.), *Reading research: Advances in theory and practice* (Vol 1, pp. 63–111). New York: Academic Press.

Ehri, L. C., & Roberts, T. (2005). The roots of learning to read and write: Acquisition of letters and phonemic awareness. In D. Dickinson & S. B. Neuman (Eds.), *Handbook of early literacy research*. New York: Guilford Press.

Eisenberg, N., Valiente, C., & Eggum, N. D. (2010). Self-regulation and school readiness. *Early Education and Development, 21*(5), 681-698. doi: 10.1080/10409289.2010.497451.

Elder, T. E., & Lubotsky, D. H. (2009). Kindergarten entrance age and children's achievement impacts of state policies, family background, and peers. *Journal of Human Resources, 44*(3), 641-683. doi: 10.3368/jhr.44.3.641.

Eldridge, S. M., Ashby, D., & Kerry, S. (2006). Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. *International journal of epidemiology*, *35*(5), 1292-1300. doi: 10.1093/ije/dyl129.

Eller, R. G., Pappas, C. C., & Brown, E. (1988). The lexical development of kindergarteners: Learning from written context. *Journal of Literacy Research, 20*(1), 5-24.

Elwood, J. & Klenowski, V. (2002). Creating communities of shared practice: The challenges of assessment use in learning and teaching. *Assessment & Evaluation in Higher Education, 27*(3), 243-256. doi: 10.1080/0260293022013860.

Entwistle, N., & Ramsden, P. (2015). *Understanding student learning.* New York: Routledge.

Fabes, R. A., Gaertner, B. M., Popp, T.K., McCartney, K., & Phillips, D. (2006). Getting along with others: Social competence in early childhood. In K. McCartney & D. Phillips (Eds.) *Blackwell handbook of early childhood development.* Malden, MA: Blackwell Publishing.

Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational Research, 38*(1), 47-65. doi: 10.1080/0013188960380104.

Fantuzzo, J., Bulotsky-Shearer, R., McDermott, P. A., & McWayne, C. (2007). Investigation of dimensions of social-emotional classroom behavior and school readiness for low-income urban preschool children. *School Psychology Review, 36*(1), 44. Retrieved from: http://repository.upenn.edu/gse_pubs/124.

Feistritzer, C.E. (2011). *Profile of teachers in the U.S. 2011.* Washington, DC: National Center for Education Information. Retrieved from: http://www.edweek.org/media/pot2011final-blog.pdf.

Figlio, D. & Loeb, S. (2011). School accountability. In E.A. Hanushek, S. Machin, & L. Woessmann, (Eds.), *Handbooks in economics* (Vol 3, pp. 383-421). Netherlands: North-Holland.

Firestone, W.A., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: the effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis, 20*(2), 95-113. doi: 10.3102/01623737020002095.

Fisher, L.D., Dixon, D.O., Herson, J., Frankowski, R.K., Hearron, M.S., & Peace, K.E. (1990) Intention to treat in clinical trials. In K.E. Peace (Ed.) *Statistical issues in drug research and development* (pp. 331–50). New York: Marcel Dekker.

Fives, H. & Buehl, M.M. (2012). Spring cleaning for the messy construct of teachers' beliefs: What are they? Which have been examined? What can they tell us? In K.R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook: Individual differences and cultural and contextual factors* (Vol. 2, pp. 471–499). Washington, DC: American Psychological Association.

Franks, A., Kelder, S., Dino, G. A., Horn, K. A., Gortmaker, S. L., & Wiecha, J. L. (2015). School-based programs: lessons learned from CATCH, Planet Health, and Not-On-Tobacco. *School Nutrition and Activity: Impacts on Well-Being, 147.* doi: 10.1201/b18227-11.

Fryer Jr, R. G., & Levitt, S. D. (2004). Understanding the black-white test score gap in the first two years of school. *Review of Economics and Statistics, 86*(2), 447-464. doi: 10.3386/w8975.

Fuchs, L.S., Fuchs, D., Hamlett, C.K., & Ferguson, C. (1992). Effects of expert system consultation within curriculum-based measurement using a reading maze task. *Exceptional Children, 58,* 436-450. doi: 10.1177/001440299205800507.

Fuchs, L.S., Fuchs, D., Hamlett, C.K., & 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*, 617-641. doi: 10.3102/00028312028003617.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., Phillips, N.B., Karns, K., & Dutka, S. (1997). Enhancing students' helping behavior during peer-mediated instruction with conceptual mathematical explanations. *Elementary School Journal, 97,* 223-249. doi: 10.1086/461863.

Fuchs, L.S., Fuchs, D., Phillips, N.B., Hamlett, C.L., & Karns, K. (1995). Acquisition and transfer effects of classwide peer-assisted learning strategies in mathematics for students with varying learning histories. *School Psychology Review, 24*, 604-620.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915-945. doi: 10.3102/00028312038004915.

Gershenson, R. A., Lyon, A. R., & Budd, K. S. (2010). Promoting positive interactions in the classroom: Adapting parent-child interaction therapy as a universal prevention program. *Education and Treatment of Children, 33*(2), 261-287. doi:10.1353/etc.0.0092.

Gibson, E., & Levin, E. (1975). *The psychology of reading*. Cambridge, MA: MIT Press.

Gijbels, D., & Dochy, F. (2006). Students' assessment preferences and approaches to learning: can formative assessment make a difference? *Educational Studies, 32*(4), 399-409. doi: 10.1080/03055690600850354.

Gijbels, D., Van de Watering, G, Dochy, F., & Van den Bossche. (2005). The relationship between students' approaches to learning and the assessment of learning outcomes. *European Journal of Psychology of Education, 4*, 327-341. doi: 10.1007/BF03173560.

Ginsburg, H. (1989). *Children's arithmetic*. Austin, TX: Pro-Ed.

Ginsburg, H.P., Galanter, M., & Morgenlander, M. (2004). B*ig Math for Little Kidsworkshops.* New York: Teachers College Innovations.

Ginsburg, H. P. & Pappas, S. (2016). Invitation to the birthday party: rationale and description. *ZDM Mathematics Education* , *48*(7), 947-960. doi: 10.1007/s11858-016-0818-4.

Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. Philadelphia, PA: Psychology Press.

Goddard, R.D. & Goddard, Y.L. (2001). A multilevel analysis of teacher and collective efficacy. *Teaching and Teacher Education, 17*, 807–818. doi:10.1016/S0742-051X(01)00032-4.

Goldstein, J., Eastwood, M., & Behuniak, P. (2014) Can teacher ratings of students' skills at kindergarten entry predict kindergarten retention? *The Journal of Education Research, 107*(3), 217-229. doi: 10.1080/00220671.2013.788986.

Goldstein, L. S. (2007). Beyond the DAP versus standards dilemma: Examining the unforgiving complexity of kindergarten teaching in the United States. *Early Childhood Research Quarterly, 22*(1), 39-54. doi:10.1016/j.ecresq.2006.08.001.

Gorsuch, R. L. (1988). Exploratory factor analysis. In *Handbook of multivariate experimental psychology* (pp. 231-258). New York, NY: Springer US.

Grant, S.G. (2003). *History lessons: Teaching, learning, and testing in U.S. high school classrooms.* Mahwah, NJ: Lawrence Erlbaum.

Gravois, T. A., & Rosenfield, S. (2002). A multidimensional framework for the evaluation of instructional consultation teams. *Journal of Applied School Psychology, 19,* 5-29. doi: 10.1300/J008v19n01_02.

Gravois, T. A., & Rosenfield, S. (2006). Impact of instructional consultation teams on the disproportionate referral and placement of minority students in special education. *Remedial and Special Education, 27*(1), 42-52. doi: 10.1177/07419325060270010501.

Gredler, G. R. (1997). Issues in early childhood screening and assessment. *Psychology in the Schools, 34(*2), 99-106. doi: 10.1002/(SICI)1520-6807(199704).

Greenberg, M. T., Domitrovich, C. E., Graczyk, P. A., & Zins, J. E. (2003). The study of implementation in school-based preventive interventions: Theory, research, and practice. *Promotion of mental health and Prevention of Mental and Behavioral Disorder, 3*, 1-62.

Grow Network. (2004). *About the Grow Network*. Retrieved from http://grownetwork.com/.

Gullo, D. F., & Burton, C. B. (1992). Age of entry, preschool experience, and sex as antecedents of academic readiness in kindergarten. *Early Childhood Research Quarterly, 7*(2), 175-186. doi: 10.1016/0885-2006(92)90003-H.

Gullo, D.F. & Hughes, K. (2011). Reclaiming kindergarten: Part I. Questions about theory and practice. *Early Childhood Education Journal, 38*(5), 323-328. doi: 10.1007/s10643-010-0429-6.

Gupta, S. K. (2011). Intention-to-treat concept: A review. *Perspectives in Clinical Research, 2(*3), 109-112. doi:10.4103/2229-3485.83221.

Guthrie. J.T. (2002). Preparing students for high-stakes test taking in reading. In A.E. Farstrup, & S.J. Samuels (Eds.), *What research has to say about reading instruction* (pp.370-391). Newark, DE: International Reading Association.

Hall, N. (1987). *Emergence of literacy*. Portsmouth, NH: Heinemann.

Hallinger, P., Bickman, L., & Davis, K. (1996). School context, principal leadership, and student reading achievement. *Elementary School Journal, 96*(5), 527-549. doi: 10.1086/461843.

Hamilton, L.S., Halverson, R., Jackson, S. S., Mandinach, E., Supovitz, J. A., & Wayman, J. C. (2009). *Using student achievement data to support instructional decision making*. Washington, DC: Institute of Education Sciences and the National Center for Education Evaluation.

Hamilton, L.S., Stecher, B.M., Marsh, J.A., McCombs, J.S., Robyn, A., Russell, J.L., et al. (2007). *Standards-based accountability under no child left behind: Experiences of teachers and administrators in three states*. Santa Monica, CA: RAND Education.

Hamre, B. K., & Pianta, R. C. (2007). Learning opportunities in preschool and early elementary classrooms. In R. Pianta, M. Cox, & K. Snow (Eds.), *School readiness & the transition to kindergarten in the era of accountability* (pp. 49-84). Baltimore: Brookes.

Hanover Research. (2013). *Kindergarten entry assessments: Practices and policy*. Washington, DC: Author. Retrieved from: http://www.hanoverresearch.com/media/Kindergarten-Entry-Assessments-Practices-and-Policies.pdf.

Hart, L. C. (2002). Preservice teachers' beliefs and practices after participating in an integrated content/method course. *School Science and Mathematics, 102*(1), 4-14. doi: 10.1111/j.1949-8594.2002.tb18191.x.

Hart, L. C., Ladd, G.W., & Burleson, B.R. (1990). Children's expectations of the outcomes of social strategies: relations with sociometric status and maternal disciplinary styles. *Child Development, 61*, 121-137. doi: 10.1111/j.1467-8624.1990.tb02766.x.

Hatch, J.A. (2005). *Teaching in the new kindergarten.* Clifton Park, NY: Thomson Delmar Learning.

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs, 76,* 408-420. doi: 10.1080/03637750903310360.

Hayes, K., & Creange, R. (2001). *Classroom Routines that really work for prek and kindergarten*. New York, NY: Scholastic Inc.

Heaton, J.B. (1975). *Writing English language test*. London: Longman.

Heaviside, S. (1993). *Public school kindergarten teachers' views on children's readiness for school. Contractor report*. Washington, DC: US Government Printing Office.

Heckman, J. J. (2008). Schools, skills, and synapses. *Economic Inquiry, 46*(3), 289. doi: 10.1111/j.1465-7295.2008.00163.x.

Herman, J. L. (2004). The effects of testing on instruction. In Susan H. Fuhrman and Richard F. Elmore, (Eds.). *Redesigning accountability systems for education*, 141-166. New York: Teachers College Press.

Heroman, C. & Copple, C. (2006). Teaching in the kindergarten year. In D.F. Gullo (Ed.) *K Today: Teaching and learning in the kindergarten year*, (pp. 59-72). Washington, D.C.: National Association for the Education of Young Children.

Hesterberg, T., Moore, D. S., Monaghan, S., Clipson, A., & Epstein, R. (2005). Bootstrap methods and permutation tests. *Introduction to the Practice of Statistics, 5*, 1-70.

Hicks, O. (1999). A conceptual framework for instructional consultation. *New Directions for Teaching and Learning, 79*, 9-18. doi: 10.1002/tl.7902.

Hill, H. C., Beisiegel, M., & Jacob, R. (2013). Professional development research consensus, crossroads, and challenges. *Educational Researcher, 42(*9), 476-487. doi: 10.3102/0013189X13512674.

Hipp, K.A. & Bredeson, P.V. (1995). Exploring connections between teacher efficacy and principals' leadership behavior. *Journal of School Leadership, 5*,136–150. doi: 1052-6846/95/02/0136-15.

Howes, C. & Matheson, C.C. (1992). Contextual constraints on the concordance of mother-child and teacher-child relationships. *New Directions for Child and Adolescent Development, 57*, 25-40. doi:10.1002/cd.23219925704.

Howes, C. (1997). Children's experiences in center based child care as a function of teacher background and adult: child ratio. *Merrill-Palmer Quarterly, 43*(3), 404-425.

Howes, V. M. (1970). *Individualization of instruction: A teaching strategy*. Washington, DC: MacMillan Publishing.

Hsieh, W. Y., Hemmeter, M. L., McCollum, J. A., & Ostrosky, M. M. (2009). Using coaching to increase preschool teachers' use of emergent literacy teaching strategies. *Early Childhood Research Quarterly, 24*(3), 229-247. doi:10.1016/j.ecresq.2009.03.007.

Huang, F. L., & Invernizzi, M. A. (2012). The association of kindergarten entry age with early literacy outcomes. *The Journal of Educational Research, 105*(6), 431-441 doi: 10.1080/00220671.2012.658456.

