2022

ASSOCIATIONS BETWEEN YEARS OF TEACHING EXPERIENCE AND CLASSROOM MANAGEMENT PRACTICES, RESPONSIVENSS, AND ACCEPTABILITY TO A COACHING INTERVENTION

Nicole M. Peterson
Virginia Commonwealth University

ASSOCIATIONS BETWEEN YEARS OF TEACHING EXPERIENCE AND CLASSROOM MANAGEMENT PRACTICES, RESPONSIVENSS, AND ACCEPTABILITY TO A COACHING INTERVENTION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University

by

Nicole M. Peterson
Bachelor of Arts, Elementary Education & Anthropology, Roger Williams University, 2004
Master of Education, Special Education, Virginia Commonwealth University, 2012

Chair: Kevin S. Sutherland, Ph.D.
Professor, Department of Counseling and Special Education

Virginia Commonwealth University
Richmond, VA
April 2022

Acknowledgements

It is both and honor and a privilege to have had the opportunity to engage in this process. Any success that I have experienced can be attributed to the long line of people who came before me, and those that will follow. I have felt support from my professors, committee members, and peers in an unlimited number of ways. I am motivated by (and in awe of) your commitment to finding ways to support learners and advocate for their needs. Thank you for making this your passion.

To be both a teacher and a learner is a beautiful circumstance that I hold in the highest regard. For this, I am forever grateful.

Contents

List of Tables

List of Figures

Abstract

YEARS OF TEACHING EXPERIENCE AND CLASSROOM MANAGEMENT PRACTICES, RESPONSIVENSS, AND ACCEPTABILITY TO A COACHING INTERVENTION

By Nicole M. Peterson, M.Ed.

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University

Virginia Commonwealth University, 2022

Director: Kevin S. Sutherland, Ph.D., Professor, Department of Counseling and Special Education

This study aimed to bridge the research-to-practice gap related to in-service supports and professional development by identifying factors associated with teacher use of evidence-based classroom management practices and responsiveness to a coaching intervention. Years of teaching experience was identified as a teacher-level intervention determinant with a strong research base in the classroom management literature; however few studies used direct observation of behavior to report difference among teacher behavior. This study used quantitative statistical methods to analyze data from a multi-year parent study investigating the effects of a coaching intervention (BEST in CLASS-Elementary) aimed at increasing the use of evidence-based practices with students who display patterns of challenging behaviors. Using data from an intervention development year plus the first two years of a four-year investigation, 83 teachers were randomized into treatment and control groups to either receive 14 weeks of a coaching intervention or continue in a business-as-usual condition. Data for this investigation was collected using both direct and indirect measures collected before, during, and after the intervention. Results of regression models suggested a non-significant relation between years of teaching experience and practice use pre-intervention, after adjusting for students within

teachers. Similarly, results of regression models post-intervention suggest a non-significant relation between years of teaching experience and practice use during coaching, after controlling for baseline scores and adjusting for students within teachers. Experience was significantly and negatively associated with teacher reports of acceptability ($B = -.27$, $p = .046$). The predictive ability of experience on understanding was approaching significance ($B = -.25$, $p = .058$) but the relation between years of experience and feasibility was not significant ($B = .06$, $p = .668$). While years of experience did not predict practice use or responsiveness, findings may position teacher-driven interventions (i.e. coaching) as a mechanism to improve teacher practice regardless of previous teaching experience.

Chapter 1
Introduction

It has long been understood that teacher quality is related to student outcomes (e.g., Darling-Hammond, 2000; Kunter et al., 2013; Stronge et al., 2007), therefore efforts to increase our understanding of teachers' acquisition of new skills and their implementation of evidence-based practices that increase student learning are essential. Students who display patterns of challenging behavior and who may have or be at risk for emotional behavior disorders (EBDs) have been shown to have particularly poor outcomes (Sutherland et al., 2008) due in part to teacher's lack of knowledge of evidence-based practices (Ficarra & Quinn, 2014) and ability to employ evidence-based practices effectively (Gable et al., 2012). While many factors may contribute to student outcomes, Marzano et al. (2003) posits "We live in an era when research tells us that the teacher is probably the single most important factor affecting student achievement—at least the single most important factor that we can do much about" (p. 1).

In the United States, teachers are licensed at the state level, with most types requiring a post-secondary degree or training (U.S. Department of Education, 2021). Further, once a teaching license is held, it must be renewed (usually after a period of 3, 5, or 10 years) and that process includes the teacher demonstrating that they have engaged in some form of ongoing professional development. This professional development is typically in the form of workshops, webinars, or university-sponsored courses. It could also be participation in mentoring or induction programs for teachers new to the field or in a new teaching role (Darling Hammond et al., 2017). This information is important to the discussion about teacher quality because it begins to illuminate the distinctions between different teacher groups, a tenet to this investigation. For

example, although there are alternative programs that provide routes to teacher licensure, most teachers at the elementary level who are newer to the field have completed a post-secondary degree program, and may have 1-3 years of experience in the classroom (Whitford et al., 2018). This is in contrast to veteran teachers, most who have completed a post-secondary degree, have amassed many years of experience in the classroom, and have engaged in on-going professional development throughout their careers. Several professional development models that are used to support teacher skill development generally fall into one of two categories: those that occur once, such as a workshop or webinar, and those that are on-going and designed to support teacher development over time. Examples of once-occurring professional development are workshops, often referred to as "grab and go" or "one shot" opportunities where teachers are exposed to a teaching strategy or content delivery method and then are expected to apply this new knowledge to their classrooms independently. Another is the classroom observation, where an administrator, colleague, or instructional specialist observes the teacher leading a lesson and provides feedback in written form or during a post-intervention meeting. Ongoing models may include multiple-day or periodic workshops that occur over several sessions. Also, collaborative or group-based opportunities where groups of teachers discuss new content, reflect on practices, or examine student growth patterns with one another or as part of a cohort or professional learning community (PLC). A final type of ongoing professional development model, which is central to this investigation is coaching, where teachers receive on-going, tailored support driven by their particular skillset.

Teacher Coaching

This form of professional development has garnered a significant amount of attention due to its documented ability to increase teacher quality and lead to improved outcomes for students

(Darling-Hammond et al., 2017; Joyce & Showers, 2002; Kraft et al., 2018; Kretlow et al., 2012). It is no surprise then that coaching has been described as the fastest rising form of professional development (Carver-Thomas & Darling-Hammond, 2017). In a coaching environment, teachers are paired with coaches who use various tools to identify specific areas of growth for the teacher and make purposeful changes to teacher behavior. These changes may be to implement evidence-based instructional or behavior support practices or to increase the quality of practice use (Kraft et al., 2018).

Implementation Science

The field of implementation science, when applied to school settings, aims to identify factors that contribute to teachers' adoption and sustained use of evidence-based practices (Sutherland et al., 2021). Within conceptual models of implementation science are distinct components that affect intervention outcomes, outcomes that include the continued use of practices after the intervention concludes. Applied specifically to students with EBD, McLeod and colleagues (2020) describe the role of implementation determinants, or naturally occurring factors that may serve to enable or obstruct the implementation of an intervention. Determinants include factors across several levels including policy, school, classroom, and the individual delivering the intervention. Related to the individual, the authors highlight the role of delivering the practice with integrity, or according to protocol, to "foster settings that provide a foundation for the long-term academic and social-emotional success of youth" (p.426). In a discussion of implementation science in the field of special education, Sutherland et al. (2021) expand on treatment integrity as it relates to teachers delivering interventions in classrooms. The authors identify five dimensions, each of which may have a different influence on student outcomes and are essential to understanding the overall effectiveness of an intervention. These dimensions

include *adherence* (the frequency and thoroughness of practice delivery), *competence* (the quality of practice delivery), *differentiation* (the extent that additional practices outside of the intended intervention are used), *adaptation* (the ability to alter the practice to match student need), and *student responsiveness* (how students respond to the teacher's practice delivery). The authors elaborate on the role of student responsiveness in their analysis of intervention outcomes, particularly how responsiveness may contribute to a more nuanced understanding of student outcomes.

One last component of implementation science pertinent to this discussion is implementation outcomes, or indicators associated with successful implementation and adoption of an intervention or practice (McLeod et al., 2020; Sutherland et al., 2021). Proctor et al., (2011) identify eight specific, measurable indicators including acceptability, adoption, appropriateness, and feasibility; these outcomes can contribute to the overall understanding of long-term adoption and integration of practices.

Applying an implementation science lens to a coaching intervention would ask distinct questions at different phases of the intervention and contribute to the overall understanding of intervention outcomes. A more in-depth analysis of teacher responsiveness or the degree to which teachers are responsive to the coach's efforts to deliver core components during the intervention would help explain the effect of the intervention on student outcomes. Similarly, an understanding about implementation outcomes (e.g., acceptability, feasibility) after the intervention is complete may also provide insight into the long-term uptake of an evidence-based practice employed by teachers.

Lastly, an increased understanding of teacher-level determinants may also provide valuable understanding. In a seminal 2005 article, Han and Weiss identified specific factors

related to the successful implementation and sustainability of school-based mental health intervention programs including the teacher-level factors of self-efficacy, burnout, and pre-treatment acceptability. Regarding treatment acceptability, the authors described several teacher characteristics that may be influenced by years of teaching experience, for example knowledge of behavior principles and previous experiences with interventions.

While Han and Weiss's (2005) work relied on a diverse set of literature, what is less understood are the factors that contribute to a teacher's engagement in an intervention that specifically targets the use of evidence-based practices to reduce problematic student behavior. One example of this type of investigation is an analysis of teacher implementation integrity published by Sutherland et al. (2018). Using data from BEST in CLASS, a tier 2 coaching intervention designed for early childhood settings, the authors examined the effect of several teacher-level factors, including education background, on implementation outcomes. Findings suggest that teachers with higher levels of education delivered the program with a higher implementation integrity initially, however those with lower initial scores demonstrated higher levels of growth over time. In the discussion of results, the authors suggest that the positive outcomes for both sets of teachers provides promising support for the intervention delivery format (i.e., coaching) as an effective support for teachers with varying levels of education. These outcomes draw a distinct parallel to themes that are directly aligned to this study, and will be discussed in future sections.

Teacher Experience as an Intervention Determinant

Teacher experience is a determinant that may play a particularly important role in our current understanding of teacher responsiveness and may become increasingly important in the future. In addition to the aforementioned differences including exposure to ongoing professional

development that have extended veteran teachers' pedagogical knowledge, the teacher workforce itself is changing. The teaching profession is currently experiencing a high rate of attrition, leading to classrooms staffed with newer and less-qualified teachers (Carver-Thomas & Darling-Hammond, 2017). It is expected that this trend will continue, especially as the nation rebounds from the current health crisis (Carver-Thomas et al., 2021). Therefore, it is necessary to consider how to effectively meet the needs of a teaching force with less experience.

In addition to the changing workforce, a shift in student need has also been identified. Developments in education legislation, policy, and human rights movements over the last several decades have led to increased numbers of students with disabilities receiving instruction in the general education setting (Francisco et al., 2020). At the most recent count, most students with disabilities spend the majority of their day in inclusive classroom settings and a distinct upward trend has be seen since the year 2000 (Hussar et al., 2020). Several studies report that teachers generally favor the inclusion of students with disabilities in general education settings, however teacher opinion tends to be based on access to training and resources, which many teachers report a lack thereof (Pearce & Forlin, 2016; Scruggs & Mastropieri, 1996). An examination of the literature by Salend et al. (1999) regarding the impacts of inclusion on students and their teachers revealed characteristics particular to general education teachers including inability to address behavioral challenges of students with disabilities, lack of funds to support instructional needs, rigid requirements of the curriculum, and limited time for collaboration with other teachers. Further, when comparing academic achievement, students with disabilities continue to lag behind their typically developing peers (OSERS, 2020).

The combination of students with increased needs and teachers unable to provide support for these needs may contribute to more challenging student behaviors in the classroom.

The ability to manage classroom behaviors has long been identified as an essential component of effective learning environments (OSERS, 2020). Novice and experienced teachers alike identify classroom management as an essential skill in high quality teaching and report seeking continuous professional development in classroom management throughout their careers (Stough et al., 2015). If teachers are not equipped with the skills to deliver evidence-based practices appropriately, the students who exhibit these behaviors and their peers will not be able to achieve to their potential.

To mitigate the impact of these problems, a better understanding is needed about how teachers with varying levels of experience approach classroom management and employ evidence-based classroom management practices. By utilizing an implementation science lens to ask questions related to teacher responsiveness and implementation outcomes within a coaching intervention, we might better understand how to design and implement interventions in ways that increase long-term teacher adoption of practices. Adoption of evidence-based practices will undoubtedly contribute to teacher quality and thus, better student outcomes.

Conceptual Framework

Several conceptualizations of how teachers develop expertise and the process that leads to high quality teaching exist. A consistent aspect of all of these models is the emphasis on change over time, specifically the role of lived classroom experiences in influencing a teacher's understanding of what works and what does not. In fact, this line of thinking is mirrored in the clinical, student teaching experiences compulsory to most teacher preparation programs. It is assumed that the completion of course work alone does not make a high quality teacher, rather learned skills must also be applied in authentic settings under the guidance of an experienced teacher (Darling-Hammond & Cobb, 1995; Ronfeldt & Reininger, 2012). In many ways, this

feature of teacher preparation programs further highlights the essential role of lived classroom experience on teacher skill development.

For example, *adaptive expertise* describes the particular ability of a high-quality teacher to respond purposefully to classroom experiences and adjust their decision-making as a result of previous lived experiences (Soslau, 2012). Further, an adaptive expert is flexible and "able to strategically move away from planned curriculum components to better support the contextual needs of their pupils, question familiar solutions to problems by noticing unique features, and recognize the need to refine, change, and try out different decisions while paying close attention to the impact on their pupils" (Soslau, 2012, p. 768). Here, the role of lived experiences is posited as an essential element to gaining teaching expertise.

In a similar model Steele's (2010) framework for teacher development establishes four distinct stages where, similar to popular theories of child development, a teacher progresses in chronological order. The stages of *unaware*, *aware*, and *capable* lead to the *inspired teacher*, which is characterized by three essential qualities:

> First, the teacher has a broad, deep understanding of the subject being taught, developed through coursework, life experience, and continuous refinement... Second, the teacher has a wide repertoire of teaching techniques—also fine-tuned over time—and is comfortable and competent with each one...Third, an inspired teacher can also "read" students, situations, settings, and reactions and can select apt responses so that learning goes smoothly...The apparently intuitive responses of expert teachers reflect the distillation of months or years of learning until the essential understanding is in their bones. (p.64)

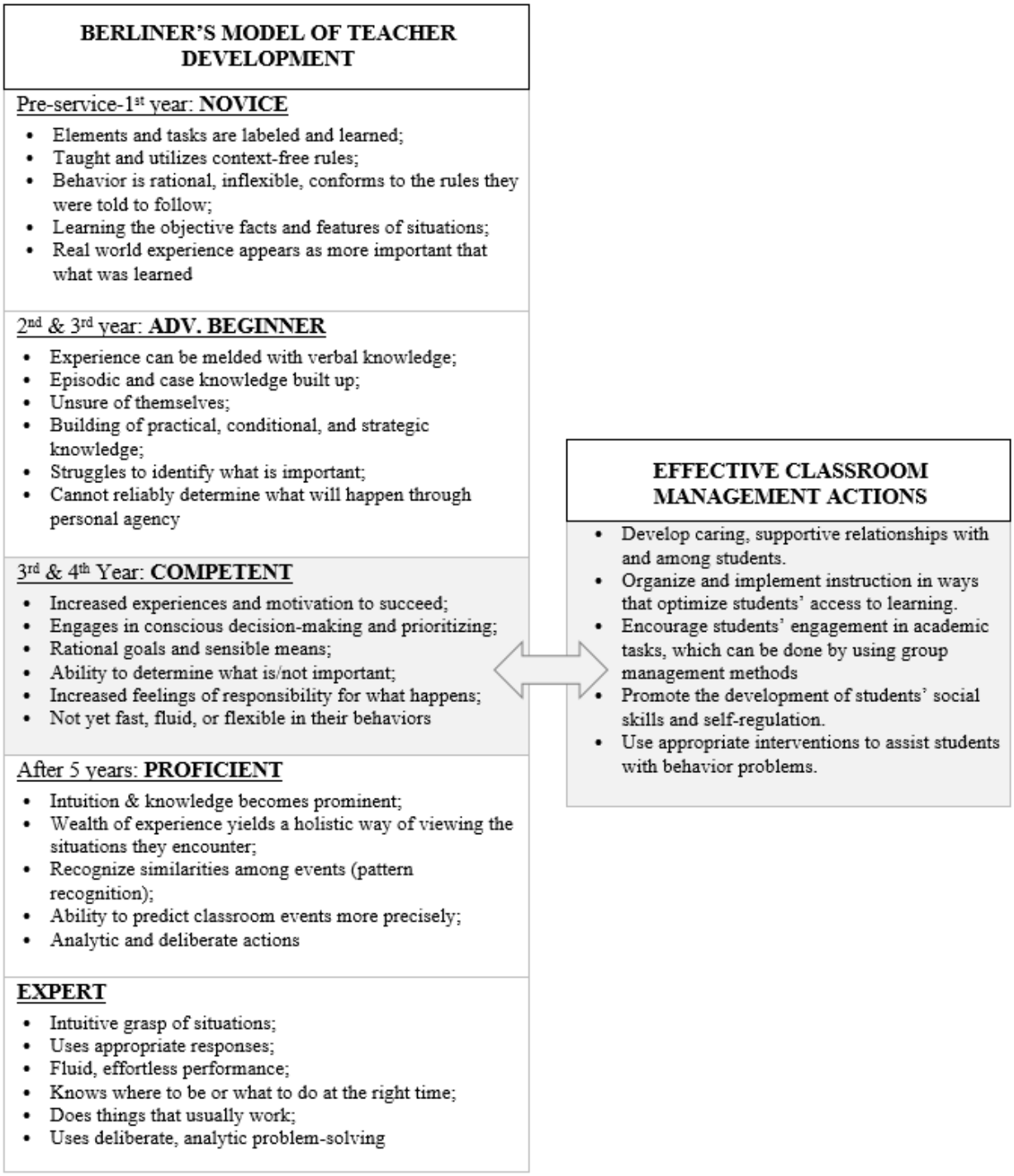
Additionally, the author describes other teaching abilities that change over time, namely the ability to notice relevant details about classroom dynamics and the students, the ability to

recognize patterns and respond meaningfully, and allowing context to drive decision-making. Here again, the role of lived experiences and an opportunity to establish and test hypotheses related to effective teaching and learning outcomes are prioritized.

While both of these conceptualizations speak to the growth teachers experience as a result of time spent in classrooms, a theoretical model that effectively addresses questions related to experience and classroom management will also propose timelines that can support a line of inquiry related to years of experience. Similar to Steele (2010), Berliner (1988) proposed a theory of teacher development using a set of chronological stages. Based on outcomes from their own research, the author identifies the cognitive understanding at each stage, examples of the ways that this cognitive understanding can translate into specific classroom actions, and the typical years of experience needed to reach each stage (see Figure 1 for a description of each stage). For example, it is not until the teacher reaches the *competent* stage that they are able to prioritize and make conscious choices about their actions and choose sensible means to obtain rational goals. Further, it is at this stage that teachers develop an increased feeling of responsibility for what happens in their classrooms, drawing a parallel between teacher actions and student outcomes. Essential to this investigation is the prioritization that Berliner places on the cognitive ability at each stage, particularly the role of lived experiences to drive cognition.

Figure 1

Relationship between Berliner's Model of Teacher Development and Evertson and Weinstein's Effective Classroom Management Actions



To better apply this model to classroom management and the implementation of evidence-based practices, additional criteria are needed. In the introduction to the Handbook of Classroom Management: Research, Practice, and Contemporary Issues (2006), Evertson and Weinstein draw on the work of several theorists to define classroom management as "the actions teachers take to create an environment that supports and facilitates both academic and social emotional learning" (p. 4). In order to accomplish this, the authors detail five required tasks of teachers:

> (1) to develop caring, supportive relationships with and among students; (2) organize and implement instruction in ways that optimize students' access to learning; (3) use group management methods that encourage students' engagement in academic tasks; (4) promote the development of students' social skills and self regulation; and (5) use appropriate interventions to assist students with behavior problems. (p. 5)

Mapped onto Berliner's model, it is not until teachers reach advanced stages of development that they have the cognitive ability to approach these tasks. Figure 1 shows the relationship between Berliner's Model of Teacher Development and Evertson and Weinstein's classroom management actions. The higher-order skills that appear to be required to establish effective classroom management provide further justification for an investigation looking at years of teaching experience and the use of classroom management practices.

Research Aims and Methodology

To investigate differences among teachers with varying levels of experience, data from a multi-year parent study investigating the effects of a coaching intervention (BEST in CLASS-Elementary) aimed at increasing the use of evidence-based practices with students who display patterns of challenging behaviors was examined using quantitative statistical methods. During an

intervention development year plus the first two years of a four-year investigation, 83 teachers were randomized into treatment and control groups to either receive 14 weeks of a coaching intervention or continue in a business-as-usual condition. Teachers were eligible for participation if they taught students in grades Kindergarten through third grade and had at least one student that showed patterns of externalizing behaviors, as identified by a behavioral screener. Data for this investigation was collected using both direct and indirect measures collected before, during, and after the intervention. The main aim of the study is to determine the predictive value of teacher experience on the use of classroom management practices, responsiveness to a coaching intervention, and intervention acceptability.

It is expected that outcomes from this study will add to the existing classroom management literature with a report of descriptive data related to teachers' use of evidence-based practices. Specific attention was placed on the use of evidence-based practices for students who have or at risk for EBD. Additionally, an analysis of teachers' responsiveness throughout the intervention will contribute to work designed to develop and test implementation strategies aimed at maximizing the use and impact of evidence-based practices. This increased understanding could guide development of supports for in-service teachers, particularly as less-experienced teachers fill more classrooms. In Chapter 2, a more specific review of the teacher experience literature will describe how teacher experience and expertise are conceptualized, the current research landscape, and a review of what is known about the relation between experience and classroom management. Additionally, gaps in current understanding will be explored, which set the stage for the present study.

Chapter 2

Review of Literature

In the previous chapter, classroom management practices were identified as essential skills for teachers, particularly as student and teacher demographics change. In Chapter 2, a review of empirical studies from the teacher experience literature will be presented, including outcomes related to perceptions and use of classroom management practices. The purpose of this review is to identify gaps in understanding by answering questions related to how teacher experience and expertise are conceptualized, the current research landscape, and a review of what is known about the relation between experience and classroom management. First, background context will be presented using a broad view of teacher experience. Then, outcomes related specifically to classroom management will be explored. The chapter will close with a presentation of the specific research questions guiding this investigation.

A Broad View of Teacher Experience Research

Looking specifically at empirical research from the United States, experience has been examined with a wide variety of dependent variables that span the breadth of topics related to instruction, teacher attitude and ability, and classroom management. Additionally, a group of studies have used longitudinal designs to measure changes over time, highlighting the role of experience on teacher development. The sections that follow provide examples of these lines of research and associated outcomes.

Instructional Outcomes

To measure differences in math instruction, Leinhardt (1989) compared data from classroom observations of two "novices" (student teachers) and four "experts" (teachers whose students had performed well on growth measures). Even with a small sample, the authors found distinct differences in instructional approaches including the expert teachers' ability to string together concepts that better support the learner, present content in a fluid structure using appropriate representational figures, and use skilled judgement in their use of practice and repetition in response to student need. Conversely, outcomes from another investigation found no statistically significant relationship between the teacher-level factors of age, sex, and years of experience on mathematics achievement (Paulsen, 1977). McDonald et al. (2005) used a complex model that included external factors such as socio-economic status, home learning experiences, and preschool preparedness. The authors reported a stronger relationship between the teacher-level factors of personality and years of education (degree type) to reading achievement than years of teaching experience. In one study particular to students with EBD, Gage et al., (2017) reported null effects when examining student achievement as it relates to the teachers' education level, certification status, and years of experience. Important to note however are the null effects for change in student achievement over time, which may skew outcomes related to teacher experience.

The variability in these findings are supported by a comprehensive report from the National Center for Analysis of Longitudinal Data in Education Research (CALDER; Rice, 2010) in which several decades of state testing data were analyzed to determine the impact of teacher experience on student achievement. Outcomes suggest that the variable of teacher experience is more impactful during the first few years of teaching, with differences between experience groups trailing off as years pass. Further, these differences were more pronounced in

math instruction than reading and at the elementary and middle school level than in high school. The authors also identify some critical barriers to the investigation, which may have direct implications for the current study; for example, the tendency for less-experienced teachers to be teaching in under-resourced schools. Implications for this investigation will be further explored in the discussion section.

Teacher Knowledge, Attitude, and Decision-making

In addition to student achievement, a large body of research has investigated teacher experience in relation to a variety of dependent variables that may be related to classroom management including instructional tendencies, self-efficacy, pedagogical knowledge, and response to student need. Similar to the aforementioned studies measuring student academic variables, this set of studies does not present conclusive outcomes. Examining differences in classroom observations of teachers in different experience groups (pre-service, novice, and experienced) across different settings (rural, suburban, urban), Everhart and Vaugh (2005) found differences in instructional tendencies including the use of movement and the proportion of managerial versus direct instruction. Similar differences between experience groups were reported by Kan and Bulut (2014) who investigated grading and scoring patterns, however these outcomes are not aligned with previous work done by Crowl and Berkowitz (1985) whose investigation revealed a small, non-significant correlation between grades, attitude scores and teacher experience.

Mixed outcomes have also been reported in studies whose focus was on teachers' response to student need. Specific to special education, MacFarlane and Woolfson (2013) examined a host of teacher-level variables and their relation to attitudes and behavior towards including students with EBD in inclusive learning settings. The authors report that teachers with

more experience showed a decreased willingness to work with students in this disability category, however this finding did not translate to significant differences in teacher behavior. Looking specifically at referral for special education services, a process that requires teachers to evaluate students based on student need and potential, Egyed and Short (2006) found a significant, positive relationship for burnout but no other teacher-level variables, including experience.

The variable of teacher experience is more consistently linked to teacher self-efficacy. For example, Belibas and Liu (2017) found that experienced teachers reported higher levels of self-efficacy, including on the classroom management subscale. To examine more specific trends related to the self-efficacy subscales, Klassen and Chui (2010) used a large sample (n = 1,430) and experience as a continuous variable, which uncovered a nonlinear relationship. A steady increase across all subscales was seen, with a peak at around 25 years of experience followed by a similar decrease over time.

Lastly, a group of studies using longitudinal designs examined change in teacher perceptions and behavior related to instructional methods and classroom organization. These studies highlight the role of classroom experience on the development of teacher ability. For example, several studies focused on the impact of clinical classroom experiences for pre-service teachers and reported an increased ability to identify appropriate responses to student scenarios (Hoy & Woolfolk,1990). A study of 87 first year teachers in urban settings specifically examined the development of classroom management skills (Kwok, 2018). Teachers reported that while their preparation programs provided knowledge about classroom management, feedback from school personnel and first-hand classroom experiences contributed more to their skill development. Adam (1982) reported changes in teachers over a six-year period and found

student discipline and motivation to be a reported concern throughout experience levels. Additionally, a decrease in concern related to one's self and an increase in concerns related to the task of teaching and the impact of instruction on students were also reported. Analysis of classroom observation data revealed significant changes in ratings of organized/systematic teacher behavior, affective oriented teacher behavior, and surgent/stimulating teacher behavior between years one and three of experience. Further, the author reported a significant relation between experience and student behavior in the elementary-level participants. In the sections that follow, analysis of literature specific to classroom management and behavior support will continue to frame our understanding of the teacher experience variable.

Classroom Management

An initial review of the literature suggests that while the effect of teacher experience on classroom management has been widely studied, closer examination reveals several gaps in understanding. These gaps set the stage for this proposed study and will be presented in the sections that follow. A scoping review of empirical research was conducted to frame the existing literature base around the topic of teacher experience level and classroom management. Studies conducted in the United States served as the focus for the review, as cultural factors may influence the role of the teacher and subsequent responses to behavior in schools. Additionally, the pathway of teacher preparation to licensure is similar across each of the states but may differ in other countries. The review process resulted in approximately 4,000 published articles, in which a collection of approximately 50 representative studies conducted in the United States were used to answer questions related to the conceptualization of experience and expertise, the research landscape, and intervention responsiveness.

Conceptualization of Experience

A discrepancy was found in the ways that experience and expertise is conceptualized in the classroom management literature. In some studies, being an "experienced" teacher was not related to the number of years of teaching experience, but rather related to results from a knowledge assessment or achieved accolade such as being recognized as the "teacher of the year," obtaining certification from national teacher organizations, or principal nomination. For example, in Pressley et al. (2020) teachers were grouped by years of teaching experience and scores received in their teaching evaluations (presumably done by administrators). "Novice" teachers had either less than three years of experience or received low scores on their evaluations. To compile a group of "highly effective" teachers, the authors used experience, social nomination, teaching evaluation scores, and membership in teaching organizations. Here, "teaching level" was confounded by years of experience, subjective performance outcomes, and other external variables. Similarly, Clarridge (1990) applied Berliner's definition of "expert" to establish a group of expert teachers who all had more than five years of experience and were nominated by their principal or superintendent. In both examples, expertise and years of experience become mutually exclusive groups; in this schema, a teacher with less than three years of experience could not be an expert and an expert must have more than five years of experience. This is important because it illustrates the various ways that the process of teacher development and expertise are viewed and acknowledges that there is not an established norm. Further, it limits the ability to generalize outcomes.