Huang, F. L., & Invernizzi, M. A. (2014). Factors associated with lowercase alphabet naming in kindergarteners. *Applied Psycholinguistics, 35*(06), 943-968. doi: 10.1017/S0142716412000604.

Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness, 2(*1), 88-110. doi: 10.1080/19345740802539325.

Hylander, I. (2012). Conceptual change through consultee-centered consultation: A theoretical model. *Consulting Psychology Journal: Practice and Research, 64*(1), 29-45. doi: 10.1037/a0027986.

IBM Corp. (2015). *IBM SPSS Statistics for Windows,* Version 22.0. Armonk, NY: IBM Corp.

Idol, L., Nevin, A., & Paolucci-Whitcomb, P. (1994). *Collaborative consultation*. New York, NY: Pro-Ed.

Ingram, D., Lewis, K. S., & Schroeder, R. G. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record, 106*(6), 1258-1287.

Invernizzi, M. (2003). Concepts, sounds, and the ABCs. *Literacy and Young Children: Research-Based Practices, 140*, 49-72.

Invernizzi, M., Justice, L., Landrum, T. J., & Booker, K. (2004). Early literacy screening in kindergarten: Widespread implementation in Virginia. *Journal of Literacy Research, 36*(4), 479-500. doi: 10.1207/s15548430jlr3604_3.

Invernizzi, M., Swank, L., Juel, C., & Meier, J. (2003). *Phonological awareness literacy screening-kindergarten.* Charlottesville, VA: University Printing.

Isaacs, J.B. (2012). *Starting school at a disadvantage: The school readiness of poor children*. Washington, DC: Center for Children and Families at Brookings.

Janus, M., & Duku, E. (2007). The school entry gap: Socioeconomic, family, and health factors associated with children's school readiness to learn. *Early Education and Development, 18*(3), 375-403. doi: 10.1080/10409280701610796a.

Jennings, P. A. & Greenberg, M. (2009). The Prosocial Classroom: Teacher social and emotional competence in relation to child and classroom outcomes. *Review of Educational Research, 79*, 491–525. doi: 10.3102/00346543083256.

Jeynes, W. H. (2002). Standardized tests and Froebel's original kindergarten model. *Teachers College Record, 108*(10), 1937-1959. doi: 10.1111/j.1467-873X.2008.00427.x.

Jitendra, A. K., Dupuis, D. N., & Zaslofsky, A. (2014). Curriculum-Based Measurement and standards-based mathematics: Monitoring the arithmetic word problem-solving performance of third-grade students at risk for mathematics difficulties. *Learning Disability Quarterly, 37*(4), 241-251. doi: 10.1177/0731948713516766.

Johnson, R. S. (2002). *Using data to close the achievement gap: How to measure equity in our schools*. Thousand Oaks, CA: Corwin Press.

Justice, L. M., Invernizzi, M. A., & Meier, J. D. (2002). Designing and implementing an early literacy screening protocol: Suggestions for the speech-language pathologist. *Language, Speech, and Hearing Services in Schools, 33*, 84-101. doi: 10.1044/0161-1461(2002/007).

Kagan, D. M. (1992). Implications of research on teacher belief. *Educational Psychologist, 27*(1), 65-90. doi: 10.1207/s15326985ep2701_6.

Kagan, S. L., & Kauerz, K. (2006). Making the most of kindergarten—trends and policy issues. In D. F. Gullo (Ed.), *K today: Teaching and learning in the kindergarten year* (pp. 161-170). Washington, DC: National Association for the Education of Young Children.

Kagan, S.L. (1990). Readiness 2000: Rethinking rhetoric and responsibility. *Phi Delta Kappan, 72*, 272-279.

Kaiser, L., Rosenfield, S., & Gravois, T. (2009). Teachers' perception of satisfaction, skill development, and skill application after instructional consultation services. *Journal of Learning Disabilities, 42*(5), 444-457. doi: 10.1177/0022219409339062.

Katz, M. B. (1968). *The irony of early school reform: Educational innovation in mid-nineteenth century Massachusetts*. New York, NY: Teachers College Press.

Kauffman, D., Moore Johnson, S., Kardos, S., Liu, E., & Peske, H. (2002). "Lost at sea": New teachers' experiences with curriculum and assessment. *Teachers College Record, 104*(2), 273-300. doi: 10.1111/1467-9620.00163.

Kelly, A., Downey, C., & Rietdijk, W. (2010). *Data dictatorship and data democracy: understanding professional attitudes to the use of pupil performance data in schools*. South Hamptom, GB: CFBT Education Trust.

Kerr, K.I., Marsh, J.A., Ikemoto, G.S., Darilek, H., & Barney, H. (2006). Strategies to promote data use for instructional improvement: Actions, outcomes, and lessons from three urban districts. *American Journal of Education, 112*(August), 496-520. doi: 10.1086/505057.

Kim, K.H., VanTassel-Baska, J., Bracken, B.A., Feng, A., & Stambaugh, T. (2014). Assessing science reasoning and conceptual understanding in the primary grades using standardized and performance-based assessments. *Journal of Advanced Academics, 25*(1), 47-66. doi: 10.1177/1932202X14520946.

Kinzie, M.B., Whittaker, J.V., Williford, A.P., DeCoster, J., McGuire, P., … & Kilday, C.R. (2014). MyTeachingPartner-Math/Science pre-kindergarten curricula and teacher supports: Associations with children's mathematics and science learning. *Early Childhood Research Quarterly, 29*(4), 586-599. doi: 10.1016/j.ecresq.2014.06.007.

Klassen, R. M., & Chiu, M. M. (2010). Effects on teachers' self-efficacy and job satisfaction: Teacher gender, years of experience, and job stress. *Journal of educational Psychology, 102*(3), 741-756. doi: 10.1037/a0019237.

Knotek, S. E., Rosenfield, S. A., Gravois, T. A., & Babinski, L. M. (2003). The process of fostering consultee development during instructional consultation. *Journal of Educational and Psychological Consultation*, 14, 303-328. doi: 10.1080/10474412.2003.9669491.

Kohlberg, L., & Mayer, R. (1972). Development as the aim of education. *Harvard Educational Review, 42*, 429-496.

Koziol Jr, S. M., & Burns, P. (1986). Teachers' accuracy in self-reporting about instructional practices using a focused self-report inventory. *The Journal of Educational Research, 79*(4), 205-209. doi: 10.1080/00220671.1986.10885678.

Kraemer, H. C., & Kupfer, D. J. (2006). Size of treatment effects and their importance to clinical research and practice. *Biological psychiatry, 59*(11), 990-996. doi: 10.1016/j.biopsych.2005.09.014

Kramer, T. J., Caldarella, P., Christensen, L., & Shatzer, R. H. (2010). Social and emotional learning in the kindergarten classroom: Evaluation of the Strong Start curriculum. *Early Childhood Education Journal, 37*(4), 303-309. doi: 10.1007/s10643-009-0354-8.

Kratochwill, T. R., & Bergan, J. R. (1990). *Behavioral consultation in applied settings*. New York, NY: Plenum Press.

Kretlow, A. G., & Bartholomew, C. C. (2010). Using coaching to improve the fidelity of evidence-based practices: A review of studies. *Teacher Education and Special Education, 33,* 279-299. doi: 10.1177/0888406410371643.

Lachat, M.A., & Smith, S. (2005). Practices that support data use in urban high schools. *Journal of Education for Students Placed at Risk, 10*(3), 333-349. doi: 10.1207/s15327671espr1003_7.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics,* 159-174.

Lee, H. J. (2005). Understanding and assessing preservice teachers' reflective thinking. *Teaching and Teacher Education, 21*(6), 699-715. doi:10.1016/j.tate.2005.05.007.

Lee, V., Dedick, R., & Smith, J. (1991). The effect of the social organization of schools on teachers' efficacy and satisfaction. Sociology of Education, 64, 190–208.

Lee, Y. S. (2016). Psychometric analyses of the Birthday Party. *ZDM Mathematics Education, 48*(7), 961-975. doi: 10.1007/s11858-016-0801-0.

Leithwood, K., Seashore Louis, K., Anderson, S., & Wahlstrom, K. (2004). *Review of research: How leadership influences student learning.* New York, NY: The Wallace Foundation. Retrieved from: http://conservancy.umn.edu/bitstream/handle/11299/2035/CAREI?sequence=1.

Lenth, R. V. (2007). *Post hoc power: tables and commentary*. Iowa City, IA: Department of Statistics and Actuarial Science, University of Iowa.

Lerner, J.S., & Tetlock, P.E. (1999). Accounting for the effects of accountability. *Psychological Bulletin, 125* (2), 255-275.

Lim, S.M., Rodger, S., & Brown, T. (2010). Assessments of learning-related skills and interpersonal skills constructs within early childhood environments in Singapore. *Infant and Child Development, 19*(4), 366-384. doi: 10.1002/icd.673.

Limbos, M. M., & Geva, E. (2001). Accuracy of teacher assessments of second-language students at risk for reading disability. *Journal of Learning Disabilities, 34*(2), 136-151.

Lin, H. L., Lawrence, F. R., & Gorrell, J. (2003). Kindergarten teachers' views of children's readiness for school. *Early Childhood Research Quarterly, 18*(2), 225-237. doi: 10.1016/S0885-2006(03)00028-0.

Lloyd, J. E., & Hertzman, C. (2009). From kindergarten readiness to fourth-grade assessment: Longitudinal analysis with linked population data. *Social Science & Medicine, 68*(1), 111-123. doi:10.1016/j.socscimed.2008.09.063.

Lonigan, C.J., Allan, N.P., & Lerner, M.D. (2011). Assessment of preschool early literacy skills: linking children's educational needs with empirically supported instructional activities. *Psychology in the Schools, 48*(5), 488-501. doi: 10.1002/pits.20569.

Looney, J. (2005). *Formative assessment: Improving learning in secondary classrooms*. Paris, France: Organisation for Economic Cooperation and Development.

Lowman, R. L. (2002). *The handbook of organizational consulting psychology*. Hoboken, NJ: John Wiley & Sons, Inc.

Luft, J. A. (1999). Teachers' salient beliefs about a problem solving demonstration classroom inservice program. *Journal of Science Teacher Education, 36*(2), 141-158.

MacSuga, A.S., & Simonsen, B. (2011). Increasing teachers' use of evidence-based classroom management strategies through consultation: Overview and case studies. *Beyond Behavior, 20*(2), 4-12.

Mandinach, E., Friedman, J. M., & Gummer, E. (2015). Data-driven decision making: Components of the enculturation of data use in education. *Teachers College Record, 117*(4).

Mansour, N. (2009). Science teachers' beliefs and practices: Issues, implications and research agenda. *International Journal of Environmental & Science Education, 4*(1), 25-48

Marsh, H.W., Hau, K-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers of overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*, 320-341. doi: 10.1207/s15328007sem1103_2.

Marsh, J. A. (2012). Interventions promoting educators' use of data: Research insights and gaps. *Teachers College Record, 114*(11), 1-48.

Marsh, J. A., Kerri, K.A., Ikemoto, G.S., Darilek, H., Suttorp, M. … Barney, H. (2005). *The role of districts in fostering instructional improvement: Lessons from three urban districts partnered with the Institute for Learning*. Monograph no. MG-361-EDU. Santa Monica, CA: Rand Corporation.

Marton, F. (1981). Phenomenography- describing conceptions of the world around us. *Instructional Science, 10*, 177-200.

Maryland Department of Education. (2009). *Maryland school readiness report; children entering school ready to learn*. Retrieved from: http://marylandpublicschools.org/NR/rdonlyres/BCFF0F0E-33E5-48DA-8F11-28CF333816C2/19574/MMSR_ReadytoLearn_200809.pdf.

Marzano, R. J., Marzano, J. S., & Pickering, D. (2003). *Classroom management that works: Research-based strategies for every teacher*. Ontario: Elementary Teachers' Federation of Ontario.

Mashburn, A. J., Downer, J. T., Hamre, B. K., Justice, L. M., & Pianta, R. C. (2010). Consultation for teachers and children's language and literacy development during pre-kindergarten. *Applied Developmental Science, 14*(4), 179-196. doi: 10.1080/10888691.2010.516187.

Massell, D. (2001). The theory and practice of using data to build capacity: State and local strategies and their effects. In S. Fuhrman (Ed.), *From capitol to the classroom: Standards-based reform in the states. The 100th yearbook of the National Society for the Study of Education, Part II* (pp. 148-169). Chicago: University of Chicago Press.

Matthews, J.S., Pointz, C.C., & Morrison, F.J. (2009). Early gender differences in self-regulation and academic achievement. *Journal of Educational Psychology, 101*(3), 689-704. doi: 10.1037/a0014240.

Maxwell, K.L. & Clifford, R.M. (2004*). Research in review: School readiness assessment*. Washingon, DC: National Association of the Education of Young Children. Retrieved from: https://www.naeyc.org/files/yc/file/200401/Maxwell.pdf.

May, H. & Robinson, M. A. (2007). *A randomized evaluation of Ohio's Personalized Assessment Reporting System (PARS)*. Philadelphia: University of Pennsylvania Consortium for Policy Research in Education.

McClelland, M. M., & Morrison, F. J. (2003). The emergence of learning-related social skills in preschool children. *Early Childhood Research Quarterly, 18*(2), 206-224.

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly, 21(*4), 471-490. doi:10.1016/j.ecresq.2006.09.003.

McClelland, M.M., Morrison, F.J., & Holmes, D.L. (2000). Children at risk for early academic problems: the role of learning-related social skills, *Early Childhood Research Quarterly, 15*(3). doi: 10.1016/S0885-2006(00)00069-7.

McFall, R. M. (1982). A review and reformulation of the concept of social skills. *Behavioral Assessment,* 8, 3–10.

McMillian, J.H. (1997). *Classroom assessment. Principles and practices for effective instruction.* Allyn & Bacon: Needham Heights, M.A.

McMillian, J.H., Myran, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research, 95*(4), 203-213. doi: 10.1080/00220670209596593.

Meisels, S.J. (1996). Performance in context: Assessing children's achievement at the outset of school. In A.J. Sameroff & M.M. Haith (Eds.), *The five to seven year shift: The age of reason and responsibility* (pp. 410-431). Chicago: University of Chicago Press.

Meisels, S.J. (1999). Assessing readiness. In R.C. Pianta & M.J. Cox (Eds.) *The transition to kindergarten,* (pp. 11-23). Baltimore, Maryland: Paul H. Brookes Publishing Co.

Melnick, S. A., & Meister, D. G. (2008). A comparison of beginning and experienced teachers' concerns. *Educational Research Quarterly, 31*(3), 39-56.

Meyer, J. (1975). Consultee-centered consultation with a teacher as a technique in behavior management. *American Journal of Community Psychology, 3*(2), 111-121.

Middendorf, C. (2007). *Differentiating instruction in kindergarten*. New York, NY: Scholastic Teaching Resources.

Miles, K. H. (1995). Freeing resources for improving schools: A case study of teacher allocation in Boston public schools. *Educational Evaluation and Policy Analysis, 17*, 476-493. doi: 10.3102/01623737017004476.

Miles, K. H., & Darling-Hammond, L. (1998). Rethinking the allocation of teaching resources: Some lessons from high-performing schools. *Educational Evaluation and Policy Analysis, 20*, 9-29. doi: 10.3102/01623737020001009.

Milesi, C., & Gamoran, A. (2006). Effects of class size and instruction on kindergarten achievement. *Educational Evaluation and Policy Analysis, 28*(4), 287-313. doi: 10.3102/01623737028004287.

Mills, G. E. (2000). *Action research: A guide for the teacher researcher*. Upper Saddle River, NJ: Prentice-Hall, Inc.

Moats, L. C. (2000). *Speech to print: Language essentials for teachers*. Baltimore, MD: Paul H. Brookes Publishing,

Moncher, F.J., & Prinz, F.J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, 11, 247–266. doi: 10.1016/0272-7358(91)90103-2.

Monfils, L.F., Firestone, W.A., Hickes, J.E., Martinez, M.C., Schorr, R.Y., & Camilli, G. (2004). Teaching to the test. In W.A. Firestone, R.Y. Schorr, & L.F. Monfils (Eds.) *The ambiguity of teaching to the test: Standards, assessment, and educational reform* (pp. 37-61). Mahwah, NJ: LEA.

Moon, T. R., Brighton, C. M., & Callahan, C. M. (2003). State standardized testing programs: Friend or foe of gifted education?. *Roeper Review, 25*(2), 49-60 doi: 10.1080/02783190309554199.

Moore, D. S., & McCabe, G. P. (1993). *Introduction to the practice of statistics*. New York: Freeman

Morison, P., McLaughlin, M. J., & McDonnell, L. M. (1997). *Educating one and all: Students with disabilities and standards-based reform.* Washington, DC: National Academies Press.

Mulligan, G.M., McCarroll, J.C., Flanagan, K.D., & Potter, D. (2014). *Findings from the first-grade rounds of the early childhood longitudinal study, kindergarten class of 2010-11* (ECLS-K:2011) (NCES 2015-109). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Retrieved from http://nces.ed.gov/pubsearch.

Murphy-Berman, V., Rosell, J., & Wright, G. (1986). Measuring children's attention span: A microcomputer assessment technique. *Journal of Educational Research, 80*, 23-28.

Murray, N. G., Low, B. J., Hollis, C., Cross, A. W., & Davis, S. M. (2007). Coordinated school health programs and academic achievement: A systematic review of the literature. *Journal of School Health, 77*(9), 589-600. doi: 10.1111/j.1746-1561.2007.00238.x.

Muthén, L. K., & Muthén, B. O. (1998-2011). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.

National Association for the Education of Young Children (NAEYC). (1988). NAEYC position statement on standardized testing of young children 3 through 8 years of age. *Young Children, 43*, 42-47.

National Association for the Education of Young Children (NAEYC). (2003). *NAEYC position statement on early childhood curriculum, assessment, and program evaluation*. Retrieved from: https://www.naeyc.org/files/naeyc/file/positions/pscape.pdf.

National Association for the Education of Young Children/National Council of Teacher of Mathematics (NAEYC/NCTM). (2002/2010). *Early childhood mathematics: Promoting good beginnings*. Washington, DC: National Association for the Education of Young Children.

National Commission on Excellence in Education. (1983). *A nation at risk*. Washington, DC: U.S. Department of Education.

National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. Reston, VA: Author

National Education Goals Panel. (1991). *The national education goals report*. Washington, DC: Author.

National Governors Association Center for Best Practices/Council of Chief State School Officers (NGACBP/CCSSO). (2010a). *Common core state standards for English language & literacy in history/social studies, science, and technical subjects*.

Washington, DC: National Governors Association Center for Best Practices, Council of Chief State School Officers.

National Governors Association Center for Best Practices/Council of Chief State School Officers (NGACBP/CCSSO). (2010b). *Common core state standards for mathematics*. Washington, DC: National Governors Association Center for Best Practices, Council of Chief State School Officers.

National Mathematics Advisory Panel [NMAP]. (2008). *Final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education Retrieved from: http://www.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf.

National Reading Panel Report. (2000). *Teaching children to read.* Washington, DC: National Institute of Child Health and Development.

Neal, D., & Schanzenbach, D. (2010). Left behind by design: Proficiency counts and test-based accountability. *Review of Economic Statistics. 92*(2), 263–283. doi: 10.1162/rest.2010.12318.

Neef, N. A., Iwata, B. A., & Page, T. J. (1980). The effects of interspersal training versus high-density reinforcement on spelling acquisition and retention. *Journal of Applied Behavior Analysis, 13*(1), 153-158.

Nelson, J. R., Martella, R. M., & Marchand-Martella, N. (2002). Maximizing student learning the effects of a comprehensive school-based program for preventing problem behaviors. *Journal of Emotional and Behavioral Disorders, 10*(3), 136-148. doi: 10.1177/10634266020100030201.

Neuman, S. B., & Roskos, K. (2005). The state of state pre-kindergarten standards. *Early Childhood Research Quarterly, 20*(2), 125-145. doi:10.1016/j.ecresq.2005.04.010.

Newmann, F. M., Smith, B., Allensworth, E., & Bryk, A. S. (2001). Instructional program coherence: What it is and why it should guide school improvement policy. *Educational evaluation and policy analysis, 23*(4), 297-321. doi: 10.3102/01623737023004297.

Niemeyer, J., & C. Scott-Little. (2001). *Assessing kindergarten children: A compendium of assessment instruments.* Tallahassee, FL: SERVE. Retrieved from: www.serve.org/publications/rdakcc.pdf.

Nir, A. E., & Bogler, R. (2008). The antecedents of teacher satisfaction with professional development programs. *Teaching and Teacher Education, 24*(2), 377-386. doi:10.1016/j.tate.2007.03.002.

Nitko, A. J. (2001). *Educational assessment of students*. Des Moines, IA: Prentice-Hall, Inc.

No Child Left Behind Act. (2001). *Qualifications for teachers and professionals*, 20 U.S.C. § 6319 (2008).

Noell, G. H., & Witt, J. C. (1999). When does consultation lead to intervention implementation? Critical issues for research and practice. *The Journal of Special Education, 33*(1), 29-35. doi: 10.1177/002246699903300103.

North Carolina Ready for School Goal Team. (2000). *School readiness in North Carolina: Strategies for defining, measuring, and promoting success for all children*. Greensboro, NC: SERVE.

Northwest Regional Educational Lab (NREL). (1994). Evaluating the quality and equity of alternative assessments. In Northwest Regional Educational Lab (Ed.) *Improving science and mathematics education: A toolkit for professional developers: Alternative assessment.* Portland, Oregon: Northwest Regional Educational Laboratory. Retrieved from: http://files.eric.ed.gov/fulltext/ED381360.pdf.

Olsen, L. (1997). *Made in America: Immigrant students in our public schools*. New York, NY: The New Press.

Ostrosky, M., Jung, Hemmeter, & Thomas, (2007). *What works brief training kit #3: Helping children understand routines and classroom schedules*. Center on the social and emotional foundations for early learning. Retrieved from http://csefel.vanderbilt.edu/kits/wwbtk3.pdf.

Pagani, L. S., Fitzpatrick, C., Archambault, I., & Janosz, M. (2010). School readiness and later achievement: a French Canadian replication and extension. *Developmental psychology, 46(*5), 984. doi: 10.1037/a0018881.

Pajares, M. F. (1992). Teachers' beliefs and education research: Cleaning up a messy construct. *Review of Education Research, 62*, 307-332.

Palmer, D.K., & Lynch, A.W. (2008). A bilingual education for a monolingual test? The pressure to prepare for TAKS and its influence on choices for language of instruction in Texas elementary bilingual classrooms. *Language Policy, 7,* 217-235. doi: 10.1007/s10993-008-9100-0.

Parker, F. W. (1902). An account of the work of the Cook county and Chicago normal school from 1883-1899. *The Elementary School Teacher and Course of Study, 2*(10), 752-780.

Payton, J., Weissberg, R. P., Durlak, J. A., Dymnicki, A. B., Taylor, R. D., Schellinger, K. B., & Pachan, M. (2008*). The positive impact of social and emotional learning for kindergarten to eighth-grade students: Findings from three scientific reviews*. Chicago, IL: Collaborative for Academic, Social, and Emotional Learning.

Pellegrino, J. W. (2004). *The evolution of educational assessment: Considering the past and imagining the future*. Educational testing service, policy evaluation and research center, policy information center. Retrieved from https://www.ets.org/Media/Research/pdf/PICANG6.pdf.

Perrone, V. (1991). *Expanding student assessment*. Alexandria, VA: Association for Supervision and Curriculum Development. Retrieved from: http://files.eric.ed.gov/fulltext/ED337489.pdf.

Pett, M. A. (2015). *Nonparametric statistics for health care research: Statistics for small samples and unusual distributions*. Thousand Oaks, CA: Sage Publications.

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of school psychology, 48*(1), 85-112. doi: 10.1016/j.jsp.2009.09.002.

Piaget J. (1926). *The language and thought of the child.* New York: Harcourt Brace.

Pianta, R.C. (2006). Teacher-child relationships and early literacy. In D. Dickinson & S. Newman (Eds.) *Handbook of early literacy research* (Vol. II, pp. 149-162). New York: The Guildford Press.

Pianta, R.C., La Paro, K.M., & Hamre, B.K. (2004). *Classroom assessment scoring system (CLASS).* Charlottesville, VA: Teachstone.

Pianta, R.C., Rimm-Kaufman, S., & Cox, M. J. (1999). Introduction: An ecological approach to kindergarten transition. In R.C. Pianta & M.J. Cox (Eds.) *The transition to kindergarten,* (pp. 3-10). Baltimore, Maryland: Paul H. Brookes Publishing Co.

Pierce, R., & Chick, H. (2011). Reacting to quantitative data: Teachers' perceptions of student achievement reports. In J. Clark, B. Kissane, J. Mousley, T. Spencer, & S. Thornton. (Eds.), *Mathematics: traditions and [new] practices: Proceedings of the 34th Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 631–639). Adelaide, SA: AAMT.

Pierce, R., Chick, H., Watson, J., Les, M., & Dalton, M. (2014). A statistical literacy hierarch for interpreting educational system data. *Australian Journal of Education, 58*(2), 195-217. 10.1177/0004944114530067.

Powell, D., Dunlap, G., & Fox, L. (2006). Prevention and intervention for the challenging behaviors of toddlers and preschoolers. *Infants & Young Children, 19*(1), 25-35.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*(3), 879-891. doi: 10.3758/BRM.40.3.879.

Preacher, K. J., & Selig, J. P. (2012). Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures*, *6*, 77-98. doi: 10.1080/19312458.2012.679848.

Pugach, M. C., & Johnson, L. J. (1988). Rethinking the Relationship between Consultation and Collaborative Problem-Solving. *Focus on Exceptional Children, 21*(4), 1-8.

Pyle, A., & DeLuca, C. (2013). Assessment in the kindergarten classroom: An empirical study of teachers' assessment approaches. *Early Childhood Education Journal, 41*(5), 373-380. doi: 10.1007%2Fs10643-012-0573-2.

Qualtrics. (2016). Version 12,018, Qualtrics. Provo, UT.

Quirk, M., Grimm, R., Furlong, M., Nylund-Gibson, K. & Swami, S. (2015). The association of Latin children's kindergarten school readiness profiles with grade 2-5 literacy achievement trajectories. *Journal of Educational Psychology.* doi: 10.1037/edu0000087.

Rafolth, M.A., Buchenauer, E.L., Crissman, K.K., & Halko, J.L. (2004). School readiness- preparing children for kindergarten and beyond: Information for parents (pp. 1-3). Bethesda, Maryland: National Association of School Psychologists. Retrieved from: http://www.nasponline.org/resources/handouts/schoolreadiness.pdf.

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science 28*, 4-13.

Ramey, C. T., & Ramey, S. L. (2004). Early learning and school readiness: Can early intervention make a difference? *Merrill-Palmer Quarterly, 50*(4), 471-491. doi: 10.1353/mpq.2004.0034.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*, 2nd ed. Newbury Park, CA: Sage.

Raver, C. C., & Zigler, E. F. (1997). Social competence: An untapped dimension in evaluating Head Start's success. *Early Childhood Research Quarterly, 12*, 363–385. doi: 10.1016/S0885-2006(97)90017-X.

Raykov, T. (2000). On the large-sample bias, variance, and mean squared error of the conventional noncentrality parameter estimator of covariance structure models. *Structural Equation Modeling, 7*, 431-441. doi: 10.1207/S15328007SEM0703_4.