In studies in which years of teaching experience have been quantified, both categorical and continuous variables were used. Gregory et al. (2014) and Simonsen et al. (2020) both examined years of teaching experience as a continuous variable, predicting student engagement

and use of evidence-based classroom management, respectively. In studies where categorical groups were formed based on years of teaching experience, authors either relied on theoretical frameworks like Berliner's or used a benchmark established from workforce trends. For example, Melnick and Meister (2008) used more than three years of teaching experience as a benchmark to examine differences between the concerns of experienced and beginning teachers. The authors established this distinction based on literature supporting the typical tenure point for the teaching profession. One final variability about the conceptualization of experience that was noted is related to the classification of the pre-service teacher, who is typically in their final year of a preparation program. Some authors grouped these teachers with a few years of experience, while others considered pre-service teachers in a category unto themselves.

Teacher Perceptions & Reported Practice Use

A greater proportion of studies focused on the examination of teacher perceptions of classroom management ability, style, and practices using survey data. Representation across both elementary and secondary levels was found, with a great deal of studies including teachers across both developmental levels. Overall, these studies had large participant groups, ranging from 51 to 1,062 participants with a few smaller studies using qualitative methods. While an overwhelming number of studies reported significant findings, a small number of studies reported null effects (e.g., Lock & Liechtenstein, 1980; Peters et al., 2014), suggesting minimal publication bias on the topic of teacher perceptions.

Regarding teacher's overall perceptions of classroom management self-efficacy, a positive association was found between reported perceptions of ability and years of experience (Melnick & Meister, 2008; Shoulders, & Scott, 2015). Regarding teachers' overall feelings towards students, one study reported experienced teachers holding more positive outlooks for all

student groups, including students who display aggressive behaviors (Podell & Tournaki, 2007). This is somewhat aligned to another set of outcomes where teachers who attended training held more positive attitudes towards students with social, emotional, and behavioral difficulties (MacFarlane & Woolfson, 2013).

A group of studies related to teacher-reported responses to student behavior (both actual and hypothetical) were also examined. In a large (n = 1,062) study investigating teachers' likelihood to intervene in bullying scenarios, Duong and Bradshaw (2013) found differences in the role that perceived threat played in a teacher's likelihood to intervene; experienced teachers were more likely to intervene based on their previous experiences with bullying at the school. Here again, teachers' behaviors appear to be shaped in part by their experiences. Interestingly, the authors found similarities in the role of perceived efficacy and intervening for both teacher groups. Differences were also reported in Noltemeyer et al. (2012), who found less-experienced teachers less-likely to ignore behaviors and choose a more reactive approach to discipline. Similarly, outcomes from Feuerborn and Chinn (2012) suggest that when responding to challenging behaviors, teachers with more experience relied more on their own direct encounters with students rather than reports from their colleagues, promoted building relationships with students, were less concerned with defiant behaviors, and tended to view a student who questions the teacher's authority as having leadership potential.

The aforementioned differences between experience groups may be better understood when we incorporate data that reports on teachers' classroom awareness and ability to contextualize classroom events. Outcomes from a small pocket of studies conducted in the United States reveal increased ability for experienced teachers to demonstrate connectedness between classroom events (Needels, 1991), increased awareness of classroom interactions and

ability to notice changes during a lesson (Cushing et al., 1992; Kearney et al., 1988). While the insights into experience groups provided by these sets of outcomes are important, they may lack validity as they rely on self-reported data to draw conclusions. In the next section, studies that use direct observation of teacher behavior are synthesized.

Direct Observation of Teacher Behavior

In contrast to studies reporting on teacher perceptions or reports, only a few studies used direct observation of teacher behavior in their data collection methods and a greater proportion reported null findings. This is unexpected as so many of the previously-mentioned studies reported significant differences among perceptions and reported practice use between experience groups. The recent publication dates of the studies in this direct observation group may suggest that this is an emerging topic or that self-reported teacher behavior differs drastically when compared to observed behavior. A third option, which will be explored in greater detail below and in the proposed investigation, are the possible mitigating effects of teacher-focused interventions.

In three observational studies, significant findings were reported when comparing the behaviors of novice and experienced teachers. In a study of 70 elementary teachers who either had less than one or more than five years of experience, researchers observed experienced teachers exhibiting increased levels of classroom communication and flexibility in their approach with students (O'Connor et al. 2004). Similarly, in a study of 36 secondary teachers Dubinski et al. (2016) observed differences between student teachers and experts with a mean of 8 years of experience using a matched samples design, which paired teachers in like grade levels, content area, and school characteristics. Analyses revealed a tendency for novice teachers to interact with students in a more managerial way versus experienced teachers who were more likely to interact

with students in an instructional way (Weber et al., 2016). Classroom management styles were also examined in a mixed-methods study of 80 secondary school teachers where an interaction effect between management style and years of experience and certification type was found (Ritter & Hancock, 2007). While experience level alone did not predict interventionist, non-interventionist, or interactionist teacher behavior, the combination of years of experience and a degree from a four-year preparation program appears to impact the degree to which student need influences teacher actions. Data collected from qualitative methods supported the author's conclusions that while a teacher preparation program may provide access to knowledge about teacher pedagogy, it's the experience in the classroom coupled with that pedagogical knowledge that influences teacher behavior, particularly in their ability to employ effective classroom management strategies.

Intervention Responsiveness

An interesting commonality was found between two studies that reported null findings related to years of teaching experience: participation in an intervention. In their 2014 randomized controlled trial measuring the effects of a coaching intervention to increase student behavioral engagement (*My Teaching Partner-Secondary*), Gregory et al. reported no change in intervention effects for any teacher-level characteristics, including years of experience. Similarly, in a cross-over design where teachers received targeted professional development of specific classroom management practices (training plus on-going self-monitoring activities), years of teaching experience was not a statistically significant predictor of practice use (Simonsen et al., 2020). In another investigation where teachers receive coaching to implement instructional practices that embed social-emotional learning into a school's existing curricula (*Responsive Classroom*), Rimm-Kaufman et al. (2014) reported a relationship between years of teaching experience and

treatment fidelity. However, it is important to note that years of experience was averaged across several years of the study and aggregated at the school level and not at the teacher-level, rendering the effects of at the teacher level unclear. In a study comparing outcomes from two professional development interventions at the preschool-level aimed at strengthening teacher-child interactions (*Making the Most of Classroom Interactions* and *My Teaching Partner)*, Early and colleagues (2017) did not report differences in intervention fidelity or outcomes associated with the variable of teacher experience.

While the degree to which teacher experience has been investigated as an independent variable cannot be assumed from this limited number of studies, these outcomes undoubtedly raise important questions about the potential for teacher-focused interventions to mitigate the impact of years of experience. For example, what impact do years of experience have on teachers' responsiveness to an intervention? Can an intervention such as coaching be just as effective for teachers of any experience level? This study may answer these questions and help us better understand the development of classroom management ability.

Discussion

The purpose of this scoping review was to identify gaps in understanding by answering questions related to how teacher experience and expertise are conceptualized, the current research landscape, and a review of what is known about the relation between experience and classroom management. A set of 50 unique articles, representing studies conducted in the United States that explored the impact of teacher experience as an independent variable on elements of classroom management, were synthesized. The literature was representative across elementary and secondary levels, included both general education and special education teachers, and spanned a time period over several decades. Analysis of the literature revealed several distinct

themes, which have contributed to both the design and development of research questions for this investigation.

First, there is extensive literature to support that teachers may approach classroom management differently as they gain professional experience, suggesting that experience may be valuable to the topic of development of teacher quality. This can be seen in the quantity of studies that report significant associations as well as the diversity of topics. For example, the breadth of dependent variables included perceptions of classroom management self-efficacy, reported use of specific practices, and actual and hypothetical responses to student needs. Additionally, outcomes from a set of longitudinal studies demonstrate change over time.

Second, it appears that the conceptualization of *experience* and *expertise* within the research landscape is not consistent, and outcomes from these studies should be carefully interpreted. This is particularly important when subjective measures are used to characterize teacher quality. For example, in studies that characterized "experienced teachers" as those which were members of a professional organization or who had received nomination from their colleagues or superiors, it should be noted that these markers of teacher quality are less connected to the research about effective classrooms than measures such as classroom climate or student growth. There is value in approaches that aim to discover characteristics associated with teaching expertise and these investigations should rely on data collected from measures that have validity within the research community. Further, a distinction is needed to either prioritize expertise or years of experience as the independent or grouping variable, so as to not allow for membership in more than one group.

Similarly, a third theme related to the participant group also emerged. A limitation described by Rice (2010) in their longitudinal analysis on teacher experience and student achievement highlights a pattern of employment in the teaching profession that may have biased their

outcomes. Teachers with less than three years of experience are more likely to teach in high-poverty, low-resourced classrooms, which may have direct consequences on student achievement rates. Second, the differing attrition rates of highly effective teachers versus less-effective teachers plays an interesting role. Ways to minimize this bias would be to eliminate differences by sampling from similar school settings or using a measure to create a descriptive variable that can be controlled for.

Fourth, outcomes from studies using direct observation of classroom management behaviors are very limited, and would contribute to our understanding of the factors that influence development of teacher quality. While teacher-report and survey measures can produce valid findings, there may be inherent bias due to inconsistencies in reporting (Biemer & Lyberg, 2003). For example, teachers may over-inflate their use of strategies, may be unable to report accurately because they must rely on their memory, or may be responding to hypothetical events in which they have no personal experience to draw from (McMillan, 2016). Empirical evidence that uses direct observation of teacher behavior in authentic settings is needed.

Last, to contribute to the field of implementation science and factors that may contribute to the long-term use of evidence-based practices in classrooms, there is a need for increased analyses and reporting on teacher-level factors that influence intervention responsiveness and implementation outcomes. Particularly, there is a need for interventions that aim to increase teacher classroom management quality. A better understanding of these two implementation components may help to close the research-to-practice gap and in turn, increase academic and behavioral outcomes for students who display challenging behavior and their peers.

Limitations

Generally, outcomes from this set of studies suggest a link between years of teaching experience and approaches to classroom management. Specifically, findings suggest significant differences in perceptions of efficacy, reported use of behavior interventions, and responses to student need. They also suggest the possible mitigating role of classroom-based interventions to produce significant outcomes despite differences in teacher-level factors. However, these results should not be interpreted without consideration of several limitations related to this review. As a scoping analysis, the set of studies synthesized do not represent a systematic collection of empirical research and are thus limited. Additionally, including only those studies conducted in the United States may discount valuable outcomes from studies whose findings may be generalizable to teachers here. While several studies reported null-findings, the potential for publication bias is likely. Finally, the review relied on only one researcher for coding of studies, which may have affected the overall reliability, specifically the possible exclusion of studies and potential misinterpretation of study characteristics.

Research Questions

Analysis of the literature on the topic of classroom management and teaching experience revealed several distinct themes, which in turn led to the development of three research questions. First, there is extensive literature to support that teachers may approach classroom management differently as they gain professional experience, suggesting that the topic of experience is valuable in a discussion of teacher development. Second, there is limited empirical evidence that uses direct observation methods regarding the classroom management and behavior support practices used by teachers across experience groups. Third, few studies in which teachers are participants in interventions focused on increasing classroom management

skills measure or report differences across experience groups, which may add purposeful knowledge to our understanding of intervention determinants and outcomes.

Guided by gaps in understanding related to teacher experience and classroom management, the aim of this study is to examine differences in observed behaviors and responsiveness of teachers within a classroom management intervention. The following research questions drove this investigation:

1. To what extent do classroom management practices differ across teacher experience groups?

2. Does teacher experience influence teacher delivery of practices within a coaching intervention?

    a. Does teaching experience influence the extensiveness/adherence of teacher delivery of practices?

    b. Does teaching experience influence the competence of delivery of practices?

3. To what extent does teaching experience influence teacher rating of usability and acceptability of coaching/intervention?

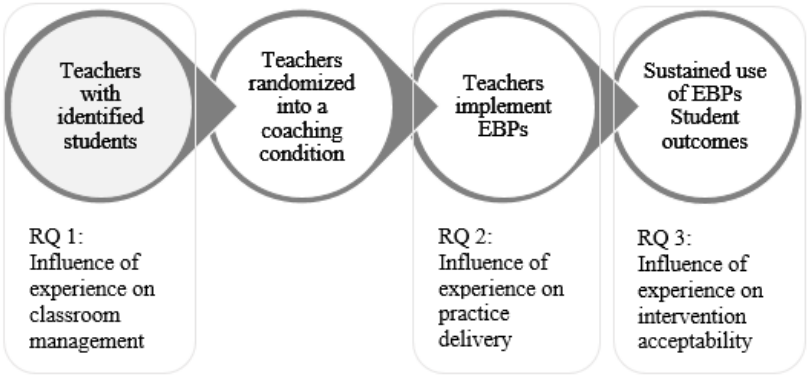
Data came from two studies involving 83 teachers who identified students displaying patterns of challenging behaviors, 45 of whom were randomized into a coaching condition. This is a meaningful data set to use for this investigation due to several characteristics of the research design. First, the study relies on the use of direct observation measurement, collected by highly-trained research staff at multiple data points throughout the intervention. Further, this direct observation measure collected data related to both adherence and competence of practice use, providing more detail about specific teacher behaviors. Second, the setting for the study is representative of both urban and rural school districts across two states. The participant group is

large, and has representation across several teacher-level factors including age, level of academic achievement, and years of teaching experience. Last, a screening tool used during the recruitment phase to determine eligibility asked teachers to rate intensity of students' externalizing behaviors. This process established automatic controls between classrooms and eliminates some bias associated with teacher reporting of student challenging behaviors. For example, during recruitment teachers identified students who have been displaying patterns of challenging behaviors in their classrooms in the first month of the school year and completed a screener. Only teachers who had at least one identified student who has met the benchmark for intensity and frequency of externalizing behaviors were included in the study.

Direct observation data collected before, during, and after the intervention was analyzed using quantitative methods to determine the relation between experience and classroom management behaviors and responsiveness to the coaching intervention. The relationship between research questions and data collection within the larger parent study is depicted in Figure 2. For example, to contribute to the need for more direct empirical evidence garnered from direct observation of teacher behavior, the first research question used data collected during pretest, before randomization occurred.

Figure 2

Mapping of Research Questions onto Parent Study



Chapter 3 will introduce the study methodology, more detailed information about participant demographics and an explanation of the grouping strategy used. The main direct observation measure will be described as well as information about training received by research staff, reliability procedures, and frequency of administration. The coaching intervention utilized in the parent study will also be described. Finally, the plan for statistical analysis will be discussed.

Chapter 3
Methodology

The purpose of this study was to examine the impact of teachers' years of experience on their use of evidence-based classroom management practices and responsiveness and acceptability to a coaching intervention. Teachers' use of evidence-based practices and responsiveness were examined using data from a direct observation measure collected during classroom instructional time. Intervention acceptability was examined using data from a teacher self-report measure. In the sections that follow, the participants, study procedures, measures, and analytic plan are discussed. The section will begin with a description of the parent study.