Richards, J., Gallo, P., & Renandya, W. (2001). Exploring teachers' beliefs and the processes of change. *PAC Journal, 1*(1), 41-58.

Richardson, V. (1990). Significant and worthwhile change in teaching practice. *Educational researcher, 19*(7), 10-18. doi: 10.3102/0013189X019007010.

Rimm-Kaufman, S. E., Pianta, R. C., & Cox, M. J. (2000). Teachers' judgments of problems in the transition to kindergarten. *Early Childhood Research Quarterly, 15*(2), 147-166. doi:10.1016/S0885-2006(00)00049-1.

Rock, D.A. & Stenner, A. J. (2005). Assessment issues in the testing of children at school entry. *The Future of Children, 15*(1), 15-34. Retrieved from: http://files.eric.ed.gov/fulltext/EJ795841.pdf.

Rog, L.J. (2011). *Read, write, play, learning: Literacy instruction in today's kindergarten*. Newark, DE: International Reading Association.

Romme, A. G. L., & van Seggelen-Damen, I. C. (2015). Taking nothing for granted in management education: A systemic perspective on the role of reflective questioning. *Organization Management Journal, 12*(2), 76-86. doi: 10.1080/15416518.2014.1001056.

Rosenfield, S. (2004). Consultation as dialogue: The right words at the right time. In N. M. Lambert, I. Hylander, & J. H. Sandoval (Eds.), *Consultee-centered consultation: Improving the quality of professional services in schools and community organizations* (pp. 337–347). New York: Erlbaum.

Rosenfield, S. (2008). Best practice in instructional consultation and instructional consultation teams. In A. Thomas & J. Grimes (Eds.) *Best Practices in School Psychology V* (pp. 1645-1660). Bethesda, MD: National Association of School Psychologists.

Rosenfield, S. (2013). Consultation in schools - Are we there yet? *Consulting Psychology Journal: Practice and Research, 65*(4), 303-308. doi: 10.1037/a0035652.

Rosenfield, S.A., Gravois, T.A. & Silva, A.E. (2014). Bringing instructional consultation to scale. In W.P. Erchul & S.M. Sheridan (Eds.) *Handbook of research in school consultation*. New York, NY: Taylor & Francis.

Ross, J. A. (1992). Teacher efficacy and the effects of coaching on student achievement. *Canadian Journal of Education/Revue Canadienne de l'Education,* 51-65.

Roth McDuffie, A. & Morrison, J. (2008). Learning about data display: Connecting mathematics and science inquiry. *Teaching Children Mathematics, 14*, 375-382.

Rowe, K. J., & Sykes, J. (1989). The impact of professional development on teachers' self-perceptions. *Teaching and Teacher Education, 5*(2), 129-141.

Sabol, T. J., & Pianta, R. C. (2012). Patterns of school readiness forecast achievement and socioemotional development at the end of elementary school. *Child Development, 83*(1), 282-299. doi: 10.1111/j.1467-8624.2011.01678.x.

Saluja, G., C. Scott-Little, & R.M. Clifford. (2000). Readiness for school: A survey of state policies and definitions. *Early Childhood Research and Practice 2* (2). Retrieved from: http://ecrp.uiuc.edu/v2n2/saluja.html.

Samuels, C.A. (2014). States far from uniform in commitment to kindergarten. *Education Week, 34*(4) 1, 14-15. Retrieved from: http://www.edweek.org/ew/articles/2014/09/17/04kindergarten.h34.html.

Sandoval, J. H. (2013). *An introduction to consultee-centered consultation in the schools: A step by-step guide to the process and skills*. London: Routledge.

Sarama, J. & Clements, D.H. (2006). Mathematics in kindergarten. In D.F. Gullo (Ed.) *K Today: Teaching and learning in the kindergarten year* (pp. 85-94). Washington, D.C.: National Association for the Education of Young Children.

Sarama, J. (2002). Listening to teachers: Planning for professional development. *Teaching Children Mathematics, 9*(1), 36.

Schmoker, M. (2003). First things first: Demystifying data analysis. *Educational Leadership, 60* (5), 22-25.

Schmoker, M., & Marzano, R.J. (1999). Realizing the promise of standards-based education. *Using Standards and Assessment, 56*(6), 17-21. Retrieved from: http://www.ascd.org/publications/educational_leadership/mar99/vol56/num06/Realizing_the_Promise_of_Standards-Based_Education.aspx.

Schneider, J., & Hutt, E. (2014). Making the grade: a history of the A–F marking scheme. *Journal of Curriculum Studies, 46*(2), 201-224. doi: 10.1080/00220272.2013.790480.

Schuele, C. M., Justice, L. M., Cabell, S. Q., Knighton, K., Kingery, B., & Lee, M. W. (2008). Field-based evaluation of two-tiered instruction for enhancing kindergarten phonological awareness. *Early Education and Development, 19*(5), 726-752 doi: 10.1080/10409280802375299.

Schuh, K. L. (2004). Learner-centered principles in teacher-centered practices? *Teaching and Teacher Education, 20*(8), 833-846 doi: 10.1016/j.tate.2004.09.008.

Scott-Little, C. & Niemeyer, J. (2001). *Assessing kindergarten children: What school systems need to know.* Tallahassee, FL: SERVE. Retrieved from: http://earlysuccess.org/sites/default/files/website_files/files/2012-10-Inter-State-Dicsussion-Time-Doc-2.M.Boyajian.pdf.

Seashore Louis, K., Dretzke, B., & Wahlstrom, K. (2010). How does leadership affect student achievement? Results from a national US survey. *School effectiveness and school improvement, 21*(3), 315-336. doi: 10.1080/09243453.2010.486586.

Seifer, R., Gouley, K., Miller, A. L., & Zakriski, A. (2004). Implementation of the PATHS curriculum in an urban elementary school. *Early Education and Development, 15*(4), 471-485. doi: 10.1207/s15566935eed1504_6.

Shepard, L., Kagan, S.L., & Wurtz, E. (1998). *Principles and recommendations for early childhood assessments*. Washington, DC: National Education Goals Panel. Retrieved from: www.negp.gov/reports/prinrec.pdf.

Shoval, E., Erlich, I., & Fejgin, N. (2010). Mapping and interpreting novice physical education teachers' self-perceptions of strengths and difficulties. *Physical Education and Sport Pedagogy, 15*(1), 85-101. doi: 10.1080/17408980902731350.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*, 422-445.

Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., … & Chen, M. (2012). Early predictors of high school mathematics achievement. *Psychological science, 23*(7), 691-697. doi: 10.1177/0956797612440101.

Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical therapy, 85*(3), 257-268. doi: 10.1093/ptj/85.3.257.

Slavin, R.E., Cheung, A., Holmes, G., Madden, N.A., & Chamberlain, A. (2011) Effects of a data-driven reform model. Baltimore, MD: The Center for Data-Driven Reform in Education. Retrieved from: http://www.cddre.org/_images/Effects%20of%20a%20Data%20Driven%20District%20Reform%20Model%20January%202011.pdf.

Smallwood, M. L. (1935). A historical study of examinations and grading systems in early American universities. *Harvard studies in evaluation*. Cambridge, MA: Harvard University Press.

Snow, C., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Snow, K. (2011). *Developing kindergarten readiness and other large-scale assessment systems: Necessary considerations in the assessment of young children*. Washington, DC: National Association for the Education of Young Children.

Spear-Swerling, L., & Zibulsky, J. (2014). Making time for literacy: teacher knowledge and time allocation in instructional planning. *Reading and Writing, 27*(8), 1353-1378. doi: 10.1007/s11145-013-9491-y.

Spear-Swerling, L., Brucker, P. O., & Alfano, M. P. (2005). Teachers' literacy-related knowledge and self-perceptions in relation to preparation and experience. *Annals of Dyslexia, 55*(2), 266-296. doi: 10.1007/s11881-005-0014-7.

Stanovich, K. E., Cunningham, A. E., & Cramer, B. B. (1984). Assessing phonological awareness in kindergarten children: Issues of task comparability. *Journal of Experimental Child Psychology, 38(*2), 175-190.

State Education Resource Center [SERC]. (2012). *Elementary assessments: Universal screening, diagnostic, & progress monitoring*. Retrieved from: http://www.sde.ct.gov/sde/lib/sde/pdf/curriculum/cali/elementary_assessments_4-9-12.pdf.

Stecher, B.M., & Barron, S. (2001). Unintended consequences of test-based accountability when testing in "milepost" grades. *Educational Assessment, 7*(4), 259-281. doi: 10.1207/S15326977EA0704_02.

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42*, 795-819. doi: 10.1002/pits.20113.

Stecker, P. M., Lembke, E. S., & Foegen, A. (2008). Using progress-monitoring data to improve instructional decision making. *Preventing School Failure: Alternative Education for Children and Youth, 52*(2), 48-58 doi: 10.3200/PSFL.52.2.48-58.

Sterdon, J.M. & Berger, A. (2010). *NCSL technical report: State approaches to school readiness assessment.* Washington, DC: National Conference of State Legislature. Retrieved from: http://www.ncsl.org/documents/Educ/KindergartenAssessment.pdf.

Stevahn, L., Johnson, D. W., Johnson, R. T., Oberle, K., & Wahl, L. (2000). Effects of conflict resolution training integrated into a kindergarten curriculum. *Child Development, 71,* 772–784 doi: 10.1111/1467-8624.00184.

Stevenson, H.W., & Newman, R.S. (1986). Long-term prediction of achievement and attitudes in mathematics and reading. *Child Development, 57*, 646-659.

Stipek, D. (2004). Teaching practices in kindergarten and first grade: Different strokes for different folks. *Early Childhood Research Quarterly, 19*(4), 548-568. doi: 10.1016/j.ecresq.2004.10.010.

Stipek, D., & Byler, P. (2001). Academic achievement and social behaviors associated with age of entry into kindergarten. *Journal of Applied Developmental Psychology, 22*(2), 175-189. doi: 10.1016/S0193-3973(01)00075-2.

Strand, P.S. & Cerna, S. (2010). The effects of a data dissemination strategy on the letter naming and object counting skills of preschoolers attending Head Start. *Journal of Behavioral Education, 19*, 298-305. doi: 10.1007/s10864-010-9116-3.

Strickland, D.S. (2006). Language and literacy in kindergarten. In D.F. Gullo (Ed.) *K Today: Teaching and learning in the kindergarten year* (pp. 73-84). Washington, D.C.: National Association for the Education of Young Children.

Suen, H. K., & Ary, D. (2014). *Analyzing quantitative behavioral observation data.* New York, NY: Psychology Press.

Supovitz, J. A. (2009). Can high stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform. *Journal of Educational Change, 10*(2), 211-227. doi: 10.1007/s10833-009-9105-2.

Sutherland, K. S., Conroy, M. A., Vo, A., & Ladwig, C. (2015). Implementation integrity of practice-based coaching: Preliminary results from the BEST in CLASS efficacy trial. *School Mental Health, 7(*1), 21-33. doi: 10.1007/s12310-014-9134-8.

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research, 83*(3), 357-385. doi: 10.3102/0034654313483907.

Thomas, W.P., & Collier, V. (1997). *School effectiveness for language minority students.* Washington, DC: National Clearinghouse for English Language Acquisition. Retrieved from: http://www.ncbe.gwu.edu/ncbepubs/resource/effectiveness/index.htm.

Timperley, H. S., & Phillips, G. (2003). Changing and sustaining teachers' expectations through professional development in literacy. *Teaching and Teacher Education, 19*(6), 627-641. doi:10.1016/S0742-051X(03)00058-1.

Tschannen-Moran, M. & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive concept. *Teaching and Teacher Education, 17*, 783-805. doi: 10.1016/S0742-051X(01)00036-1.

Tschannen-Moran, M., & McMaster, P. (2009). Sources of self-efficacy: Four professional development formats and their relationship to self-efficacy and implementation of a new teaching strategy. *The Elementary School Journal, 110*(2), 228-245. doi: 10.1086/605771.

Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*, 202-248.

U.S. Department of Education & U.S. Department of Health and Human Services [U.S. DOE & U.S. DHHS]. (2014). The race to the top – Early learning challenge year two progress report. Washington, DC: Author.

U.S. Department of Education, National Center for Education Statistics. (2009). *Early childhood longitudinal study, kindergarten class of 1998–99 (ECLS-K) eighth grade restricted-use data file and electronic codebook* (DVD). (NCES 2009-006). Washington, DC: Author.

U.S. Department of Education. (2013). *Strategic plan for fiscal years 2014-2018.* Washington, DC: U.S. Department of Education. Retrieved from: https://www2.ed.gov/about/reports/strat/plan2014-18/strategic-plan.pdf.

U.S. Department of Health & Human Services. (2002). *A tribal guide to the good start, grow smart early learning initiative.* Washington, DC: Author. Retrieved from: http://files.eric.ed.gov/fulltext/ED493427.pdf.

U.S. House of Representatives and Senate. (2007). *Bill H.R.1429, the improving Head Start for school readiness act*. P.L. 110-34.

Verhoeven, L., van Leeuwe, J., Irausquin, R., & Segers, E. (2016). The unique role of lexical accessibility in predicting kindergarten emergent literacy. *Reading and Writing, 29*(4), 591-608. doi: 10.1007/s11145-015-9614-8.

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*(1), 80-91. doi: 10.1016/ j.tate.2007.01.004.

Vonk, J.H.C. (1991). Becoming a teacher, brace yourself. In H.W. Kam & R.Y.L. Wong (Eds.), *Improving the quality of the teaching profession* (pp. 63-81). Singapore: The Institute of Education.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, Mass.: Harvard University Press.

Wahlstrom, K. L., & Seashore Louis, K. (2008). How teachers experience principal leadership: The roles of professional community, trust, efficacy, and shared responsibility. *Educational Administration Quarterly, 44*(4), 458-495. doi: 10.1177/0013161x08321502.

Walston, J., Rathbun, A., and Germino Hausken, E. (2008). Eighth grade: First findings from the final round of the early childhood longitudinal study, kindergarten class of 1998-99. (NCES 2008-088). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Wanless, S. B., McClelland, M. M., Tominey, S. L., & Acock, A. C. (2011). The influence of demographic risk factors on children's behavioral regulation in prekindergarten and kindergarten. *Early Education and Development, 22*, 461-488. doi: 10.1080/10409289.2011.536132.

Warren, E., & Nisbet, S. (1999). The relationship between the purported use of assessment techniques and beliefs about the uses of assessment. In J.M. Truran, & K.M. Truran (Eds.), *22nd Annual conference of Mathematics Education and Research Group of Australasia* (Vol. 22, pp.515-521). Adelaide, SA: MERGA.

Watkins, D. (2001) Correlates of approaches to learning: a cross-cultural meta-analysis. In R. J. Sternberg & L. F. Zhang (Eds.), *Perspective on thinking, learning and cognitive styles*. Mahwah, NJ: Lawrence Erlbaum).

Wayman, J. C., Jimerson, J. B., & Cho, V. (2012). Organizational considerations in establishing the data-informed district. *School Effectiveness and School Improvement 23*(2), 159-178. doi: 10.1080/09243453.2011.652124.

Wayman, J.C., & Jimerson, J.B. (2013). Teacher needs for data-related professional learning. *Studies in Educational Evaluation, 42*, 25-24. doi:10.1016/j.stueduc.2013.11.001

Wayne, A. J., Yoon, K. S., Zhu, P., Cronen, S., & Garet, M. S. (2008). Experimenting with teacher professional development: Motives and methods. *Educational researcher, 37*(8), 469-479.

Webster-Stratton, C., & Reid, M. J. (2004). Strengthening social and emotional competence in young children- the foundation for early school readiness and success: Incredible years classroom social skills and problem-solving curriculum. *Infants & Young Children, 17*(2), 96-113. doi: 10.3102/0013189X08327154.

Wesson, C.L., King, R.P., & Deno, S.L. (1984). Direct and frequent measurement: If it's so good for us, why don't we use it*? Learning Disability Quarterly, 7*, 45-48.

Whiting, B. B., Edwards, C. P., & Edwards, C. P. (1992). *Children of different worlds: The formation of social behavior.* Cambridge, MA: Harvard University Press

Wild, C. J., Pfannkuch, M., Regan, M., & Horton, N. J. (2011). Towards more accessible conceptions of statistical inference. *Journal of the Royal Statistical Society, 174*(2), 247-295. doi: 10.1111/j.1467-985X.2010.00678.x.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37*(1), 3-14. doi: 10.1016/j.stueduc.2011.03.001.

Williams, K. T. (1997). Expressive vocabulary test second edition (EVT™ 2). Journal of the *American Academy of Child & Adolescent Psychiatry, 42*, 864-872.

Williford, A.P., Downer, J.T., Hamre, B.K., & Pianta, R.C. (2014). *The Virginia kindergarten readiness project: Executive summary & legislative report*. Charlottesville, VA: Elevate Early Education.

Witt, J. C., Noell, G. H., LaFleur, L. H., & Mortenson, B. P. (1997). Teacher use of interventions in general education settings: Measurement and analysis of the independent variable. *Journal of Applied Behavior Analysis, 30*(4), 693-696. doi: 10.1901/jaba.1997.30-693.

Wohlstetter, P., Datnow, A., & Park, V. (2008). Creating a system for data-driven decision-making: Applying the principal-agent framework. *School Effectiveness and School Improvement, 19*(3), 239-259. doi: 10.1080/09243450802246376.

Woolfolk Hoy, A., & Spero, R. B. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education, 21*(4), 343-356. doi: 10.1016/j.tate.2005.01.007.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics, 14*(4),1261-1295.

Xue, Y., & Meisels, S. J. (2004). Early Literacy instruction and learning in kindergarten: Evidence from the Early Childhood Longitudinal Study: Kindergarten Class of 1998-1999. *American Educational Research Journal*, *41*(1), 191-229. doi: 10.3102/00028312041001191.

Yeh, S.S. (2005). Limiting the unintended consequences of high-stakes testing. *Education Policy Analysis Archives, 13*(43).

Ysseldyke, J., & Christenson, S. (1993). *TIES-II: The instructional environment system-II. A system to identify a student's instructional needs*. Longmont, CO: Sopris West, Inc.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist, 25*(1), 3-17.

Zins, J. E., & Elias, M. J. (2006). Social and emotional learning. In G. G. Bear & K. M. Minke (Eds.), *Children's needs III: Development, prevention, and intervention* (pp. 1–13). Bethesda, MD: National Association of School Psychologists.

Zinsser, K.M., Weissberg, R.P., Dusenbury, L. (2013). *Aligning preschool through high school social and emotional learning standards: A critical (and doable) next step*. Chicago, IL: Collaborative for Academic, Social and Emotional Learning. Retrieved from: https://static1.squarespace.com/static/513f79f9e4b05ce7b70e9673/t/52ac836ce4b0532c165d4f2e/1387037548200/Zinsser+et+al++brief+on+state+standards--20131214.pdf.

Table 1

Crosswalk of Best Practices in Instructional and Data-Driven Consultation Models

	Instructional Consultation					
Data-Driven Consultation	(1) Entry and Contracting	(2) Problem Identification and Analysis	(3) Intervention Design and Planning	(4a) Intervention Implementation	(4b) Intervention Evaluation	(5) Closure
(1) Data Acquisition		X				
(2) Data Reflection		X				
(3) Program Design			X			
(4) Program Alignment			X			
(5) Formative Feedback					X	

Note. Instructional consultation principles were adapted from Rosenfield (2013) and Rosenfield, Gravois, and Silva, (2014). Best practices of data-driven consultation were surmised from Dimmitt, Carey, and Hatch (2007), Halverson and colleagues (2006) and Slavin and colleagues (2011)

Table 2

Completion and Demographic Data for the Selected Division, Current Study, Treatment Group, and BAU Group

	Frequency (Percent)			
Scale	Division (N = 116)[a]	Current Study (N = 72)	Treatment (*n* = 38)	BAU (*n* = 34)
Survey Completion	--	--	--	--
Demographic	*n* = 96 (80%)	*n* = 60 (83%)	*n* = 34 (89%)	*n* = 26 (76%)
Pretest	--	*n* = 67 (94%)	*n* = 38 (100%)	*n* = 29 (85%)
Posttest	--	*n* = 67 (94%)	*n* = 37 (97%)	*n* = 30 (88%)
Female	*n* = 93 (97%)	*n* = 70 (97%)	*n* = 37 (97%)	*n* = 33 (97%)
Race		--	--	--
White	*n* = 76 (80%)	*n* = 62 (86%)	*n* = 32 (84%)	*n* = 30 (88%)
Black	*n* = 15 (19 %)	*n* = 10 (14%)	*n* = 6 (15%)	*n* = 4 (12%)
Identified as Hispanic	*n* = 3 (3%)	*n* = 2 (3%)	*n* = 1 (3%)	*n* = 1 (3%)
Education	--	--	--	--
Bachelor's	*n* = 32 (33%)	*n* = 21 (29%)	*n* = 13 (34%)	*n* = 8 (24%)
Master's or Higher	*n* = 64 (67%)	*n* = 48 (71%)	*n* = 22 (66%)	*n* = 26 (76%)
Worked in Current School Before	*n* = 83 (89%)	*n* = 64 (88%)	*n* = 33 (86%)	*n* = 31 (91%)
Number of Schools Represented	N = 25	N = 21	*n* = 18 (86%)	= 18 (86%)

Table 2 (continued)

	Mean (*SD*)			
Scale	Division (N = 116)	Current Study (N = 72)	Intervention (*n* = 38)	BAU (*n* = 34)
Avg # Teachers/School	5.16 (1.79)	3.54 (1.30)	2.11 (1.07)	1.89 (0.65)
Birth Year	1973 (12.73)	1974 (12.44)	1974.0 (11.86)	1976.0 (12.37)
Experience (in years)	--	--	--	--
Kindergarten	7.00 (7.25)	6.73 (7.36)	6.78 (7.86)	7.21 (6.58)
Total	13.92 (10.25)	13.31 (10.71)	13.71 (10.71)	11.92 (9.75)

[a] Because demographic data were not complete for the selected division, percentages were calculated as the percent of responders (*n* = 96) with the exception of the Survey Completion- Demographic percent.

Table 3

Procedural Timeline of Intervention

Timeframe (2016)	Activity	Participants: All VKRP Teachers	Participants: Intervention Group	Participants: BAU Group
August	VKRP Training	X	X	X
	Demographic Survey	X	X	X
	Consent Procedures		X	X
September	Pretest Survey		X	X
October	VKRP Administration	X	X	X
	Notified of intervention condition		X	X
	Consultation session scheduled		X	
November	Consultation session held		X	
December	Posttest Survey		X	X

Table 4

Original Items the Learning from Leadership Project Survey and Alterations Made for Current Study

Teacher Survey Item	Alterations for Current Study [a]
[4-1] How often in this school year has your school administrator buffered teachers from distractions to their instruction	[Deleted]
[4-10] My school administrator clearly defines standards for instructional practices.	My school administrator clearly defines standards *for using assessment to inform* instructional practices.
[4-13] How often in this school year has your school administrator discussed instructional issues with you?	How often in this school year has your school administrator discussed issues *around data use* with you?
[4-16] How often in this school year has your school administrator observed your classroom instruction?	How often in this school year has your school administrator observed *your assessment practices in the classroom*?
[4-18] How often in this school year has your school administrator attended teacher planning meetings?	How often in this school year has your school administrator attended teacher planning meetings *related to assessment or data-use*?
[4-19] How often in this school year has your school administrator made suggestions to improve classroom behavior or classroom management?	How often in this school year has your school administrator made suggestions to improve the *use of assessment in the classroom*?
[4-21] How often in this school year has your school administrator given you specific ideas for how to improve your instruction?	How often in this school year has your school administrator given you specific ideas for how to improve your instruction *related to data*?

Note: Survey was adapted with permission from Dr. Karen Seashore-Louis. The full survey can be seen in *Leithwood, K., Seashore Louis, K., Anderson, S., & Wahlstrom, K. (2004). Review of research: How leadership influences student learning. Executive summary prepared for The Wallace Foundation.*
[a]Alterations are denoted in the table by italics.

Table 5

Factor Loadings of Teacher Reported Self-Efficacy and Practices

Survey Item	Factor 1: Data-Use Efficacy	Factor 2: Data-Use Practices
I feel confident in interpreting individual-level assessment data (e.g., a student's answers to a math test) to inform my instruction.	1.016	
I feel confident in interpreting classroom-level assessment data (e.g., scores across students' math tests) to inform my instruction.	.733	
I alter my teaching practice as a result of information obtained from assessments throughout the school year.		.878
I use assessments to inform classroom practice.		.710
I individualize instruction based on results of assessments.		.708
I use assessments to inform grouping children according to needs.	.216	.688
I alter my teaching practice as a result of information obtained from entry assessments.		.630

Table 6

Pooled Unstandardized Betas and Standard Deviations of Linear Regression Models

Predictor	*B* (*SD*)						
	Data Int. Skills-Classroom	Data Int. Skills-Individual	Improvement Belief	School Acct. Belief	Irrelevance Belief	Data-Use Efficacy	Data-Use Practices
Constant	15.94 (12.92)	6.27 (9.86)	138.70 (170.18)	20.33 (62.20)	198.20 (221.40)	24.48 (42.60)	18.65 (80.40)
Teacher Demographics							
Age	.01 (.01)	-.01 (.01)	.06 (.09)	.01 (.03)	.09 (.11)	.01 (.02)	.01 (.04)
Race (White)	-.04 (.12)	-.04 (.09)	4.22 (1.88)*	.99 (.63)	.16 (1.62)	-.15 (.38)	-.18 (.77)
Education (MA/MS)	.15 (.10)	-.06 (.08)	1.38 (1.45)	.22 (.57)	3.75 (2.17)$^{\tau}$	.61 (.31)*	.69 (.63)
K Experience (years)	-.01 (.01)	-.00 (.01)	.12 (1.22)	.09 (.05)	.07 (.17)	-.01 (.03)	-.03 (.05)
Total Experience (years)	.02 (.01)	-.00 (.01)	.02 (.12)	.06 (.05)$^{\tau}$	.04 (.17)	.02 (.03)	.02 (.06)
Taught in current school before	.04 (.15)	-.08 (.12)	.25 (2.12)	.04 (.80)	.11 (2.88)	-.80 (.50)	-1.61 (.99)

$^{\tau}$ $p < .10$, * $p < .05$, ** $p < .01$

Table 6 continued

Predictor	*B* (*SD*)						
	Data Int. Skills-Classroom	Data Int. Skills-Classroom	Improvement Belief	School Acct. Belief	Irrelevance Belief	Data-Use Efficacy	Data-Use Practices
Perception of Climate	.01 (.01)$^{\tau}$	.01 (.01)	-.28 (.11)*	.02 (.04)	.17 (.12)	-.05 (.02)*	-.04 (.05)
Pretest Response	.33 (.15)*	.05 (.10)	.58 (.10)**	.67 (.11)**	.67 (.13)**	.30 (.15)$^{\tau}$	.70 (.14)**
Tre. Condition (Intervention)	.37 (.08)**	.09 (.07)	2.06 (1.28)	.49 (.45)	1.05 (1.40)	.54 (.29)*	.25 (.55)

$^{\tau}$ $p < .10$, * $p < .05$, ** $p < .01$

Table 7

Group Means, t Values, and Local Effect Sizes for Outcome Variables

Construct	$\bar{x}_{TRE}$ (*SD*)	$\bar{x}_{BAU}$ (*SD*)	*t* value	*p* value	Cohen's *d*	Power
Data Interpretation Skills- Classroom	0.78 (.30)	0.39 (0.30)	4.43**	0.000	1.06	1.00
Data Interpretation Skills- Student	0.64 (.25)	0.55 (0.21)	1.42	0.156	0.34	0.41
Improvement Belief	15.13 (6.55)	13.04 (7.67)	1.59	0.112	0.38	0.24
School Accountability Belief	4.17 (1.97)	4.15 (2.19)	1.08	0.279	0.26	0.03
Irrelevance Belief	12.77 (6.28)	13.24 (6.67)	0.75	0.454	0.18	0.07
Data-Use Efficacy	3.13 (1.20)	2.44 (1.03)	2.04*	0.041	0.49	0.69
Data-Use Practices	5.76 (2.01)	5.23 (2.59)	0.51	0.609	0.24	0.31

Table 8

Indirect Effects of the Intervention on Data-Use Practices Through Changes in Data Interpretation Skills at the Classroom-Level, Improvement Belief, and Data-Use Efficacy

Construct	a path	b path	c path	*ab*	P_m
Data Interpretation Skills- Classroom	0.06 (0.04)*	0.04 (.03)*	0.02 (0.02)	0.01 (0.01)	--
Improvement Belief	0.01 (0.03)	0.00 (0.00)	0.02 (0.02)	0.00 (0.00)	--
Data-Use Efficacy	0.13 (0.03)*	0.20 (0.01)*	0.02 (0.02)	0.03 (0.01)*	0.12

Note: P_m represents the percent of mediation, or the proportion of the total effect for which the indirect effect accounts.