Parent Study: BEST in CLASS Elementary

BEST in CLASS is a Tier 2 intervention designed to increase teachers' use of key practices that decrease problem behaviors demonstrated by students in the classroom. BEST in CLASS is recognized as an evidence-based practice by What Works Clearinghouse (https://ies.ed.gov/ncee/wwc/Study/86120), the National Institute of Justice (National Institute of Justice, 2018), and the Robert Wood Johnson Foundation (Bierman et al., 2016). Originally designed as an intervention designed for preschool-age children, BEST in CLASS-Elementary was adapted for delivery in early elementary school grades (Sutherland et al., 2019) and initially tested in a small randomized control trial (Sutherland et al., 2020). BEST in CLASS – PK has demonstrated positive outcomes for teachers (Conroy et al., 2019) and children at-risk for emotional/behavioral disorders (EBD) (Sutherland et al., 2018b) and served as the foundation for the development of BEST in CLASS-E. The intervention is distinct from other Tier 2

interventions in that it is not a pullout program, rather it uses a practice-based coaching model to support teachers' use of specific practices with students identified as at risk for EBD and their peers during ongoing classroom instruction. This focused use of practices is considered a "value-added" model because teachers are increasing the intensity of targeted practices that they are likely already using in their classrooms (e.g., praise, opportunities to respond). In learning to intensify BEST in CLASS-E practices, teachers are trained and coached to increase the quality and quantity of their use of each practice with focal students and to link the practices together during authentic learning activities throughout the day.

BEST in CLASS consists of three components, the first being a full-day training workshop where teachers learn the specific BEST in CLASS practices. Second, there is a manual for teachers to use as a reference for additional information and support. Last, there are fourteen weeks of coaching to support teachers in applying the specific BEST in CLASS practices in the classroom. BEST in CLASS-E uses an adapted form of practice-based coaching (Snyder et al., 2015), an evidence-based framework to support teachers as they implement changes to their teaching methods and strengthen relationships with students. Teachers engage in two-week coaching cycles where teachers receive feedback on practices using a staggered approach. For example, after completing the initial training workshop, teachers engage in a coaching cycle which includes self-reflection, goal-setting, and guidance from their coach on a particular practice for two weeks before moving onto the next practice (see Table 1 for the coaching sequence). Additionally, a mid-point cycle is included where teachers choose a previously-learned practice to review and a culminating cycle at the end where the focus is on linking practices together to increase effectiveness. By the end of the intervention, teachers will have

received coaching across all practices (for more information about the BEST in CLASS coaching model see Sutherland et al., 2015).

Table 1

BEST in CLASS Teaching Practices and Coaching Schedule

Practice	Weeks	Definition
Home-school Partnership	Baseline-14	Strategies to help enhance positive teacher-family interactions, foster productive conversations, and provide support to both teachers and families to help proactively problem solve.
Supportive Relationships	1-3	Teacher behavior that conveys warmth, closeness and interest when listening to and interacting with the focal student.
Rules	4, 5	The design and implementation of classroom rules to effectively support students' appropriate behaviors.
Precorrection	6, 7	A proactive practice to prevent predictable challenging behaviors by reminding or instructing students to engage in a more appropriate, alternative behavior before the challenging behaviors occurs.
Opportunities to Respond	9, 10	Providing focal children with prompts, questions, and opportunities to engage in positive, appropriate behaviors during instructional activities.
Praise	11, 12	The provision of praise for specific student behaviors or responses intended to build strong positive relationships with students and reinforce appropriate behaviors.

BEST in CLASS coaches are highly trained study staff with many having previous classroom experience. Coaches engage in systematic training and development activities before and during the intervention to ensure a high degree of coaching fidelity. During the pre-intervention phase, goals of coach development include the ability to deliver coaching following the prescribed intervention framework, collection of live observation data, and use of a self-monitoring tool to calibrate the quality of their coaching practices. Also included in this phase is a checkout

process, where coaches must meet a pre-determined threshold in order to complete training. During the intervention, coaches continue to engage in activities to maintain a high level of treatment integrity including formalized integrity scoring using video artifacts of coaching sessions, coach self-monitoring and reporting, and on-going within-study professional development to provide additional performance feedback and address any barriers to implementation.

BEST in CLASS-E data is an appropriate parent study for this investigation due to several distinct factors including its use of direct observation measures at multiple time points and a group of participants across urban and rural settings with variability in teacher-level factors. Additionally, screening tools used during recruitment for study eligibility, where teachers are asked to rate the intensity of student behaviors, have established homogeneity in behavior across student participants. For example, only teachers who have at least one identified student who has met the benchmark for intensity and frequency of externalizing behaviors are part of the study. Importantly, the intervention uses coaching as the change mechanism to increase the quality of practice use with students who display patterns of challenging behaviors. These two characteristics, coaching and challenging student behavior, represent contemporary topics in education: coaching, for its recent rise in use (Carver-Thomas & Darling-Hammond, 2017) and promise as an effective PD model (Kraft et al., 2018), and student behavior, for its longstanding barrier to student success due to in part to teacher-level factors (Ficarra & Quinn, 2014; Gable et al., 2012).

Participants

Volunteers for the parent study were recruited from school districts in two eastern states, one predominantly urban and the other predominantly rural. As a multi-year randomized control

trial, the parent study was conducted at different schools in identified districts each year. As a result, the data is representative of teachers in a variety of different schools, teaching at the Kindergarten through third grade level. Table 2 provides demographic information for participants in which data were analyzed for this study. Teachers are distributed fairly evenly across treatment and control groups, 45 and 38 teachers, respectively. This makes the sample size for the first research question 83 participants, and the sample size for research questions two and three, 45 participants. There is also a mostly even distribution across grade levels, including with each grade level and across condition. While there is representation across all age groups, the majority of teachers fall into the 26-35 and 36-45 years old categories. Similarly, there is representation in all categories of race; however 65% of the sample are Caucasian and 27% are African American. The sample is heavily skewed in regards to gender, with 99% of the population female. Compared to national averages, the population is mostly aligned with trends related to categories of race and gender for public school teachers. For example, 79% of teachers in the United States identify as Caucasian, 7% African American, 9% Hispanic, and 89% female (Hussar et al., 2020).

Table 2

Participant Demographic Cross Tabulation

		Study Group		
		Control (*n*, %)	Treatment (*n*, %)	Combined (*n*)
Full Study		38, 46	45, 54	83
		Control (*n*)	Treatment (*n*)	Combined (*n*, %)
Grade				
	K	7	11	18, 22
	1st	10	15	25, 30
	2nd	9	8	17, 20
	3rd	9	9	18, 22
	Other	3	2	5, <1
Age				

18-25	3	8	11, 13
26-35	14	16	30, 36
36-45	11	9	20, 24
46-55	7	9	16, 19
Over 55	2	3	5, <1
No answer	1	0	1, <1
Race			
African American	11	11	22, 27
Asian/Pacific Islander	0	2	2, <1
Caucasian/White	24	30	54, 65
Native American	1	0	1, <1
Other/Multiple	2	2	4, <1
Gender			
Female	37	45	82, 99
Male	1	0	1, 1
	Control *(M, SD)*	Treatment *(M, SD)*	Combined (*M, SD)*
Years of Experience	8.1, 9.34	10, 9.67	9.16, 9.57

*Grade, *Other* = Special Education, mixed-grade classrooms

Procedures

Systematic training of data staff and coaches occurred annually during the pre-intervention phase. A pre-determined checkout process ensured that research staff were trained to a high level of fidelity on the measures and procedures associated with their role. For coaches, this includes the ability to collect data using the direct observation measure as well as to conduct coaching meetings following the coaching framework.

Phase 1 of the parent study had two goals: recruitment and pre-test data collection. Initial recruitment began with an information session held the week before the school year began where the opportunity to participate in the study was discussed. Approximately one month into the school year, after teachers had time to familiarize themselves with their students and put initial classroom behavior supports in place, teacher participants were identified and consented following the procedures approved by the University's Internal Review Board (IRB). After

consent, teachers then nominate up to five students who have displayed patterns of challenging behavior. Because the study aims to answer questions about the teacher's use of specific practices with certain students as well as student-level outcomes, caregiver consent and student assent is also obtained. Following this, teachers completed the externalizing behavior form of the Systematic Screening for Behavior Disorders (SSBD; Walker et al., 1990), to identify critical behavior tendencies and rate the frequency of adaptive and maladaptive behaviors. During the development year of the parent study, eligibility was determined by comparing teacher responses against specific critical events aligned with the goals of the intervention (i.e. disruption, aggression). In subsequent years, a benchmark score was determined and students who meet the benchmark SSBD score are eligible for study participation. Across all study years however, teachers with one to three consented students who met the predetermined SSBD criteria participated in the study.

The goal of phase 2 of the parent study was pretest data collection. Identical sets of data were collected for all teachers at the teacher and student level, regardless of intervention condition. Additionally, several direct observation measures were conducted during classroom instruction. Pertinent to this investigation are a researcher-created teacher demographic measure and a direct observation measure of specific practice use. Following pre-test data collection, teachers were randomized within school at the grade level to promote representation at each grade and within each school.

During the intervention phase, teachers randomized into the coaching condition received a one-day training, an intervention manual, and 14 weekly cycles of coaching. During this time, several measures were used to collect data related to teacher and student behavior. These measures were completed by both data and coach research staff and the teachers themselves and

differed between treatment and control groups. Pretest direct observation measures were repeated at midpoint for all participants by data staff who were blind to condition. Also, both teachers and coaches completed an alliance measure to rate their experience within the coaching dyad. This measure was also used as a progress-monitoring tool to identify any potential barriers to coaching. Pertinent to this investigation are the on-going direct classroom observations conducted by coaches to rate specific practice use. These direct observations are student-specific; a weekly observation was completed for each student, with the focus on the teacher-student dyad specifically. For example, a teacher who has two identified students was observed two times per week, for the duration of the 14-weeks of coaching.

During the posttest data collection phase, direct observation and student-level measures were repeated by data staff blind to condition. Additionally, several teacher and coach level measures were completed. This investigation used data collected during the pretest, intervention, and posttest phase. These measures, along with other study variables are fully described in the sections that follow.

Variables and Measures

Data for the current study were collected using several measures, at different time points and by different members of the research team. In this section, the variable and associated measure and data collection procedures are described. Additionally, Table 3 displays the data collection timeline as it relates to the coaching intervention.

Table 3

Data Collection Points for Included Measures

Measure (Completed by)	Pretest Phase	Intervention Phase	Posttest Data Collection
Teacher Demographic Survey (teacher)	C, I		
TIES (data staff)	C, I		
Coach Alliance Form (teacher, coach)			I
Coach Integrity Form (data staff, coaches)		I	
TIES (coaches)		I	
Usage Rating Profile-Intervention (teacher)			I

Note: I = Intervention Group (teachers receiving coaching); C = Control Group

Independent Variable

Years of teaching experience, the independent variable for this investigation, was obtained from a teacher demographic measure which collected information about teacher grade level assignment, age, gender, degrees held, licensure status, and years of teaching experience. This measure was completed during the pretest data collection phase. Teacher years of experience was assessed with a single item that asks, "*Not counting the current school year, how many years have you been a teacher?"* This designation is important as it denotes the number of years of experience prior to the start of the intervention. This question requires a written response, allowing teachers to report using partial years when necessary and thus, increasing specificity of the response. Before analysis, the variable was centered using a Grand Mean approach (Enders & Tofighi, 2007). This process eased the interpretation of model results by using the sample group average as the intercept.

For the purposes of this investigation, years of teaching experience was treated as a continuous variable. It is important to note that this decision contrasts the majority of literature

included in the scoping review presented in Chapter 2, which treated years of experience as a binary variable. However, a continuous variable provided several distinct advantages, particularly the opportunity to use a regression analysis rather than analysis of variance (ANOVA) and the ability to refrain from dichotomizing the variable. Lazic (2008) compared outcomes using regression analysis and ANOVA on identical data with the independent variable in both continuous and categorical structure and found the later to result in more favorable analysis, specifically the increased number of parameters needed to run an ANOVA (and thus, degrees of freedom lost), which can reduce the power needed to detect relationships, resulting in a Type II error. Particular to this investigation, the author also argues that a regression provides a more informative interpretation, which considers the numerical ordering of the data. Further, Altman and Royston (2006) present an argument against categorizing data from a continuous variable, the action that would be required to create a categorical independent variable for this study. They argue (a) that there may be reduced statistical power to detect a relation between the independent and dependent variable due to the loss of data when categorization occurs, (b) it may be challenging to estimate the variability between groups (for example, teachers who fall close to the established cut point on either side may be more similar to one another than to their assigned experience group), and (c) that using two groups may make a non-linear relationship unclear.

A final characteristic of years of experience that makes a continuous approach appropriate are the units of measurement that are maintained throughout the scale. For example, a one-unit increase between three and four years is equal to a one-unit increase between 13 and 14 years. What is unknown however, is if these one-unit increases impact the dependent variable equally; a continuous variable allows for this type of analysis. Further, the sample used for this

investigation demonstrates a great deal of variability, with representation across years of experience (*M* 9.16; *SD* 9.57, Range 0-38). Lastly, some empirical support for looking at differences using a continuous scale of experience has been found. As previously reported, Klassen and Chui (2010), in a large sample of teachers (*n* = 1,430), found a linear, positive trend across all three self-efficacy subscales (instructional strategies, student engagement, and classroom management) until approximately 25 years of teaching experience, after which a negative linear trend was found. These outcomes highlight the importance of retaining a continuous approach as the range of teaching experience may differently influence teacher behavior as they progress through their careers.

Dependent Variables

Teacher Practice Use. The Treatment Integrity Instrument for Elementary Settings Adherence and Competence Scale (TIES; Sutherland et al., 2017) is a direct observation measure used by trained research staff to report on teacher use of specific practices used with identified focal students during a 15-20 minute classroom observation. The measure uses a Likert scale where observers report on two dimensions of practice use, adherence and competence. The report of adherence, or the extensiveness of practice use is done on a Likert-type scale ranging from 1 (none) to 7 (extensively). Observers also report on competence, or the skillfulness of practice use including responsiveness to student need, using a Likert-type scale ranging from 1 (very poor) to 7 (excellent; the measure also includes student responsiveness items, which were not used for this investigation). In a publication of preliminary effects of the BEST in CLASS-E Intervention (Sutherland et al., 2020), the TIES measure was shown to be reliable, with mean intraclass correlations (ICCs) of double-coded observations reported for the adherence scale of

.82 (with all items reflecting “good” to “excellent” agreement) and .61 reported for the competence scale (with items representing “fair” to “excellent” agreement).

While different variations of the TIES are used by data staff and coaches, both report the use of the predetermined practices that the coaching intervention is based on. The specific practices that were central to this investigation are supportive relationships, rules, precorrection, opportunities to respond, and praise (see Table 1 for a description of each practice). These practices represent all of the core modules of the intervention with the exception of linking and mastery, which is an opportunity for teachers to receive feedback during the last two weeks of the intervention as they use several practices together to increase effectiveness of practice use. Data for linking and mastery was only collected anecdotally by coaches and is therefore unable to be investigated in the same way as the other practices. Similarly, home-school partnership is a set of efforts to increase collaboration between the teacher and the caregiver. Data for home-school partnership is also collected anecdotally by coaches and cannot be investigated in the same way as the other practices.