Table 9

Achieved Relative Strength (Hedge's g) Calculations for Outcome Variables

Construct	Hedges' *g*
Data Interpretation Skills- Classroom	1.41
Data Interpretation Skills- Student	0.56
Improvement Belief	0.44
School Accountability Belief	0.33
Irrelevance Belief	0.19
Data-Use Efficacy	0.82
Data-Use Practices	0.39

Table 10

Frequency of Cycles by Instructional Domain and Instructional Level

Instructional Domain	Instructional level			
	Individual	Small-group	Whole-group	Total
Math	5	13	7	25
Geometry	1	2	4	7
Numeracy	1	7	3	11
Computation	0	5	1	6
Patterning	1	0	0	1
Literacy	1	0	0	1
Social Skills	3	3	6	12
Self-regulation	12	4	7	23
Total	21	20	20	61

Table 11

Frequency of Cycles by Instructional Domain and Instructional Level for Each Consultant

	Consultant		
Domain	1	2	Total
Math	14	11	25
Literacy	1	0	1
Social Skills	12	0	12
Self-regulation	19	6	23
Level			
Individual	11	12	21
Small Group	20	0	20
Whole Group	15	5	20
Total	46	17	60

Figure 1

Theory of Change Model for the Instructional-Focused, Data-Driven Consultation Model

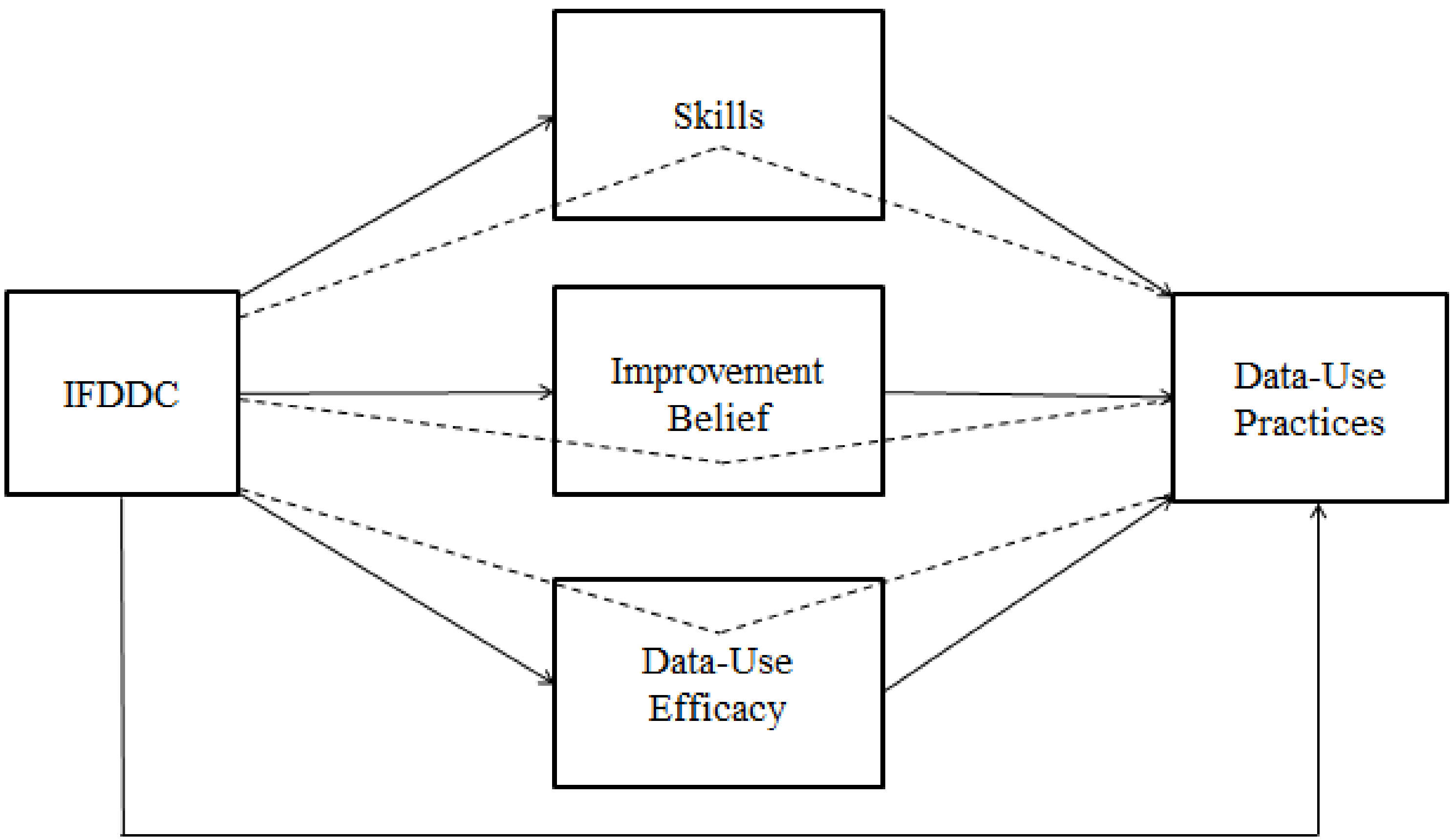
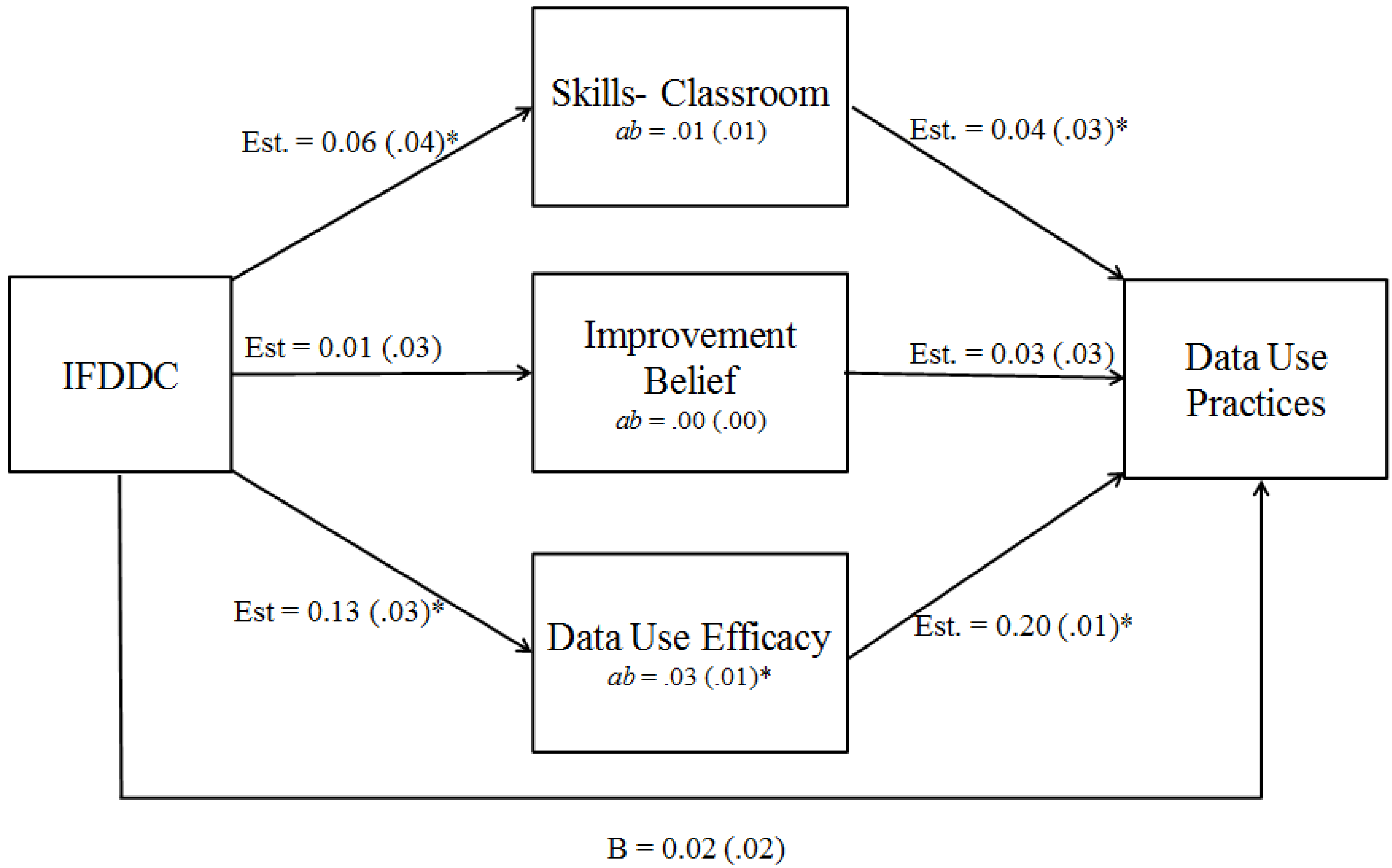


Figure 2

Direct and Indirect Effects of Full Analytical Model



** p < .00, * p < .05,

Appendix 1

Explanation of A Priori Power Analysis

An a priori power analysis was conducted using a common method for handling nested randomization in which clustering is adjusted for by multiplying the sample size of a one-level power analysis by the variance inflation factor (VIF) across levels (Campbell, Thomson, Ramsay, MacLennan, & Grimshaw, 2004; Davis, 2001; Eldrige, Ashby, & Kerry, 2006). In a two-level model, where level 1 represents classrooms and level 2 represents schools, the VIF uses the estimated intra-cluster correlation (ICC) between the two levels and the average number of observations in level 1 per level 2 group to estimate the effects of nesting (VIF = $1 + [m\text{-}1]\ \rho$, where m is the sample size per cluster and ρ is the ICC). The ICC was estimated to be 0.3 based on average of findings reported from similar consultation models (e.g., Fuchs et al., 1992; Fuchs et al., 1991) following an estimation model suggested in Higgins and Green (2011). With an average of four classrooms per school participating in VKRP, this created a VIP of 1.9. The one-level power analysis conducted with an alpha of 0.05 and beta of 0.20 by standards puts forth in Cohen (1977, 1995) revealed a necessary total sample size of at least 56, which when multiplied by the VIF resulted in a necessary sample size of at least 106. Although the pilot year had nearly 100 percent participation (Williford et al., 2014), literature detailing participation in school-based consultation usually cites attrition rates of approximately ten to 16 percent (Fuchs, Fuchs, Hamlett, & Stecker, 1991; Mashburn, Downer, Hamre, Justice, & Pianta, 2010). As such, 15 percent was added to the outcome of the power analysis, resulting in a sample size of 122.

Appendix 2

Explanation of Kindergarten Entry Assessments Implemented During the Virginia Kindergarten Readiness Program

The following literacy, mathematics, and social skills and approaches to learning KEAs were administered by all teachers to all children in their classroom as a part of VKRP.

Literacy. To assess children's incoming literacy abilities, *the Phonological Awareness and Literacy Screening-Kindergarten* (PALS-K; Invernizzi, Swank, Juel, & Meier, 2003) will be administered. The PALS-K is a state-wide, one-on-one assessment of early literacy skills such as phonological awareness, alphabet knowledge, and grapheme-phoneme correspondence (letter writing and sounds). Children's performance is scored across six subtests (rhyme awareness, beginning sound awareness, alphabet knowledge, letter-sounds, spelling, and concept of word-word list) which are summed to create a cumulative score. The subtests are moderately to highly correlated, with correlations ranging from r = .33 to r = .93 (Invernizzi et al., 2004). Evaluations of the PALS-K reveal predictive validity (Invernizzi et al., 2004; Justice, Invernizzi, & Meier, 2002), as well as sound inter-rater and test-retest reliability (Invernizzi et al., 2004).