Observational data was collected using the TIES at three different points in the parent study and by different research staff to serve different purposes. These distinctions are important for this particular study. Highly trained research staff used the TIES to collect pre-test data before teachers were randomized and before the intervention began. Data from these observations was used to answer RQ 1. These observations were double-coded using two observers who met reliability standards during training and maintained a high inter-observer reliability throughout the intervention. Mean intra class correlation (ICC) values for adherence (M = 0.73, SD = 0.06) and competence (M= 0.52, SD = 0.20) across practices demonstrated moderate alignment between observers (Koo & Li, 2016).

Highly trained coaches collected data using the TIES up to 14 times with each teacher-student dyad. For example, a teacher with one identified student had between 12 and 14 time points for each of the practices and each practice was given an adherence and competence score. These observations were single-coded by the coach, who met reliability standards during training; however, coach reliability of the TIES was not maintained after the start of the intervention. Data from these observations was used to answer RQ 2.

Intervention Acceptability. The Usage Rating Profile- Intervention (URP-IR; Chafouleas et al., 2011) is a teacher self-report measure designed to collect data related to a teacher's perception of intervention usability and acceptability across the six dimensions of acceptability, understanding, home school collaboration, feasibility, system climate, and system support and shown to be reliable (Briesch et al., 2013). At post-test, participants rated the degree to which they align with intervention-related statements using a Likert-type scale with 1 (strongly disagree) to 6 (strongly agree). Data from the URP-IR subscales related to individual-level factors (acceptability, understanding, and feasibility) were used to answer RQ 3, with subscale scores providing information about possible implementation barriers.

Coaching Integrity

Two measures were used throughout the study to ensure that teacher participants experienced high quality, standardized coaching throughout the intervention. The BEST in CLASS Coaching Integrity Form is a Likert-style measure with two identical forms, *Coach Report* and *Observer Report*, which describes the degree to which specific behaviors and practices are used by coaches during coaching meetings. The Coach form was completed by coaches after each coaching session as a self-reflection and self-monitoring tool. The Observer Report was used by trained

data staff to contribute to the report of overall intervention integrity and as a progress-monitoring tool to provide feedback to coaches during the intervention. Coaching integrity was obtained at four points during the intervention for each teacher-coach dyad using double coding of video-recorded coaching sessions. For this investigation, overall coaching integrity data from the *Observer Report* is presented in the Results section to describe the degree to which teachers had access to the coaching intervention as designed. Based on previous analysis and due to the nature of the on-going accountability measures of the study, the integrity scores were high with limited variability across teacher-coach dyads. As a result, this information will be presented as data that describes the high quality, systematic coaching that study participants received rather than a variable of interest.

An essential component of effective coaching is the relationship between the teacher and the coach. To describe the relationship within each teacher-coach dyad, the Practice-based Coaching Working Alliance Inventory- Coach and Teacher Forms were completed at two points throughout the study (mid-point and post-test). Both forms use a seven-point Likert scale in which the teacher and coach report on the frequency with which certain behaviors characterize the work done within the dyad between 1 (almost never) and 7 (almost always). The teacher form aggregates data into the subscales of rapport and client focus. For example, a teacher would rate the statement, "*My coach encourages me to talk about my work with the focal students in ways that are comfortable for me*." The coach form uses the subscales of child focus, rapport, and identification with similar statements, for example *"I encourage my teacher to talk about the work in ways that are comfortable for him/her.*" For this investigation, post-test alliance scores were used to report on the strength of the teacher-coach relationship throughout the intervention. Throughout the previous years of the project, teacher and coach alliance scores have been scored

high, with low variability in the data. Therefore, this variable is used as a descriptor of the high-quality relationships between the teacher and the coach and as a second indicator of coaching integrity.

Analytic Plan and Hypotheses

The following section outlines the analytic plan for each research question and associated sub-questions. Included in this discussion are the methods that were used during data exploration and the modeling approaches for each research question. Before data analysis began, several steps were taken to check for data accuracy. First, although the data have been doubled-entered by research staff, it was necessary to check for out of range values and identify and remove outliers. Outliers were identified using Cook's distance (*D*; Cook, 1977), a method particular to regression analysis, which identifies the influence of a data point on the outcome variable. Any value of *D* greater than 4/*n* (where n equals number of observations) was identified as an outlier and removed (Field, 2018). Further, a cross-check of data inclusion criteria was necessary. For example, this investigation followed the parameters established by the parent study to only include students who were present for 80% of observations and teachers who were present for 80% of coaching cycles. Therefore, the sample used for RQ 2 and 3 included only teachers who participated in at least 10 coaching cycles and students who had at least 10 data collection time points.

To increase the precision in which scores of competence for each practice can be analyzed and interpreted, adherence scores of one (indicating that the practice was not observed and therefore not assigned a value) were coded as missing values for competence, since competence could not be scored for a practice that was not observed. This procedure was followed for the regression models reported in research questions one and two. It was also

necessary to explore the distribution of the independent variables to determine normality. This included the use of visuals to provide an initial understanding of the variables and use of the Shapiro-Wilk test of normality (Royston, 1992). Lastly, teacher years of experience was grand mean centered to allow the interpretation of outcomes against the sample average.

***Research Question 1*: To what extent does years of teaching experience predict teacher use of classroom management practices?**

Data for this question came from the TIES direct observation measure, collected by data staff during the pretest data collection phase for teachers, before randomization into treatment groups. Two linear regression models for each teaching practice were run, one for adherence and one for competence (10 models, total). Each model used the following model building strategy. First, a visual representation explored the impact of the independent variable on the dependent variable and provided support for interpretations. Next, a null model with no predictors was run to obtain an intraclass correlation coefficient (ICC) which described the amount of variance in each practice accounted for at level one (student level) and two (teacher level). ICCs larger than .05 were produced across all models (in both adherence and competence) for all five practices, justifying the need to account for the grouping of student-level data within teachers. As a result, all models were nested across all practices. To account for nesting of student data within teachers, linear regression with cluster (teacher) robust standard error estimation was used (Huber, 1967; White, 1980). Using this specification accounts for the nonindependence of students by adjusting the standard error estimates and, therefore, removes variance due to the repetition of children across teachers (Asparouhov & Muthen, 2006). Without accounting for the nonindependence in the data, the estimated standard errors would be inflated resulting in a greater chance of committing a Type I error. The null models also provide Akain Information

Criterion (AIC) and Schwarz's Bayesian Information Criterion (BIC), which were used as a baseline to measure model fit as predictors were added. Next, the teacher-level predictor of years of experience was added to the model, clustered at the teacher level, and AIC and BIC monitored in comparison to the null model. For example, models with lower AIC and BIC scores explain the largest amount of variation of the dependent variable using the fewest possible independent variables, suggesting a better model fit.

Research Question 1 Hypotheses

It was hypothesized that years of teaching experience will impact adherence and competence differently for three of the five practices. For pre-correction, rules, and opportunities to respond which require more forethought, planning, and require other classroom systems to be in place, a positive association was expected between experience and both adherence and competence. This suggests that teachers with more experience are using these practices more frequently and with a higher level of skill and efficiency. For the practices of supportive relationships and praise, it was hypothesized that experience will be associated with increased competence of these practices but not adherence. This suggests that more experienced teachers were using the practices less extensively, but with a higher level of competence when they are used. Conversely, it was predicted that less experienced teachers are using these practices more frequently (adherence) but with a lower level of skill (competence). Both hypotheses were drawn from the supporting theories introduced in Chapter 2, which suggests that as teachers develop in their careers they are better able to embed classroom management practices into their established routines and systems, and teach using a fluid style. This may include using practices that are not being measured such as proximity, where a teacher uses purposeful positioning or non-verbal cues to quell behavior disruptions.

Research Question 2: ***To what extent does teacher experience influence teacher delivery of practices within a coaching intervention?***

a. Does teacher experience influence the extensiveness/adherence of teacher delivery of practices?

b. Does teacher experience influence the competence of delivery of practices?

Data for this analysis came from the TIES direct observation measure, collected by coaching staff during the intervention with teachers in the treatment group only. Two multiple regression models for each teaching practice were run, one for adherence and one for competence (10 models total). The relation between experience and adherence and competency scores collected during the final week of coaching, after controlling for baseline score (collected during week 1 of coaching, before participating in coaching cycles) were examined.

A similar model building strategy was used here however, this model included an additional predictor, teachers' baseline score on adherence or competence. First, a visual representation was created to explore the impact of the independent variable on the dependent variable and provide support for interpretations. Next, a null model with no predictors was run to obtain an intraclass correlation coefficient (ICC) which described the amount of variance at level one and two. Similar to the previous research question, ICCs larger than .05 were produced for both adherence and competence across all practices with the exception of one instance, therefor a nested approach was used for consistency. To account for nesting of student data within teachers, linear regression with cluster (teacher) robust standard error estimation was used (Huber, 1967; White, 1980). Null models also provided Akain Information Criterion (AIC) and Schwarz's Bayesian Information Criterion (BIC), which was used as a baseline to measure model fit as predictors

were added. Next, the predictors (years of experience and baseline scores) were added to the model, clustered at the teacher level, and AIC and BIC monitored as an indicator of model fit.

Research Question 2 Hypotheses

Similar to the previous hypothesis, it was expected that novice teachers will have higher adherence scores for each practice at the conclusion of the intervention; however, veteran teachers are hypothesized to have higher competence scores. As identified by Berliner's model, newer teachers may be more impressionable by their environment, making them more likely to take advice from a coach assuming an "expert" role. Conversely, teachers with more experience, who have already begun to develop their pedagogy may struggle to implement new practices into their established routines.

Research Question 3: To what extent does teaching experience influence teacher rating of usability and acceptability of a coaching intervention?

Data for this analysis came from the URP-IR measure, self-reported by teachers in the treatment group during the post-test data collection phase. A regression model tested the impact of years of experience on the subscale scores of acceptability, understanding, and feasibility. Since these subscales were likely to be related, a single structural model was used that included each of these subscales as a dependent variable (and accounted for covariance between these subscales).

The following model building strategy was used. First, visuals were used to explore the impact of the independent variable on the dependent variables. Next, to specify the structural model, teaching experience was regressed on the subscales of acceptability, understanding, and feasibility in a single model. Covariances were estimated between the three dependent variables. The r-squared values were examined to assess the percent of variance accounted for by teaching

experience in each of the dependent variables. Several statistical tests were employed to evaluate overall fit of the model to the data including the chi square test, root means square error of approximation (RMSEA), comparative fit index (CFI) and Tucker-Lewis index, and the Standardized Root Mean Square Residual (SRMR).

Research Question 3 Hypothesis

Due to the personalized, teacher-driven nature of the coaching intervention and the empirical research supporting it as an effective professional development tool, it was hypothesized that all participants will score the intervention with a high level of acceptability. However, it is expected that differences may be seen in feasibility scores, with less experienced teachers scoring higher than veteran teachers, who may struggle to incorporate new approaches into their teaching style long-term that may already be fluid and automatic.

Regression Diagnostics

To determine the degree to which findings from the investigation of these data can be generalized to larger populations, the following assumptions pertinent to regression models with a single predictor were identified and tested (Field, 2018). First, the presence of unusual and potentially influential data was examined using visual inspection of a box and whisker plot for the independent variables, years of teaching experience. Once identified, regressions were run with and without these data points to determine their influence on the data. To test the assumptions of linearity (the linear relationship between the independent and dependent variables) and homoscedasticity (the comparison of the variance across all values of the independent variable), residual versus fitted plots were examined across all models in all three research questions. Visual and statistical analysis of skewness and kurtosis were used to determine if the assumptions had been violated. Robust standard error estimation corrected for

the violations of these assumptions.

Chapter 4
Results

Chapter 4 will present results, organized by research question. First, descriptive statistics related to the variables as well as other demographic information for each participant group are presented. For RQ 2 and 3, coaching integrity and alliance data will also be presented.

Descriptive Statistics

RQ1 Participant Group

The data set used to answer this research question included 83 teachers who were mostly evenly distributed across grades Kindergarten through third grade (see Table 2). The independent variable (years of teaching experience) for this participant group had a range of 38 with the minimum value at zero years of previous experience and a max at 38 years. Teachers in this group had a mean of 9.16 (SD = 9.64) years of experience; although visual analysis suggested that the variable was skewed towards teachers with less experience. This unequal distribution was confirmed with a significant test of skewness *(p* < .001) and a kurtosis value approaching significance (p = .09). The Shapiro-wilk test of Normality (Royston, 1992), which considers both skewness and kurtosis was significant, indicating a non-normal distribution.

The dependent variables (adherence and competence scores from each practice) are displayed in Table 4. Comparison of mean scores of practice use for all participants suggest that certain practices were used more extensively and with a higher level of skill than others. For example, the mean adherence and competence score for the evidence-based practice of opportunities to respond (4.78 and 4.29, respectively) is much higher than the adherence and competence scores for the practice of rules (1.21 and 0.44, respectively). Most standard

deviations are around 1.0, suggesting a low dispersion of scores and low variability. There were no systematic trends within or across variables related to dispersion.

RQ 2 and 3 Participant Group

The data set used to answer this research question included 45 teachers who participated in the coaching intervention, with a mostly even distribution across grades Kindergarten through third grade (see Table 2). Post-test coach alliance scores reported by the teacher suggest strong collaboration for both subscales of client focus and rapport (M = 6.85, SD = 0.41 or 98% and M = 6.82, SD = 0.41 or 97% respectively). Similar scores were reported by coaches for both subscales, M = 6.37, SD = 0.57 or 91% for client focus and M = 6.38, SD = 0.33 or 91% for rapport. These data suggest that teacher-coach dyads had strong working relationships with one another and the reporters were aligned in their efforts to meet intervention aims. Mean scores from the coaching integrity coach report measures revealed high percentages of implementation throughout the intervention, with 83-99% fidelity reported across all coaching elements. This suggests that teachers in the intervention group received the intervention as designed.

Teachers in the sample had on average 10 years of teaching experience (SD = 9.67, range 0-27). Visual analysis suggested that the variable was skewed towards teachers with less experience. The unequal distribution was confirmed with a significant test of skewness ($p < .01$) and a non-significant kurtosis value ($p = .07$). The Shapiro-wilk test of Normality (Royston, 1992), which considers both skewness and kurtosis was significant ($p < .001$), indicating a non-normal distribution.

Teachers' Use of Classroom Management Practices

Research Question 1 asked *To what extent does years of teaching experience predict teacher use of classroom management practices?* Data for this question came from the TIES direct observation measure, collected by data staff during the pretest data collection phase for teachers, before randomization into treatment groups. Two linear regression models for each teaching practice were run, one for adherence and one for competence (10 models, total).

RQ 1 Results

The intraclass correlation coefficients (ICC) for both adherence and competence for each practice met the pre-determined benchmark of .05 (Raudenbush & Bryk, 2002), providing justification of the need to account for grouping of students within teachers (see Table 4). Results of the regression models predicting each practice suggested a non-significant relation between years of teaching experience and practice use, after adjusting for students within teachers (see Table 4). These findings deny the hypothesis that years of teaching experience will have an impact on practice scores, albeit differently across practice delivery.

Table 4
Use of Classroom Management Practices

Practice	Practice Delivery	*n*	*M; SD*	ICC	*t*	*p*
Opportunities to Respond	Adherence	160	4.74; 0.99	.41	-0.28	0.78
	Competence	159	4.30; 0.78	.50	-0.67	0.51
Praise	Adherence	160	3.03; 1.38	.44	0.88	0.38
	Competence	131	3.74; 0.84	.32	-0.48	0.63
Supportive Relationships	Adherence	160	2.90; 1.19	.24	-1.17	0.37
	Competence	141	3.50; 1.10	.48	-0.40	0.69
Precorrection	Adherence	160	1.63; 0.88	.07	0.03	0.76
	Competence	63	3.42; 0.77	.65	-0.42	0.63
Rules	Adherence	160	1.21; 0.60	.20	0.03	0.98
	Competence	23	3.09; 1.00	.64	-1.90	0.07

Note: Adherence is scored on a 1-7 point Likert-type scale; Competence is scored on a 0-7 point Likert-type scale, with zero values removed.

Coached Teachers' Practice Delivery

Research Question 2 asked *To what extent does teacher experience influence teacher delivery of practices (i.e., adherence and competence) within a coaching intervention?* Data for this analysis came from the TIES direct observation measure, collected by coaches during the intervention with teachers in the treatment group. Two multiple regression models for each teaching practice were run, one for adherence and one for competence (10 models total). The relation between experience and adherence and competence scores collected during the final week of coaching, after controlling for baseline score (collected during week 1 of coaching, before participating in coaching cycles) were examined.

RQ 2 Results

Similar to RQ 1, the intraclass correlation coefficients (ICC) for both adherence and competence for each practice on the coach-reported measure were examined and met the pre-determined benchmark of .05, providing justification of the need to account for grouping of student-level data within teachers (see Table 5). Before running the regressions, paired samples t-tests were used to compare differences in practice use pre- and post-intervention to determine if there was significant change in teacher behavior. Table 6 shows significant differences in mean scores for both adherence and competence of all practices. Due to the nature of the Likert scale used to collect data, where a score of one for adherence indicates that the practice "did not occur" and the conversion of these corresponding values to missing for competence, increases in *n* values for adherence indicate the number of instances in which the practice was observed. For example, the practice rules was used to some degree during nine observations at pretest and 43 observations at post-test. Coupled with the significant results from the t-tests, this suggests that teachers on average increased both the quantity and quality of practice use, and some began using practices that they may have been using before.

Results of the regression models for each practice suggest a non-significant relation between years of teaching experience and posttest practice use, after controlling for baseline scores and adjusting for students within teachers (see Table 5). Thus, while data suggests significant changes in teacher behavior post-intervention, years of teaching experience was not a significant predictor of this change. These findings deny the hypothesis that years of teaching experience will be a significant, positive predictor of competence scores across all practices and a negative predictor of adherence scores.

Table 5
Practice Delivery, Coached Teachers

Practice	**Practice Dimensions**	***n***	***M***	**ICC**	***t***	***p***
Opportunities to Respond	Adherence	70	5.50	.27	-0.51	0.61
	Competence	70	5.50	.27	0.64	0.52
Praise	Adherence	70	5.11	.36	-1.36	0.18
	Competence	59	4.09	.71	0.70	0.49
Supportive Relationships	Adherence	70	3.60	.52	0.71	0.48
	Competence	55	3.75	.71	0.29	0.79
Precorrection	Adherence	70	2.90	<.00	-1.88	0.07
	Competence	34	3.76	.81	-0.62	0.54
Rules	Adherence	70	2.57	.12	-0.75	0.46
	Competence	9	3.75	.93	-1.15	0.30

Note: Adherence is scored on a 1-7 point Likert-type scale; Competence is scored on a 0-7 point Likert-type scale, with zero values removed.

Table 6
Practice Delivery, Coached Teachers Pre-and Post-intervention

Practice	**Practice Dimensions**	**Pre-test**		**Post-test**		***t***	***p***
		n	***M, SD***	***n***	***M, SD***		
Opportunities to Respond	Adherence	70	4.93, 0.98	70	5.49, 0.09	4.14	<0.01
	Competence	70	4.59, 1.10	70	5.30, 0.77	4.96	<0.01
Praise	Adherence	70	3.21, 1.50	70	5.13, 1.16	9.99	<0.01
	Competence	59	4.02, 1.10	70	5.10, 1.05	7.30	<0.01
Supportive Relationships	Adherence	70	2.89, 1.52	70	3.59, 1.38	3.30	<0.01

	Competence	55	3.92, 1.18	66	4.35, 1.10	3.64	<0.01
Precorrection	Adherence	70	2.03, 1.29	70	2.90, 1.44	4.22	<0.01
	Competence	34	3.79, 1.14	53	4.35, 1.02	4.82	<0.01
Rules	Adherence	70	1.29, 0.10	70	2.54, 1.53	6.96	<0.01
	Competence	9	3.77, 1.09	43	4.11, 0.98	7.58	<0.01

Note: Adherence is scored on a 1-7 point Likert-type scale; Competence is scored on a 0-7 point Likert-type scale, with zero values removed for calculating means but included for t-tests.

Coached Teachers' Perception of the Intervention

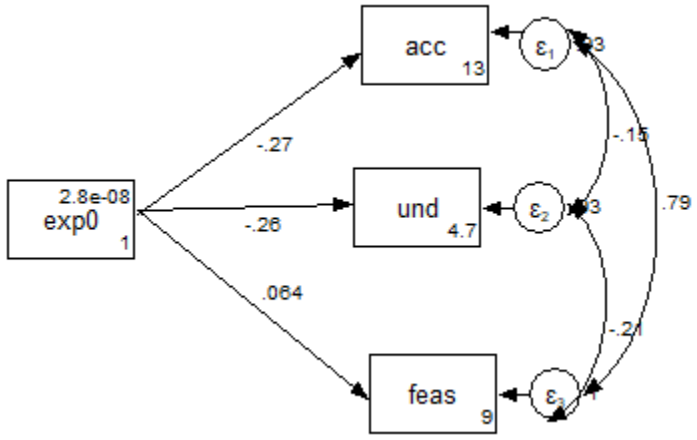
Research Question 3 asked *To what extent does teaching experience influence teacher rating of usability and acceptability of a coaching intervention?* Data for this analysis came from the URP-IR measure (Chafouleas et al., 2011), self-reported by teachers in the treatment group during the post-test data collection phase. A regression model tested the impact of years of experience on the subscale scores of acceptability, understanding, and feasibility.

RQ 3 Results

Combined scores of the URP subscales of interest were high, suggesting a high level of acceptability for all participants (M = 95.64, SD = 7.21; 108 points possible). Analysis of a scatter plot of the correlation between the independent variable (years of experience) and the dependent variables (acceptability, understanding, and feasibility) revealed no distinct patterns, however some clustering of scores was seen in quadrant II, suggesting that less experienced teachers tended to rate these components high. Covariance paths included in the structural model (see figure 3), revealed a strong relation between the constructs of feasibility and acceptability (B = .79, $p < .01$), providing justification for use of a single model. Further, experience was significantly and negatively associated with teacher reports of acceptability (B = -.27, p = .046). The predictive ability of experience on understanding was approaching significance (-.25, p = .058) but the relation between years of experience and feasibility was not significant (.06, p =

.668). These results confirm the hypothesis that participants would rate an intervention that is teacher-driven and personalized with a high level of overall acceptability however, they deny the hypothesis that years of experience would be a negative predictor of feasibility scores.

Figure 3
Path Model for RQ 3: URP Subscales



Several test statistics were used to evaluate the overall fit of the model to the data. First, the chi square goodness of fit test was significant ($p < .01$) suggesting poor model fit (Snedecor & Cochran, 1989). Second, a root means square error of approximation (RMSEA) value of 0.0 indicated close model fit (Steiger, 1990). Next, both the comparative fit index (CFI) and Tucker-Lewis index were 1.0, suggesting very good model fit (Bentler, 1990; Tucker & lewis, 1973). Lastly, the Standardized Root Mean Square Residual (SRMR) value was 0.0, suggesting a close fitting model (Hu & Bentler, 1999). With the exception of the Chi-square test which may be influenced by sample size, all indices suggest a strong model fit.

Regression Diagnostics

To determine the degree to which findings from the investigation of these data can be generalized to larger populations, the following assumptions pertinent to regression models with

a single predictor have been identified and tested (Field, 2018). First, the presence of unusual and potentially influential data was examined using visual inspection of a box and whisker plot for the independent variable, years of teaching experience. In the participant set for research question 1, three values were determined to be outliers (35, 35, and 38 years of experience) as they fell outside of 1.5 times the maximum interquartile range. Regressions were run with and without these three data points with no changes to the results; therefore, they did not appear to be influential and were included. No outliers were identified for the variable of years of teaching experience in the participant set for research questions 2 and 3.

To test the assumptions of linearity (the linear relationship between the independent and dependent variables) and homoscedasticity (the comparison of the variance across all values of the independent variable), residual versus fitted plots were used across all models in all three research questions. Visual analysis of these plots revealed no clear pattern of funneling or curving, meeting the assumptions of homoscedasticity and linearity; however statistical analysis of skewness and kurtosis suggests some violation of the linearity assumption. The use of robust standard error estimation corrected for the violations of these assumptions.

Chapter 5
Discussion

The purpose of this study was to examine the impact of teachers' years of experience on their use of evidence-based classroom management practices as well as their responsiveness and acceptability to a coaching intervention. Examination of data from a direct observation measure collected during classroom instructional time with teachers who have students identified as showing patterns of challenging behaviors revealed a non-significant relation between years of teaching experience and practice use, after adjusting for students within teachers. Examination of data collected from direct observations of teachers before and after participating in a coaching intervention revealed a non-significant relation between years of teaching experience and practice use, after controlling for baseline scores and adjusting for students within teachers. Years of teaching experience was found to be a significant, negative predictor of intervention acceptability subscale scores for coached teachers with understanding subscale scores approaching significance. The sections that follow will explore the implications for these findings and the limitations of the study.

Teacher Practice Use (RQ1)

The aim of Research Question one was to add empirical evidence to the literature gathered from direct observation of teacher behavior. Contrary to a large set of studies using self-report that suggests teachers may approach classroom management differently as they gain professional experience (e.g., Feuerborn & Chinn, 2012, Noltemeyer et al., 2012), results from this study suggest that teaching experience was not a significant predictor of practice use. These findings are also contrary to the theoretical model presented, which suggested that as teachers

gain experience, they also develop cognitive understanding related to classroom functions, interactions, and the ways in which their behaviors impact student outcomes (Berliner, 1988). Further, they develop the flexibility to employ strategies in more appropriate ways that take into consideration timing, the needs of the particular student, and organizational factors.

This dissonance may be the result of several important factors situated in both methodology and the nature of the intervention in the parent study. First, data for this study was collected using an interval scale that allowed years of experience to be treated as a continuous variable. This is in contrast to the categorical scales used in the majority of previous investigations (e.g., Clarridge, 1990; PressleyF et al., 2020). While grouping participants can be a valuable approach to identify patterns and trends, it is widely understood that continuous variables yield stronger statistical analysis (Altman & Royston, 2006; Laziac, 2008). The categorization of the experience variable in other studies may have increased the likelihood of a type I error, indicating a difference in groups that may be heavily influenced by scores around the mean. Second, the nature in which data was collected for this study (i.e., direct observation) is an important distinction. The scoping review presented in Chapter 2 highlighted a need for empirical evidence collected using direct observation measures to contend with possible threats related to the validity of data collected via self-report methods (Biemer & Lyberg, 2003; McMiIllan, 2016). These methodological differences may help account for this dissonance and provide further justification to respond to this research gap.

Even though experience was not found to be a significant predictor of practice use, findings revealed pertinent information regarding teachers' use of evidence-based practices when responding to patterns of challenging student behavior. First, high intraclass correlation scores (ICCs) for the nested models suggest significant variability in practice use between teachers. By

extension, this also suggests that teachers are using practices across their students in a somewhat similar manner. This uniqueness between individuals suggests that the consideration of factors at the teacher-level is a valuable topic of study. In order words, while teacher experience did not explain this difference, some other variable likely does. This line of questioning illuminates the implementation science framework presented in Chapter 1 and the identification of teacher-level factors as intervention determinants by Han and Weiss (2005).

Second, certain practices were observed to be used at a higher rate, or with a greater level of extensiveness, than others. For example, mean adherence scores for the practice opportunities to respond were higher than mean scores for the practice rules. Because these data were collected in classrooms with students who are known to display patterns of challenging behaviors, this insight helps identify practices used regularly when responding to challenging behavior versus ones they are not. Relatedly, differences in mean competence scores across practices provide insight into the level of skill in which the teacher uses the practice, including the timing of practice use and appropriateness given the student and situation. Differences in these scores suggests teachers may have less awareness of certain evidence-based practices or certain practices may be harder to implement with a high level of quality as teachers respond to ongoing challenging behaviors.

These findings may inform the way that we support teachers to increase use of evidence-based practices. Additional resources or time may be needed to support teachers' use of practices that are being implemented less extensively or with a lower level of quality (i.e. precorrection and rules). This could mean front-loading practices earlier in the intervention or diversifying the models or type of feedback that teachers receive as they integrate these particular skills into their instructional methods. Further, high ICCs illuminate the unique behaviors between participants,

providing evidence for professional development activities that are teacher-driven and are responsive to the needs of the individual.

These findings may also inform next steps for research. While this study adds to the existing literature, more investigations using direct observation of teacher behavior are needed. Given the differences discovered between observed teachers, it would be appropriate to utilize data collection and analysis methods that continue to allow a nested approach to account for this variability. Further, the nature of the theoretical model suggests changes in teacher skills over time; thus, it may be purposeful to use methods that allow for exploration of data that is non-linear, or a growth model. This design would account for participant maturity, and may speak more clearly to use of practices as their relationships with students develop. Lastly, findings from this study support the continued investigation of teacher-level factors, or determinants, that may explain differences in practice use. Some of this work has already be done within the BEST in CLASS project; for example researchers have examined elementary teachers' attributions of student behavior (McCullough et al. 2022), early childhood teachers' level of teacher education (Sutherland et al., 2018), and an examination of the intersection of self-efficacy and race/ethnicity match between teachers and students (Kunemund et al, 2020). Next steps for research could include behavioral content knowledge, previous experience working with students with patterns of challenging behaviors, and quality of Tier 1 (i.e. universal) classroom supports and structures.

Coached Teacher Behavior (RQ 2)

Despite a great deal of literature that supports years of experience as a predictor of teacher perceptions, self-reported practice use, and responses to hypothetical scenarios, teacher experience was not a significant predictor in changes in teacher behavior after participation in a

coaching intervention. Given that the mean adherence and competence scores across all practices increased significantly as a result of the intervention, these results provide evidence to support what was proposed in Chapter 2 regarding the ability for teacher-driven interventions (i.e., coaching) to mitigate these differences. The teacher-focused nature of coaching as a professional development model may be tailored in such a way that skill development is likely to occur despite previous teaching experience. This is an essential understanding as we seek to increase teacher effectiveness, particularly in classroom management and behavior support.