Mathematics. To assess children's incoming mathematics abilities, *the Happy Birthday: Early Mathematics Assessment System* (Birthday Party: EMAS; Ginsburg & Pappas, 2016) will be administered. The Birthday Party: EMAS is a one-on-one direct assessment draws from literature detailing the developmental trajectory of early mathematic skills in three-, four-, and five-year-old children. This measure is comprised of 27 items that factor into four categories (RMSEA = .05, SRMR = .02, CFI/TLI = 1.0/.99; Ginsburg & Pappas, 2016; Lee, 2016): Number and Operation, Shape, Space, and Pattern. At age five, the Birthday Party: EMAS has demonstrated sound test-retest and interrater reliability, as well as concurrent and predictive validity (Ginsburg & Pappas, 2016; Lee, 2016).

Social skills and approaches to learning. To assess children's social skills and approaches to learning, the *Child Behavior Rating Scale* (CBRS; Bronson, Goodson, Layzer, & Love, 1990) will be completed for all children. The CBRS is a teacher rating scale consisting of 17 items rated on a Likert scale (1: never to 5: usually/always) about a child's behavior in the classroom. Items fall into two subscales: social skills (7 items) and approaches to learning (10 items). The social skills subscale include items such as "[child is] willing to share toys or other things with other children when playing," while the approaches to learning subscale includes items such as "completes learning tasks involving two or more tasks (e.g. cutting and pasting) in an organized way." Both scales have shown concurrent and predictive validity across diverse samples (Lim, Rodger, & Brown, 2010; Matthews et al., 2009; Wanless, McClelland, Tominey, & Acock, 2011) and have strong internal validity (social skills alpha = .94, approaches to learning alpha =.94; Williford et al., 2014).

Appendix 3

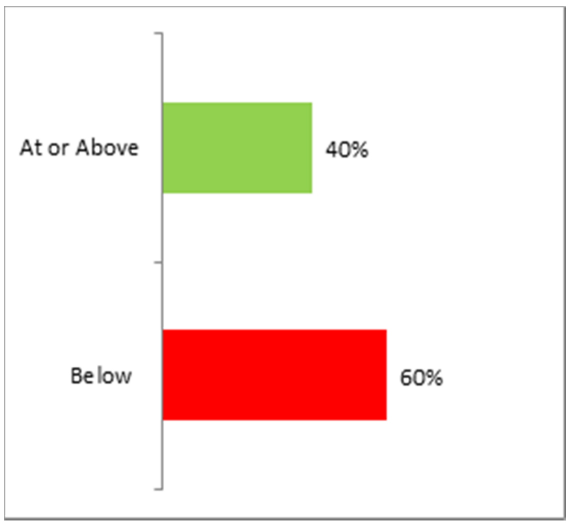
Example Virginia Kindergarten Readiness Program Report

Virginia Kindergarten Readiness Program Teacher Report

Teacher Name: Jane Doe

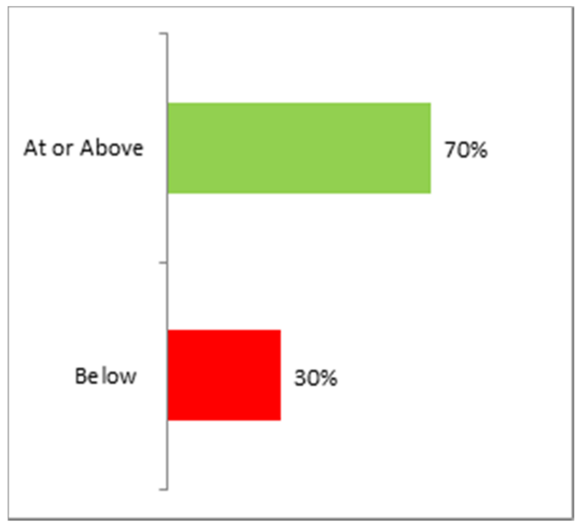
Classroom Summary

Math- Birthday Party



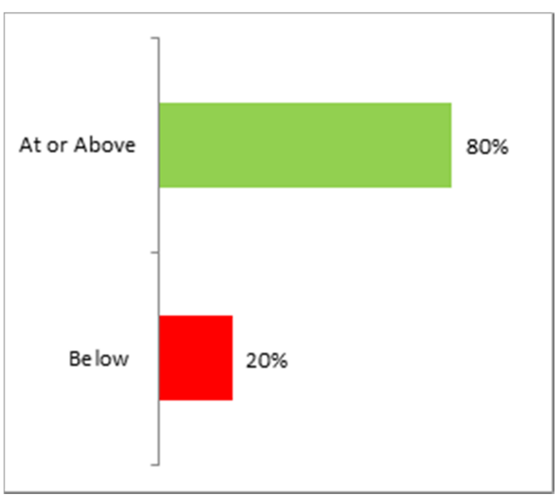
Students receiving a score of 23 or above meet the benchmark for math readiness at the beginning of kindergarten

Literacy- PALS



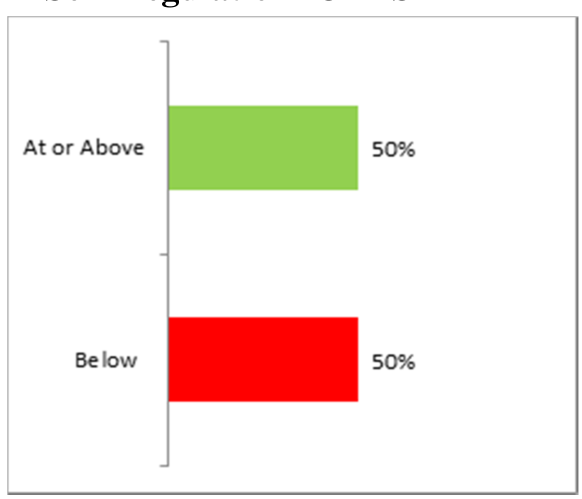
Students receiving a score of 29 or above meet the benchmark for literacy readiness at the beginning of kindergarten

Social Skills- CBRS



Students receiving a score of 3.71 or above meet the benchmark for social skills readiness at the beginning of kindergarten

Self-Regulation- CBRS



Students receiving a score of 2.90 or above meet the benchmark for self-regulation readiness at the beginning of kindergarten

Virginia Kindergarten Readiness Program
Classroom Summary Report

Teacher Name: Jane Doe

Student's Scores on each early learning domain are provided below. Within each learning domain, students whose score is at or above the benchmark/threshold are shaded in green *and* students whose score is below the benchmark/threshold are shaded red.

	Math	Literacy	Social Skills	Self-Regulation
Class Average	**22**	**31**	3.8	**2.7**
Aja	22	22	3.1	4.0
Bella	20	29	4.0	*2.3*
Emily	19	29	3.7	2.8
Eric	30	34	6.5	6.3
Hector	22	34	2.3	2.9
Kareem	21	35	5.0	6.6
Lucy	25	38	3.8	2.1
Ryan	29	30	*6.0*	2.5
Vivian	20	18	3.8	4.3
Zhang	31	23	3.9	1.5

Virginia Kindergarten Readiness Program Teacher Report

Teacher Name: Jane Doe

Classroom Summary: Math

	Total Score	Numeracy*	Patterning*	Computation*	Geometry*
Class Average	**22**	**85%**	**25%**	**70%**	**50%**
Aja	20	30%	25%	90%	15%
Bella	22	45%	60%	90%	30%
Emily	19	55%	30%	75%	75%
Eric	30	90%	90%	95%	85%
Hector	22	75%	25%	25%	75%
Kareem	21	50%	40%	25%	60%
Lucy	25	75%	80%	75%	75%
Ryan	29	90%	65%	95%	85%
Vivian	20	50%	55%	45%	30%
Zhang	31	90%	45%	95%	85%

*Reported as percent correct for each domain

Virginia Kindergarten Readiness Program
Student Math Report

School Name XXXXXXX   Teacher Name: Jane Doe   Student Name: Hector

Class Average Math Score: 77.5

Child Math Score: 50%

Math Subscale Scores (% Correct): Numeracy: 25% Patterning: 25% Computation: 50% Geometry: 50%

Item Level Report

NUMERACY Numeracy Class Average: 85 Child Numeracy Score: 25					
ITEM	**QUESTION**	**CONSTRUCT**	**CORRECT ON FIRST TRY**	**# OF CONTINGENCIES NEEDED TO ARRIVE AT RIGHT ANSWER**	**STRATEGY**
N&O 1	Highest Number Counted Correctly	Counting	24 (3 Errors)	N/A	N/A
N&O 2A	Which is more 2 or 3?	Comparing & Ordering Numbers	YES		XX
N&O 2B	Which is more 5 or 4?		YES		XX
N&O 2C	Which is more 1 or2?		NO		XX
N&O 2D	Which is more 9 or 10?		NO	1/3	NS
N&O 2E	Which is more 9 or 8?		NO		XX
N&O 2F	Which is more 6 or 7?		NO		XX
N&O 3A	1 to 1 Counting (6)	Enumeration	YES		XX
N&O 3A	Cardinality Principle		YES		N/A
N&O 3A	Identity Principle		NO	2/3	N/A
N&O 3B	1 to 1 Counting (8)		NO		XX
N&O 3B	Cardinality Principle		NO		N/A
N&O 3B	Identity Principle		NO	3/3	N/A
N&O 5A	Legibly writes the number 5	Numerals	NO	2/3	Pictographic
N&O 5B	Legibly writes the number 9		NO	3/3	XX
N&O 5C	Legibly writes the number 7		NO	3/3	XX
PATTERNING Patterning Class Average: 75 Child Patterning Score: 25					
ITEM	**QUESTION**	**CONSTRUCT**	**CORRECT ON FIRST TRY**	**# OF CONTINGENCIES NEEDED TO ARRIVE AT**	**STRATEGY**

				RIGHT ANSWER	
SSP 3A-1 **SSP 3A-2** **SSP 3B-1** **SSP 3B-2**	Pattern Reproduction (ABABAB) Pattern Extension (ABABAB**ABAB**) Pattern Reproduction (ABBABB) Pattern Extension (ABBABB**ABBABB**)	Patterns	NO NO NO NO		N/A N/A N/A N/A
COMPUTATION Computation Class Average: 75 Child Computation Score: 50					
ITEM	**QUESTION**	**CONSTRUCT**	**CORRECT ON FIRST TRY**	**# OF CONTINGENCIES NEEDED TO ARRIVE AT RIGHT ANSWER**	**STRATEGY**
N&O 4A **N&O 4B** **N&O 4C** **N&O 4D**	3+1=4 3-2=1 4-1=3 4+2=6	Adding & Subtracting	YES NO NO NO	 1/3 2/3 2/3	XX XX XX XX
GEOMETRY Geometry Class Average: 80 Child Geometry Score: 50					
ITEM	**QUESTION**	**CONSTRUCT**	**CORRECT ON FIRST TRY**	**# OF CONTINGENCIES NEEDED TO ARRIVE AT RIGHT ANSWER**	**STRATEGY**
SPS 1A **SPS 1B** **SPS 1C** **SPS 1D**	Identification of a Rectangle Identification of a Triangle Identification of a Square Identification of a Triangle	Recognizing Shapes: Identification	YES YES NO NO		N/A N/A N/A N/A
SPS 2A **SPS 2B** **SPS 2C**	Attribute Identification: Square Attribute Identification: Rectangle Attribute Identification: Triangle	Recognizing Shapes: Attributes	YES YES YES		N/A N/A N/A

SPS 4A **SPS 4B** **SPS 4C** **SPS 4D**	Identification of **Closest to** Identification of **Farthest From** Identification of **In Between** Identification of **Next to**	Spatial Sense	NO NO NO NO	1/2 2/2 1/2 1/2	N/A N/A N/A N/A
SPS 5A **SPS 5B** **SPS 5C** **SPS 5D**	Construction of a Vertical Column Construction of a Horizontal Row Construction of a Diagonal Row Correct Placement of One Counter	Spatial Sense	NO NO NO YES	2/3 2/3 3/3	N/A N/A N/A N/A

Virginia Kindergarten Readiness Program
Student Social-Emotional Report

School Name XXXXXXX Teacher Name: Jane Doe Student Name: Hector

SELF-REGULATION **Child1 Self-Regulation Score: 2.3 Class Average Self-Regulation Score: 2.5**	
Item	**Child's Score**
Observes rules and follows directions without requiring repeated reminders	Never
Responds to instructions and then begins an appropriate task without being reminded	Rarely
Completes learning tasks involving two or more steps (e.g. cutting and pasting) in organized way	Frequently/ Usually
Finds and organizes materials and works in an appropriate place when activities are initiated	Never
Attempts new challenging tasks	Sometimes
Concentrates when working on a task; is not easily distracted by surrounding activities	Sometimes
Completes tasks successfully	Sometimes
Takes time to do his/her best on a task	Always
Sees own errors in a task and corrects them	Sometimes
Returns to unfinished tasks after interruption	Rarely

SOCIAL SKILLS **Child1 Self-Regulation Score: 2.9 Class Average Self-Regulation Score: 3.3**	
Item	**Child's Score**
Willing to share toys or other things with other children when playing; does not fight or argue with playmates in disputes over property	Sometimes
Cooperative with playmates when participating in a group play activity; willing to give and take in the group, to listen to or help others	Rarely
Takes turns in a game situation with toys, materials, and other things without being told to do so	Rarely
Expresses hostility to other children verbally (teasing, threats, taunts, name calling, "I don't like you," etc.)	Never

Expresses hostility to other children physically (hitting, pinching, kicking, pushing, biting)	Never
Complies with adult directives, giving little or no verbal or physical resistance, even with tasks that he/she dislikes	Frequently/Usually
Does not fuss when he/she has to wait briefly to get attention from teacher or other adult; child may be asked once to wait by the teacher or adult	Always

Appendix 4

Instructionally Focused Data-Driven Consultation Protocol

An instructional manual for consultants to use with teachers in the instructionally focused, data-driven consultation condition of the Virginia Kindergarten Readiness Program (VKRP), fall of 2015.

Introduction

The present school-based, consultee-centered consultation was created from literature on instructional (Rosenfield, 2013; Rosenfield, Gravois, & Silva, 2014) and data-driven consultation (Berndhardt, 2003; Slavin et al., 2011; Wayman & Jimerson, 2014). By drawing on the best practices from both fields, this protocol was created to ensure consultants are intervening in ways that have proven effective in classrooms. To accomplish this, three central components must be covered: preparation, discussion meeting, and follow-up. Consultation preparation includes examining the teacher's classroom-and individual-level data and reviewing the VKRP universe of strategies accessible to all teachers participating in VKRP. The discussion meeting refers to the face-to-face consultation, which occurs between the consultant and the teacher, and involves reviewing the data and determining strategies to be used. This phase also include completion of the Action Plan, which provides teachers with concrete steps to take in the classroom. The final, crucial piece of this consultation model involves following up with teachers around progress on their action plan while offering additional guidance. To implement these components effectively consultants must learn this model and continue to practice it consistently, which can be ensured through structured training and supervision. When implemented effectively, instructionally focused, data-driven consultation is expected to improve teachers' use of data-based teaching practices derived from assessment data.

I. Training/Supervision

Before a consultant can begin, he/she must be trained on the protocol. This includes reading through the Consultation Manual and participating in a role-play session with the lead consultant. The role-play session will be based off of a de-identified report provided by the lead consultant. The consultant is expected to review the data from this report according to the Consultation Manual and conduct a discussion meeting with the lead consultation as the teacher. The lead consultant then decides if the consultant is ready to begin on his/her own. If not, the lead consultant will review areas for improvement and assign a second report to review, after which the two will conduct a subsequent role-play.

Once a consultant has been trained, he/she must meet regularly with the lead consultant for supervision and discussion. In addition, supervision with a trained clinical psychologist with experience with school-based consultation and VKRP will be provided, and supervision sessions will include evaluating discussion meetings as well as reviewing teachers' fidelity checklists.

II. Consultation Preparation

The first component of instructionally focused, data-driven consultation involves preparing for the discussion meeting. After a teacher is assigned to a consultant, the consultant will send the initial email which will introduce the consultant, describe contact to be expected through consultation, and ask the teacher to review his/her Integrated Report. The consultant should then document this contact and any reply the teacher provides on the individual teacher's documentation of intervention (see Appendix A).

After sending the introductory email, the consultant is expected to prepare for the discussion meeting by reviewing the teacher's integrated classroom summary.

III. Discussion Meeting

The second component of instructionally focused, data-driven consultation is the discussion meeting. This meeting should last 45 to 60 minutes and should be held in a confidential location (i.e., without any children or adults outside the consultant and the teacher present). In addition, the consultant should be equipped with a laptop (preferably with internet connectivity), paper, writing utensils, a blank Action Plan (see Appendix 5), and the teacher's documentation of intervention. If any of these conditions is not met, the consultant should note the discrepancy on the Discussion Meeting Checklist, located at the end of the individual's teacher's documentation of intervention. Throughout or after the meeting, the consultant should indicate the portions of the meeting that occurred and any notes about the content covered.

There are four primary pieces of the discussion meeting: introduction (5 minutes), data review and completion of the action plan (25-35 minutes), review of the action plan and description of follow-up (5-10 minutes), all of which are described in detail below.

A. Discussion Meeting Introduction

The consultation will begin with the consultant introducing herself and the purpose of the consultation: to help teachers use the data from VKRP to inform their practice. The consultant will then ask how administration went and any questions or concerns. Then the consultant will introduce the overall goal for the meeting: **to review his/her data and to identify methods of intervention through the data**. Then, the consultant will indicate that to do this, they will work together to review the data and identify interventions. This portion of the consultation should last for approximately 5 minutes.

B. Discussion Meeting Data Review and Completion of Action Plan(s)

Reviewing the teachers' VKRP data will constitute the majority of the discussion meeting (estimated 25 to 30 minutes). The central goal for the data review portion is to isolate areas of interest in the data and to form interventions based on this data. To begin, the consultant will ask the teacher for his/her preference to begin with either classroom-or child-level data. The consultant will follow a similar process in either case:

1. Ask the teacher about the data:
    i. What do you notice? What stands out to you?"
    ii. In which domain were your students most successful? Which was most challenging for them?
    iii. Does this match with what you see in the classroom?
2. Engage teacher in a discussion about their interest:
    i. Where does that score come from?
    ii. What do you think that means?
    iii. Have you had similar experiences with past children/classrooms?
3. Engage in collaborative problem-solving with the teacher to form interventions around that data-interest. Collaborative problem-solving involves simultaneous and collaborative brainstorming in which the consultee and the consultant use the other's ideas to solve a problem (Pugach & Johnson, 1988; Idol et al., 1994). The "problem" surrounds the data interest the teacher identified and the "solution" will be one or more instructional strategies. To do this, the teacher and consultant will discuss how the ideal "solution" and work backwards to decide on approaches to achieve this solution. Using a laptop, the consultant should bring up the VKRP universe of strategies and strategy guide for the pair to use as a resource. NOTE: decided-upon strategies do not need to come from this list.

4. The consultant will show the Action Plan to the teacher, and together they will complete the form with their agreed upon "problem" and "solution." Additionally, the consultant will ask the teacher to specify a time, location, and child/group with which to try the strategy.

These four steps can be repeated up to three times, so that up to three "problems" with associated "solutions" can be identified and recorded on the Action Plan. The number of problem-solution cycles will likely depend on time.

C. Discussion Meeting Review Action Plan and Description of Follow-Up

After finishing the data review, the consultant and teacher will review the Action Plan they created to ensure the teacher has a concrete plan for intervention upon leaving. At this point, the consultant will describe the remainder of the intervention. The consultant will tell the teacher that she will be contacting her again in two weeks to inquire about the teacher's progress on the action plan and that she will be available to discuss any additional strategies the teacher may need via email for the next month. After that time, the teacher will be receiving the post-measures to complete and the consultation will be finished.

IV. Consultation Follow-up

One weeks after the first date on a teacher's Action Plan the consultant will send a follow-up prompt via email inquiring about progress on the action plan. The prompts will be individualized to each teacher, asking him or her to describe successes and/or challenges of chosen strategies. Emails will also include links to strategy resources that are available in an effort to reinforce what was discussed during the discussion meeting.

Consultant Log to be completed during or directly after in-person meeting. Any fields with the word “NOTE” are required. Additional notes can be added as the consultant needs.

	Y/N	**Notes**
Consultation held		
Logistics		
60 minute duration		
Confidential location		
Consultant has all materials		
Introduction		
Consultant		
Goal of consultation		
Administration questions		
Data-Review and Action Plan		
Teacher identifies area of interest in data		
Teacher and consultant arrive at a solution for problem		
Strategy implementation described in detail		
Recorded on action plan		
Review action plan		
Described follow-up contact		

Notes:

Appendix 5

Example of Action Plan Used During Consultation Sessions

To be completed during consultation meeting

Description of Data/"Problem": ____

____

____

____

Description of Strategy/ "Solution": ____

____

____

Plan for Intervention

Date: __________________ Setting: ___________________ Target: __________________

To be completed after consultation meeting

Outcome/Evaluation: ____

____

____

____

Plan for future intervention: ____

____

____

____

Appendix 6

Example of Follow-Up Prompt Sent to Treatment Teachers

Hi [TEACHER NAME],

I wanted to follow up with you regarding the action plans we created during our consultation meeting. We discussed using the tucker turtle and problem solving social stories to improve your students' general self-regulation abilities. Following your Action Plan, we set XX/XX/XXXX as the date you would read the Tucker Turtle social story and XX/XX/XXXX as the date to introduce the Problem Solving Toolkit social story. A few questions about these...

With regards to Tucker Turtle:

1) Were you able to implement this activity?

2) If yes, how did it go? Do you feel like it was successful (i.e., the targeted children participated and learned)? Were there any aspects of the activity that were not successful?

3) If no, what stopped you from implementing this activity?

With regards to the Problem Solving Toolkit:

1) Were you able to implement this activity?

2) If yes, how did it go? Do you feel like it was successful (i.e., the targeted children participated and learned)? Were there any aspects of the activity that were not successful?

3) If no, what stopped you from implementing this activity?

I enjoyed working with you a few weeks ago and I hope we can continue to work together to improve learning outcomes in your classroom! Please let me know if there is anything else I can do to help.

Thank you,

[CONSULTANT NAME]

Appendix 7

Items Associated with Outcome Constructs

Measure	Factor	Item	Available Response
Data-Use Survey	Data-Use Efficacy	I feel confident in interpreting classroom-level assessment data (e.g., scores across students' math tests) to inform my instruction.	1-6
		I feel confident in interpreting individual-level assessment data (e.g., a student's answers to a math test) to inform my instruction.	1-6
	Data-Use Practices	I alter my teaching practice as a result of information obtained from entry assessments.	1-6
		I alter my teaching practice as a result of information obtained from assessments throughout the school year.	1-6
		I individualize instruction based on results of the assessments.	1-6
		I use assessments to inform grouping children according to needs.	1-6
		I use assessments to inform classroom practice.	1-6
		I alter my teaching practice as a result of information obtained from entry assessments.	
COA-III Survey	School Accountability	Assessment provides information on how well schools are doing	1-6
		Assessment is an accurate indicator of a school's quality	1-6
		Assessment is a good way to evaluate a school	1-6
	Student Accountability	Assessment places students into categories	1-6
		Assessment is assigning a grade or level to student work	1-6
		Assessment determines if students meet qualifications standards	1-6
	Improvement	Assessment is a way to determine how much students have learned from teaching	1-6
		Assessment provides feedback to students about their performance	1-6
		Assessment is integrated with teaching practice	1-6
		Assessment establishes what students have learned	1-6
		Assessment feedbacks to students their learning needs	1-6
		Assessment information modifies ongoing teaching of students	1-6
		Assessment results are consistent	1-6

		Assessment measures students' higher order thinking skills	1-6
		Assessment helps students improve their learning	1-6
		Assessment allows different students to get different instruction	1-6
		Assessment results can be depended on	1-6
	Irrelevance	Assessment forces teachers to teach in a way against their beliefs	1-6
		Teachers conduct assessments but make little use of the results	1-6
		Assessment results should be treated cautiously because of measurement error	1-6
		Assessment is unfair to students	1-6
		Assessment results are filed & ignored	1-6
		Teachers should take into account the error and imprecision in all assessment	1-6
		Assessment interferes with teaching	1-6
		Assessment has little impact on teaching	1-6
		Assessment is an imprecise process	1-6
Data Intepret-ation Skills[a]	Math	Grab Bag Addition/Subtraction- Demonstrate placing a specified amount of counters in bag/bucket/hat/etc. Have children pull counters out and answer questions about counters either in or out of the bag.	0,1
	Math	Patterns in the Environment- Encourage children to look for and draw patters in the classroom and/or outside.	0,1
	Literacy	How Many Letters are in Your Name- Have children count the number of letters in their, family members', and other students' names.	0,1
	Literacy	Sentence Sleuth- Write nursey rhyme lines on cards, changing one word, and read the sentences out loud. Then have children determine which word was changed and write it correctly.	0,1
	Literacy	Personal Dictionary- Give children words and associated pictures to cut and paste into their own dictionary.	0,1
	Literacy	Take a Break- highlight stops (e.g., commas and periods) while reading a book to a child.	0,1
	Social Skills	Pair a student with low social skills with a student with higher social skills and ask them to complete a task together.	0,1
	Social Skills	Provide explicit social skills instruction through a story and/or role-play.	0,1

Self-Regulation	Prompt children to use cue cards with potential solutions to help children solve a conflict.	0,1
Self-Regulation	Provide physical and/or verbal praise after a child engages in a desired behavior	0,1
Self- Regulation	Use simple commands (e.g., “it’s time to clean-up the blocks” rather than “how about we clean up the blocks?”	0,1

[a] The Data Interpretation Skills Measure items were identical across classroom- and student-level reports.

Appendix 8

Domains, Subdomains, and Associated Strategy Options

Domain	Subdomain	Strategy
Math	Geometry	Triangle/Rectangle Scavenger Hunt Building Rectangles and Triangles 3D Shape Bracelets
	Patterning	Patterns I Patterns II
	Numeracy	Introduction to the 5-Frame Counting with Gummy Bears Matching Numbers & Animals
	Computation	Snap Cube Counting
Literacy	Beginning Sounds	Going on a Picnic Sound Bags
	Blending Sounds	Turtle Talk
	Sound-to-Letter Segmentation	Head, Waist, Toes
	Rhyme	Draw-a-Rhyme Stories
	Alphabet Recognition	Feely Sock
	Comprehension	Brainstorm-Reflect-Reformulate
	Concept of Word	Sentence Sleuth
	Letter Sounds	Photo Line
	Print Knowledge	Take a Break
	Vocabulary	Personal Dictionary Word Search
	Writing	Squiggle and Write
	Spelling	Picture Hunt
Social Skills	--	Supporting Friendship Skills

		Supporting Problem-Solving Skills
Self-Regulation	--	Reinforcement Cues & Visuals Consequences Effective Commands Supporting Emotion Regulation

Appendix 9

Teacher Satisfaction Survey Items

Please indicate your agreement with the following sentences (1 = Strongly Disagree; 6 = Strongly Agree)

I felt comfortable interpreting the data from VKRP on my own.

I found the consultant's help with interpreting the VKRP data useful.

It was useful to cover the classroom-wide data during the consultation.

It was useful to cover individual students' data during the consultation.

I found working with consultant to brainstorm strategies easy.

I found working with consultant to brainstorm strategies useful.

It was useful to discuss classroom-wide strategies during the consultation

It was useful to discuss strategies for individual children during the consultation.

I implemented the strategies/activities we discussed in my classroom.

I utilized action plans after the consultation session.

Overall I found the consultation useful in improving students' learning/behavior.

Overall I found the consultation useful in informing my teaching practices.