These results are aligned with the small set of findings from intervention studies using direct observation measures that report null findings of teacher experience as a predictor of teacher behavior (e.g., Gregory et al., 2014, Simonsen et al., 2020). Distinct factors related to the intervention and methodology of the parent study extend this particular literature base in valuable ways. The nature of BEST in CLASS, designed as a Tier 2, teacher-delivered intervention, required teachers to identify students showing patterns of challenging behavior. Screeners were used to identify students that had met established benchmarks related to the intensity and frequency of externalizing behaviors. This is in contrast to other studies that did not have this inclusion criteria and whose interventions were designed as Tier 1, universal approaches. For example, in the coaching intervention *My Teaching Partner- Secondary* (Gregory et al., 2014) aimed at increasing student behavioral engagement, teachers identified a 'focal class' in which coaching efforts and data collection were based. Similarly, in a PD model with training plus on-going self-monitoring activities for elementary teachers, data collection was focused on the teachers' use of specific practices during a recorded lesson (Simonsen et al., 2020). In both studies, baseline data collection provided descriptive features of teacher behavior pre-intervention, however it was not used as inclusion criteria for participation. The essential

distinction here is a resulting participant group of teachers used for this study who all had students who displayed ongoing behaviors that meet a similar threshold for frequency and intensity. These unique factors contribute to the potential for coaching to mitigate teacher and student difference, adding to the growing literature base positioning coaching as an effective professional development method.

The results also revealed important distinctions as it relates to the adherence and competence scores of each practice compared at pre- and post-test. Adherence and competence are two of several dimensions of treatment integrity that influence student outcomes and are essential to understanding the overall effectiveness of an intervention (Sutherland et al., 2021). The TIES coding manual describes adherence and competence as two unique dimensions. Both are scored using a Likert-type scale; however the scale values and scoring methods are unique, making score comparison inappropriate. It is worth noting however, that there were no instances where mean adherence scores increased for a practice where competence did not, suggesting that the development of these two aspects of the skill may happen simultaneously. For example, as teachers develop how to employ the practice of precorrection throughout their teaching (adherence), they may also have the opportunity to develop their ability to use the practice skillfully in a way that is responsive to student need (competence). This association is a direct parallel to the theoretical model, which emphasizes the role of lived experiences on teachers' ability to be responsive to student need.

Similar to the previous results, these results have direct implications for research and practice. Null findings from this investigation support the positive associations discovered in the coaching literature. To illustrate, a meta-analysis of teacher coaching program outcomes reported positive effects on student achievement and large, positive effects on instructional quality (Kraft

et al., 2018). The authors propose that these effects are due to the specific characteristics of coaching that separate it from less effective professional development models, mainly that it is job-embedded, of intense and sustained duration, based on a discrete skill set, and focuses on active learning. What has yet to be done, however, is a review of literature that examines coaching as an impetus for the development of classroom management and behavior supports, exclusively. Essential understanding here would also include examination of specific coaching components, for example the use of goal setting, self-reflection, feedback, and coaching dosage.

Within coaching interventions and other professional development models, it would also be prudent to examine student-level factors such as intensity and duration of problematic behaviors prior to the intervention and the potential mediating role of student responsiveness. Applied to classroom settings, Sameroff's (1995) transactional model describes how continuous interactions between key players contribute to changes in both teacher and student behavior over time (Sutherland & Oswald, 2005). Here, the impact of ongoing, problematic behaviors on both the student and the teacher are acknowledged. For example, the relationship between a teacher and a student who continues to display challenging behaviors despite the teacher's efforts will likely be one of frustration, mistrust, or resentment. Similarly, students themselves may develop hostile feelings toward their teachers, which may be either the result or the cause of the behavior. It would be purposeful to examine the intensity and duration of the behavior prior to the intervention to better understand how this factor may impact intervention outcomes, particularly in cases where the duration of student behavior has persisted for multiple school years.

Related to teacher-student relationships, Sutherland et al. (2020) position student responsiveness as a dimension of treatment fidelity which could lead to essential understanding within an implementation science framework. Defined in the TIES manual as "the extent to

which the student's behavior indicates responsiveness to the teacher's attempts to engage the student" (p.37), a high responsiveness score indicates a strong relationship between the teacher's actions (including the use of practices) and the student's participation; effectively, demonstrating that what the teacher is doing is working to minimize problematic behaviors and increase engagement in instruction. Student responsiveness becomes particularly important when we consider the theoretical framework which is based on teachers' lived experiences. For example, if a teacher employs a practice and it is perceived as effective, (i.e., observable increases in student responsiveness) the teacher is likely to continue to employ this practice. Conversely, if it is perceived as *ineffective*, they are likely to discontinue its use. This is supported in the self-efficacy literature and often described as *success breeds success* (See Bandura, 1993). Therefore, the teacher's perception of the effectiveness of the practice may be a determining factor in their future behavior.

Lastly, investigations related to change in teacher behavior over time must include a comparison group. While experience was a null finding in this study, we should not exclude it from future investigations. Based on the theoretical framework it should, however be paired with a comparison group. This design choice would minimize the threat of maturity, or change in participant behavior due to lived experiences- the crux of the theoretical underpinning.

Perceptions of the Intervention (RQ 3)

In considering the results from RQ 3 related to perceptions of acceptability of the coaching intervention, it is important to remember that mean scores in both adherence and competence across all practices increased significantly as a result of the intervention. While teachers with various amounts of experience successfully integrated practices into their classroom routines and combined scores of the URP subscales of interest were high, more

experienced teachers perceived the intervention to be less acceptable (i.e., less effective, less commitment to using, harder to integrate with current practices). This is surprising given their outcomes, however pertinent understanding may be gleaned when we consider Berliner's (1988) teacher development model which suggests that over time the practice of teaching becomes more intuitive and fluid, requiring less effort. The change in behavior that the intervention promotes contrasts this, and may be perceived as a barrier for experienced teachers who have reached this fluid state. Given the mean years of experience for the participant group was 8.22, it can be assumed that the majority of coached teachers had reached the "expert; greater than 5 years" designation in Berliner's model. This may suggest that there is a threshold at which point teachers begin to perceive integrating new skills into their repertoire as a challenge.

This distinction may also highlight the role of examining student responsiveness as discussed in the previous section. For example, student outcomes (i.e., responsiveness) may be a predictor of intervention acceptability. Related to the previous discussion of self-efficacy, teacher perception of student outcomes may also be pertinent. For example, a student's responsiveness may have in fact increased significantly by posttest, but ultimately it is the teachers' perceptions of student outcomes that may impact their reporting of acceptability, including a commitment to future implementation. These hypotheses will be unknown unless tested and thus, serve as appropriate next steps for research. A thorough understanding of intervention acceptability would include data collection during a maintenance phase, both with the teachers' current students and into subsequent school years. Additionally, qualitative methods may help to identify other intervention determinants not captured by other study measures.

These findings may also provide specific guidance for practice. Aligned with the previous discussion of self-efficacy and the existing research base, self-efficacy appears to be a strong predictor of teacher behavior when implementing new instructional techniques (e.g., McKinney et al., 1999; Tschannen-Moran et al., 1998) and student outcomes (e.g., Anderson et al., 1988; Ross, 1992). Therefore, it may be purposeful to increase opportunities for teachers (particularly more experienced veteran teachers) to engage with data in a way that highlights positive changes over time. For example, the BEST in CLASS coaching framework includes the use of some self-actualization exercises including self-reflection, review of video artifacts, and goal setting and feedback. The use of these tools provides access to some student behavior data, but heavy emphasis is placed on teacher behaviors. It may be useful to include student responsiveness data in these activities, and also as a part of the post-test data collection efforts. For example, a final coaching session could be held with an expressed focus on the presentation of student data from the various measures used in the study to provide a more in-depth and thorough examination of outcomes. These activities may affect self-efficacy and intervention acceptability, ultimately leading to increased adoption and sustainment of the intervention practice elements.

Limitations and Implications for Future Research

Findings from this study should be interpreted with the following limitations in mind. Related to participant selection, teachers included in the study were volunteers who may have been more motivated to increase their teaching abilities than the larger population of teachers. Additionally, the act of volunteering for a study may encourage natural self-reflection, which may be responsible for some changes in behavior for teachers in both conditions. Lastly, all participants are teachers who have students who continue to display patterns of challenging

behaviors despite their efforts. As a result, there may be impacts to self-efficacy, classroom climate, and motivation that are particular to this participant group.

There may also be limitations related to methods used for data collection. For example, the TIES measure used for RQ 2 is collected by highly trained coaches, serving as a single observer. This is in contrast to data collected for RQ1, which used two trained observers. Further, for RQ2, the observer (i.e., coach) conducts the direct observations throughout the intervention. Thus, this observer becomes very familiar with the classroom, which could lead to more reliable data collection but also allows the possibility for bias. As stated in the next steps for research, the lack of a control group when examining teacher responsiveness in RQ 2 may be a threat to maturation. Lastly, each practice was tested in a separate model which did not allow analysis of the relations that may exist between the practices. Previous work has provided evidence that the use of some practices may be related to one another (e.g., opportunities to respond and praise; see Sutherland et al., 2002).

While the findings from each research question have specific implications for future research, some distinct themes have emerged. In order to increase the validity of outcomes, certain methodological characteristics that are aligned with the theoretical model should be used. These include the continued use of direct observation methods, a nested approach to account for differences between teachers, and years of experience as a continuous variable. To contend with the threat of maturity, or the natural change in skill that the theoretical model assumes, longitudinal data collection (coupled with non-linear analysis) should be used. Further, a comparison or control group would strengthen this design. Related to the identification of variables, the continued examination of teacher-level factors (i.e. perception of student behaviors, self-efficacy) and addition of student-level factors (i.e. intensity and duration of

behavior, responsiveness) has been suggested. Outcomes from these investigations may indicate that a revision to the theoretical model is needed, perhaps with a greater emphasis on the types of experiences teachers accumulate, rather than the number of years in the profession. Lastly, to better understand the degree to which teacher-driven professional development methods such as coaching can mitigate teacher differences, it would be purposeful to extend this line of research.

Conclusion

This study aimed to bridge the research-to-practice gap related to in-service supports and professional development by identifying factors associated with teacher use of evidence-based classroom management practices and responsiveness to a coaching intervention. While years of experience did not predict practice use or responsiveness, findings may position teacher-driven interventions (i.e., coaching) as a mechanism to improve teacher practice regardless of previous teaching experience. Evidence of this can be seen in the alignment between the results of this study and intervention outcomes identified in the literature review. Recommendations for future studies include a continued emphasis on direction observation measures, a nested approach that can test and account for teacher-level factors, and the importance of longitudinal models and comparison groups to mitigate the threat of maturity. Implications for practice include the continued use of teacher-driven interventions, with more purposeful engagement with student and teacher data by the teacher themselves. Given the strong link between teacher quality and student outcomes, this study adds important perspective to our understanding of teacher development and ultimately, student success.

References

Acock, A. C. (2018). A gentle introduction to STATA (6th ed.). Stata Press.

Adam, R. D. (1982). Teacher development: A look at changes in teacher perceptions and behavior across time. *Journal of Teacher Education, 33*(4), 40-43.

Altman, D. G., & Royston, P. (2006). The cost of dichotomising continuous variables. *BMJ, 332*(7549), 1080. https://doi.org/10.1136/bmj.332.7549.1080

Asparouhov, T., Muthén, B. (2006). Multilevel modelling of complex survey data. Proceedings of the Survey Research Methods Section, American Statistical Association 2006, 2718–2726.

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

Bellibas, M. S., & Liu, Y. (2017). Multilevel analysis of the relationship between principals' perceived practices of instructional leadership and teachers' self-efficacy perceptions. *Journal of Educational Administration, 55*(1), 49-69.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Berliner, D. C. (1988). The development of expertise in pedagogy [Paper presentation]. Annual Meeting of the American Association of Colleges for Teacher Education, New Orleans, LA.

Biemer, P. P., & Lyberg, L. (2003). *Introduction to survey quality*. Wiley.

Bierman, K.L., Greenberg, M.T., & Abenavoli, R. (2016). Promoting social and emotional learning in preschool: Programs and practices that work. Edna Bennet Pierce Prevention Research Center, Pennsylvania State University.

Briesch, A. M., Chafouleas, S. M., Neugebauer, S. R., & Riley-Tillman, C. T. (2013). Assessing influences on intervention implementation: Revision of the Usage Rating Profile-Intervention. *Journal of School Psychology*, *51*(1), 81–96. https://doi.org/10.1016/j.jsp.2012.08.006

Carver-Thomas, D., Leung, M., & Burns, D. (2021). California teachers and COVID-19: How the pandemic is impacting the teacher workforce. Learning Policy Institute.

Chafouleas, S. M., Briesch, A. M., Neugebauer, S. R., & Riley-Tillman, T. C. (2011). Usage Rating Profile – Intervention (Revised). Storrs, CT: University of Connecticut.

Clarridge, P. B. (1990). Multiple perspectives on the classroom performance of certified and uncertified teachers. *Journal of Teacher Education, 41*(4), 15-25.

Conroy, M. A., Sutherland, K. S., Algina, J., Ladwig, C., Werch, B., Martinez, J., Jessee, G., & Gyure, M. (2019). Outcomes of the BEST in CLASS intervention on teachers' use of effective practices, self-efficacy, and classroom quality. *School Psychology Review, 48*(1), 31-45. https://doi.org/10.17105/SPR-2018-0003.V48-1

Cook, R.D., 1977. Detection of influential observation in linear regression. *Technometrics 19*(1), 15–18.

Crowl, T. K, & Berkowitz, M. (1985). Regular and special education teachers' marking attitudes and behavior. *Contemporary Educational Psychology, 10*(3), 260-267.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives, 8*(1). https://doi.org/10.14507/epaa.v8n1.2000

Darling-Hammond, L., Hyler, M. E., Gardner, M. (2017). Effective Teacher Professional Development. Palo Alto, CA: Learning Policy Institute.

Darling-Hammond, L., & Cobb, V. L. (1995). Teacher preparation and professional development in APEC members: A comparative study. Washington, D.C.: U.S. Department of Education.

Duong, J. & Bradshaw, C. P. (2013). Using the extended parallel process model to examine teachers' likelihood of intervening in bullying. *The Journal of School Health, 83*(6), 422-429.

Early, D. M., Maxwell, K. L, Ponder, B. D., & Pan, Y. (2017). Improving teacher-child interactions: A randomized controlled trial of Making the Most of Classroom Interactions and My Teaching Partner professional development models. *Early Childhood Research Quarterly, 38*, 57-70.

Egyed, C. J., & Short, R. J.. (2006). Teacher self-efficacy, burnout, experience and decision to refer a disruptive student. *School Psychology International, 27*(4), 462-474.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models. *Psychological Methods, 12*(2), 121–138. https://doi.org/10.1037/1082-989X.12.2.121

Everhart, B. & Vaugh, M. (2005), A comparison of teaching patterns of student teachers and experienced teachers in three distinct settings: Implications for preparing teachers for all settings. *Education, 126*(2) 221-239

Evertson, C. M., & Weinstein, C. S. (2006). *Handbook of classroom management : research, practice, and contemporary issues*. Lawrence Erlbaum Associates.

Feuerborn, L. & Chinn, D. (2012). Teacher perceptions of student needs and implications for positive behavior supports. *Behavioral Disorders, 37*(4), 219-231.

Ficarra, L., & Quinn, K. (2014). Teachers' facility with evidence-based classroom management practices: An investigation of teachers' preparation programmes and in-service conditions. *Journal of Teacher Education for Sustainability, 16*(2), 71–87. https://doi.org/10.2478/jtes-2014-0012

Field, A. (2018). *Discovering statistics using IBM SPSS Statistics (5th Ed.).* Los Angeles, CA: Sage Publications.

Francisco, M. P., Hartman, M., & Wang, Y. (2020). Inclusion and special education. *Education Sciences, 10*(9), 238.

Gable, R., Tonelson, S., Sheth, M., Wilson, C., & Park, K. (2012). Importance, usage, and preparedness to implement evidence-based practices for students with emotional disabilities: A comparison of knowledge and skills of special education and general education teachers. *Education and Treatment of Children, 35*(4), 499-519.

Gage, N. A., Adamson. R., MacSuga-Gage, A. S. & Lewis, T. J.. (2017). The relation between the academic achievement of students with emotional and behavioral disorders and teacher characteristics. *Behavioral Disorders, 43*(1), 213-222.

Gregory, A., Allen, J. P., Mikami, A.Y., Hafen, C. A, & Pianta, R. C. (2014). Effects of a [professional development program on behavioral engagement of students in middle and high school. Psychology in the Schools, 51(2), 143-163.

Han, S. S., & Weiss, B. (2005). Sustainability of teacher implementation of school-based mental health programs. *Journal of Abnormal Child Psychology, 33*(6), 665-679. https://doi.org/10.1007/s10802-005-7646-2

Hoy, W. K. & Woolfolk, A. E. (1990). Socialization of student teachers. *American Educational Research Journal, 27*(2), 279-300.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions," in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (1st ed.) by L. M. LeCam and J. Neyman. Berkeley: University of California Press, 1967.

Hussar, B., Zhang, J., Hein, S., Wang, K., Roberts, A., Cui, J., Smith, M., Bullock Mann, F., Barmer, A., and Dilig, R. (2020). The condition of education 2020 (NCES 2020-144). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Joyce, B. R., & Showers, B. (2002). Student achievement through staff development. Alexandria, VA: Association for Supervision and Curriculum Development.

Kan, A., & Bulut, O. (2014). Crossed random-effect modeling: Examining the effects of teacher experience and rubric use in performance assessments. *Eurasian Journal of Educational Research, 14*(57).

Kearney, P., Plax, T. G., Sorensen, G., & Smith, V. R. (1988). Experienced and prospective teachers' selections of compliance-gaining messages for "common" student misbehaviors. *Communication Education, 37*(2), 150-164.

Klassen, R. M., & Chiu, M. M. (2010). Effects on teachers' self-efficacy and job satisfaction. *Journal of Educational Psychology, 102*(3), 741-756.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kraft, M., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research, 88*(4), 547-588. https://doi.org/10.3102/0034654318759268

Kretlow, A., Cooke, N., & Wood, C. (2012). Using in-service and coaching to increase teachers' accurate use of research-based strategies. *Remedial and Special Education, 33*(6), 348-361. https://doi.org/10.1177/0741932510395397

Kunemund, R. L., Nemer McCullough, S., Williams, C. D., Miller, C. C., Sutherland, K. S., Conroy, M. A., & Granger, K. (2020). The mediating role of teacher self-efficacy in the relation between teacher–child race mismatch and conflict. *Psychology in the Schools, 57*(11), 1757–1770. https://doi.org/10.1002/pits.22419

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology, 105*(3), 805–820. https://doi.org/10.1037/a0032583

Kwok, A. (2018). Promoting "quality" feedback: First-year teachers' self-reports on their development as classroom managers. *The Journal of Classroom Interaction, 53*(1), 22-36.

Lazic, S. E. (2008). Why we should use simpler models if the data allow this: Relevance for ANOVA designs in experimental biology. *BMC Physiology, 8*(6). https://doi.org/10.1186/1472-6793-8-16

Leinhardt, G. (1989). Math lessons: A contrast of novice and expert competence. *Journal for Research in Mathematics Education, 20*(1), 52-75.

Lock, C., & Lichtenstein, J. (1980). Some research implications for the development of teacher inservice programs. *Kappa Delta Pi Record, 17*(1), 11.

MacFarlane, K., & Woolfson, L. M. (2013). Teacher attitudes and behavior toward the inclusion of children with social, emotional and behavioral difficulties in mainstream schools: An application of the theory of planned behavior. *Teaching and Teacher Education, 29*, 46-52.

Marzano, R., Marzano, J. S., & Pickering, D. (2003). *Classroom management that works: Research-based strategies for every teacher*. Alexandria, VA: Association for Supervision and Curriculum Development.

McCullough, S. N., Granger, K. L., Sutherland, K. S., Conroy, M. A., & Pandey, T. (2022). A preliminary study of BEST in CLASS–Elementary on teacher self-efficacy, burnout, and attributions. *Behavioral Disorders*, *47*(2), 84–94. https://doi.org/10.1177/01987429211010672

McDonald C., Carol, S., Seung-Hee, H., Annemarie H., & Morrison, F. J. (2005). Teacher qualifications, classroom practices, family characteristics, and preschool experience: Complex effects on first graders' vocabulary and early reading outcomes. *Journal of School Psychology*, *43*(4), 343-375.

McLeod, B. D., Kunemund, R., Nemer, S., & Lyon, A. R. (2020). Leveraging implementation science and practice to support the delivery of evidence-based practices in services for youth with emotional and behavioral disorders. In T. W. Farmer, M. A. Conroy, E. M. Z., Farmer, & K. S. Sutherland (Eds), Handbook of research on emotional and behavioral disorders: Interdisciplinary developmental perspectives on children and youth (pp. 417-432). New York: Routledge/Taylor & Francis.

McMillan, J. H. (2016). Fundamentals of Educational Research (7th Edition). Pearson Education, Inc. ISBN-10: 0133579166

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

National Institute of Justice. (2018, November 21). Evidence Based Programs. https://crimesolutions.gov/ProgramDetails.aspx?ID=627

Needels, M. (1991). Comparison of student, first-year, and experienced teachers' interpretations of a first-grade lesson. *Teaching and teacher education 7*(3) 269-278.

Noltemeyer, A., Kunesh, C., Hostutler, C., Frato, P., Sarr-Kerman, B. J. (2012). The effects of student and teacher characteristics on teacher impressions of--and responses to--student behaviors. *International Education Studies*, v5 n4 p96-111.

O'Connor, E. A., Fish, M. C., & Yasik, A. E. (2004). The Influence of teacher experience on the elementary classroom system: An observational study. *The Journal of Classroom Interaction, 39*(1), 11-18.

Office of Special Education and Rehabilitative Services, U.S. Department of Education. (2020). 41st annual report to Congress on the implementation of the individuals with disabilities education act, 2019.

Paulsen, James R. (1977). Hey teacher, your personality's showing! *School Science and Mathematics, 77*(3) 237-240.

Pearce, M., & Forlin, C. (2016). Challenges and potential solutions for enabling inclusion in secondary schools. *Australasian Journal of Special Education, 29*(2).

Peters, C. D., Kranzler, J. H., Algina, J., Smith, S. W., & Daunic, A. P. (2014). Understand disproportionate representation in special education by examining group differences in behavior ratings. *Psychology in the Schools, 51*(5), 452-465.

Pressley, T., Croyle, H., & Edgar, M. (2020). Different approaches to classroom environments based on teacher experience and effectiveness. *Psychology in the Schools, 57*(4), 606-626.

Podell, D. M., & Tournaki, N. (2007). General and special educators' predictions of student success as a function of learner characteristics and teacher experience. *Teacher Education and Special Education, 30*(4), 249-263.

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(2), 65–76. https://doi.org/10.1007/s10488-010-0319-7

Raudenbush, & Bryk, A. S. (2002). *Hierarchical linear models : applications and data analysis methods* (2nd ed.). Sage Publications.

Rice, J.K. (2010). The impact of teacher experience: Examining the evidence and policy implications. National Center for Analysis of Longitudinal Data in Education Research.

Ronfeldt, M., & Reininger, M. (2012). More or better student teaching? *Teaching and Teacher Education*, *28*(8), 1091–1106. https://doi.org/10.1016/j.tate.2012.06.003

Royston, P. (1992). Approximating the shapiro-wilk W-test for non-normality. *Statistics and Computing, 2(*3), 117-119. https://doi.org/10.1007/BF018912.

Ritter, J. T., & Hancock, D. R. (2007). Exploring the relationship between certification sources, experience levels, and classroom management orientations of classroom teachers. *Teaching and Teacher Education, 23*(7), 1206-1216.

Salend, S. J., & Garrick D., Laurel M. (1999). The impact of inclusion on students with and without disabilities and their educators. *Remedial and Special Education, 20*(2), 114-126.

Sameroff, A. J. (1995). General systems theory and developmental psychopathology. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology: Vol. 1. Theory and Methods* (pp. 659–695). New York: Wiley and Sons.

Scruggs, T. E, & Mastropieri, M. A. (1996). Teacher perceptions of mainstreaming/inclusion, 1958–1995: A research synthesis. *Exceptional Children*, *63*(1), 59–74. https://doi.org/10.1177/001440299606300106

Shoulders, T. & Scott K. M. (2015). Rural high school teachers' self-efficacy in student engagement, instructional strategies, and classroom management. *American Secondary Education, 44*(1), 50-61.

Simonsen, B., Freeman, J., Myers, D., Dooley, K., Maddock, E., Kern, L., & Byun, S. (2020). The effects of targeted professional development on teachers' use of empirically supported classroom management practices. *Journal of Positive Behavior Interventions, 22*(1), 3-14.

Snedecor, G. W. & Cochran, W. G. (1989), *Statistical Methods, Eighth Edition*. Iowa State University Press.

Snyder, P., Hemmeter, M., & Fox, L. (2015). Supporting implementation of evidence-based practices through practice-based coaching. *Topics in Early Childhood Special Education, 35*(3), 133-143. https://doi.org/10.1177/0271121415594925

Soslau, E. (2012). Opportunities to develop adaptive teaching expertise during supervisory conferences. *Teaching and Teacher Education, 28*(5), 768–779. https://doi.org/10.1016/j.tate.2012.02.009

Steele, C. F. (2010). Inspired Responses. *The Effective Educator, 64*(4), 64-68.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*, 173–180.

Stronge, J.H., Ward, T.J., Tucker, P.D. (2007). What is the relationship between teacher quality and student achievement? An exploratory study. *Journal of Personnel Evaluation in Education 20*, 165–184. https://doi.org/10.1007/s11092-008-9053-z

Stough, L.M., Montague, M.L., Landmark, L.J., & Williams-Diehm, K. (2015). Persistent classroom management training needs of experienced teachers. *The Journal of Scholarship of Teaching and Learning, 15*(5), 36-48.

Sutherland, K. S., Conroy, M. A., Algina, J., Ladwig, C., Jessee, G., & Gyure, M. (2018). Reducing child problem behaviors and improving teacher-child interactions and relationships: A randomized controlled trial of BEST in CLASS. *Early Childhood Research Quarterly, 42*, 31–43. https://doi.org/10.1016/j.ecresq.2017.08.001

Sutherland, K. S., Conroy, M. A, McLeod, B. D., Algina, J., & Kunemund, R. L. (2018). Factors associated with teacher delivery of a classroom-based tier 2 prevention program. *Prevention Science, 19*(2), 186–196. https://doi.org/10.1007/s11121-017-0832-y

Sutherland, K. S., Conroy, M. A., McLeod, B. D., Algina, J., & Kunemund, R. L. (2017). Factors associated with teacher delivery of a classroom-based tier 2 prevention program. *Prevention Science, 19*(2), 186–196. https://doi.org/10.1007/s11121-017-0832-y

Sutherland, K. S, Conroy, M. A, McLeod, B. D, Granger, K., Broda, M., & Kunemund, R. (2020). Preliminary study of the effects of BEST in CLASS–Elementary on outcomes of elementary students with problem behavior. *Journal of Positive Behavior Interventions, 22*(4). https://doi.org/10.1177/1098300719900318

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

Sutherland, K. S., Lewis-Palmer, T., Stichter, J., & Morgan, P. L. (2008). Examining the influence of teacher behavior and classroom context on the behavioral and academic outcomes for students with emotional or behavioral disorders. *The Journal of Special Education, 41*(4), 223–233. https://doi.org/10.1177/0022466907310372

Sutherland, K. S., McLeod, B. D., & Kunemund, R. L. (2017). Scoring manual for the observational treatment integrity instrument for elementary school classrooms [Unpublished manuscript]. Virginia Commonwealth University

Sutherland, K. S., McLeod, B. D., Conroy, M. A., Lyon, A. R., & Peterson, N. (in press). Implementation science in special education: Progress and promise. In B. Talbott, K. McMaster, D. Lee, & T. Farmer. (Eds.), *Handbook of special education research: Theory, methods, & leveraging developmental processes.* Routledge.

Sutherland, K. S, & Oswald, D.P. (2005). The relationship between teacher and student behavior in classrooms for students with emotional and behavioral disorders: Transactional processes. *Journal of Child and Family Studies, 14*(1), 1–14. https://doi.org/10.1007/s10826-005-1106-z

Sutherland, K. S., Wehby, J. H., & Yoder, P. J. (2002). Examination of the relationship between teacher praise and opportunities for students with EBD to respond to academic requests. *Journal of Emotional and Behavioral Disorders, 10*(1), 5-13. https://doi.org/10.1177/106342660201000102

Teo, T., Tsai, L. T., & Yang, C.-C. (2013). Applying structural equation modeling (SEM) in educational research: An introduction. In *Application of Structural Equation Modeling in Educational Research and Practice* (pp. 3–21). SensePublishers. https://doi.org/10.1007/978-94-6209-332-4_1

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*, 1–10.

U.S. Department of Education. (2021). *Certification requirements by state*. https://www2.ed.gov/teachers/jobs/reqs/edpicks.jhtml

Walker, H. M., Severson, H. H., Todis, B. J., Block-Pedego, A. E., Williams, G. J., Haring, N. G., & Barckley, M. (1990). Systematic Screening for Behavior Disorders (SSBD): Further Validation, Replication, and Normative Data. *Remedial and Special Education*, *11*(2), 32–46. https://doi.org/10.1177/074193259001100206

Weber, N. D, Waxman, H. C., Brown, D. B. & Kelly, L. J.. (2016). Informing teacher education through the use of multiple classroom observation instruments. *Teacher Education Quarterly 43*(1), 91-106.

Whitford, D. K., Zhang, D., & Katsiyannis, A. (2018). Traditional vs. alternative teacher preparation programs: A meta-analysis. *Journal of Child and Family Studies*, *27*(3), 671–685. https://doi.org/10.1007/s10826-017-0932-0

White, H. (1980). A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. Econometrica 48(4), 817-838.