# Student Perceptions of the Classroom Environment: Actionable Feedback as a Catalyst for Instructional Change

A DISSERTATION SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF THE UNIVERSITY OF MINNESOTA BY

Peter M. Nelson

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

James E. Ysseldyke, Adviser

May 2014

## Acknowledgements

I wish to thank my dissertation committee for their support throughout this project. Without their expertise, many important components of this work might have been overlooked and left to the fray. In particular, I would like to thank my adviser, Dr. James Ysseldyke, whose previous work, ongoing mentorship, and insight on educational issues brought me to the University of Minnesota. His support was invaluable in the development of this project. I also wish to thank Dr. Theodore Christ for his mentorship throughout this project and during my time at the University of Minnesota. Dr. Christ's rigorous standards for scholarship and work ethic contributed immensely to the growth of this project and have played a large role in my development as a researcher.

I also owe a great deal to the teachers who agreed to participate in this project. The classroom is a personal place for all teachers, and I am very grateful for their willingness to open their doors to me and this project. Finally, I am forever indebted to the best editor and wife in the world, Anna Nelson. As an English major and former English teacher, Anna's perspective and keen attention to detail continue to improve the clarity and relevance of my work.

## Dedication

For their unconditional willingness to pick me up, dust me off, and set me on my way, this project is dedicated to my mom and dad.

## Abstract

The current study explored the impact and feasibility of using students' perceptions of the classroom teaching environment as an instructional tool for teachers. Data were collected using the Responsive Environmental Assessment for Classroom Teaching (REACT)—a questionnaire assessing students' perceptions of specific components of instructional support. A total of 31 suburban teachers serving 797middle-school students were assigned to an experimental feedback group or a control group. Students' responses on the REACT served as the primary dependent variable and were collected at three time points throughout the fall semester (Time 1, 2, and 3). The experimental group participated in a feedback meeting following the first data collection and teachers assigned to the control group participated in a feedback meeting following the second data collection. All teachers completed a short survey evaluating the REACT and procedures for implementation.

A multi-level approach to data analysis (HLM) was used to adjust for the natural clustering of students by teacher. Controlling for other variables in the final model, students' responses at Time 2 were significantly higher in the classrooms of teachers who received feedback after Time 1 relative to those of teachers who did not receive feedback. Students' self-reported trouble in class, initial REACT score, and gender were also significant predictors of REACT scores at Time 2 and Time 3. No group differences were observed at Time 3. The descriptive and inferential data observed in the current study offer preliminary support for the REACT as a tool for instructional support.

# Table of Contents

**List of Tables**

## List of Figures



.

# Chapter 1

## Introduction

Most researchers agree that academic and behavioral outcomes among students are partially influenced by teachers. Accordingly, a great deal of research seeks to identify specific components of classroom environments that are associated with positive student outcomes. Indeed, entire books have been dedicated to research on such components (e.g., Hattie, 2009). As a natural byproduct of such inquiries, the manner in which information is collected on relevant instructional variables is well established. It is important to note, however, that research on which instructional variables matter and how they should be measured is necessary, but not sufficient. Given the abundance of research on effective instructional strategies across all methodologies (e.g., observation, interview, surveys), it is disheartening to see far less work on ways to translate those methods into meaningful changes in teacher behavior and student outcomes.

The available research on the use of student perceptions to assess the classroom environment suggests that this method of data collection, like many others, could benefit from renewed attention toward translating student perceptions into action. Here, the argument is not that research using students' perceptions of the classroom environment is lacking. Nor is it that the methods for collecting data on student perceptions are flawed. In fact, it is because there is extensive research on these topics that the field stands to benefit from empirical work exploring student perceptions as a mechanism for formative feedback. Few researchers have addressed how student ratings of the instructional environment can be used as catalysts for instructional change. Even fewer projects have done so while guarding against potential threats to internal and external validity. Thus, more rigorous examinations of student perception as a guide for making changes to instruction are practically relevant and will significantly contribute to the empirical literature on student-teacher feedback—regardless of the outcome.

The general body of research on students' perceptions of the classroom environment offers some evidence for the use of student ratings as a feedback tool (Blose & Fisher, 2003; Fraser et al., 1982; Waldrip et al., 2008). First, the demonstrated relationship between students' perceptions of different aspects of the classroom environment and motivation, attitude toward school, and achievement suggests that student ratings may be worthwhile as an outcome variable in empirical research (Eccles et al., 1993; Fast et al., 2010; Goodenow, 1993; Ryan & Patrick, 2001). Second, it is clear that student ratings of the classroom environment offer a unique look into instruction as it is experienced by those who matter most. Teachers and students often differ in their perceptions of the classroom environment, but student views are more predictive of future academic performance (Bernaus & Gardner, 2008; Desimone, et al., 2010).

From the available research, it follows that students' perceptions of the classroom environment are likely to be useful as a platform for teacher feedback and as an outcome measure in their own right. What remains for discussion are questions regarding what components of the classroom environment should be examined and how student reflections on those components are expected to change teacher behavior. There is a need for research that values student ratings as a tool for meaningful professional development—but what aspects of that process will promote the largest effect?

If the focus is on using measurement to inform practice, the literature on effective professional development (PD) is perhaps the best source of inquiry. Yoon and colleagues (2008) advocate for careful thought into the *theory of instruction* and *theory of instructional change* underlying any given professional development initiative. In this case, the theory of instruction outlines the connection between the skills emphasized in the PD and the intended outcomes of the PD. Conversely, the theory of instructional change refers to the components of the PD that are expected to promote changes in teacher knowledge or behavior. Taken together, these theories explicitly address the "what" and "how" of approaches to professional development.

Although previous examinations of students' perceptions of the classroom teaching environment rarely provide an explicit discussion of theories of instruction and instructional change, the distribution in content among the instruments included in Table 1 indicates that there are a wide variety of factors deemed to be most important for improving student outcomes. Given that the current project differs somewhat from previous work, it will be helpful to briefly explain the theory behind the content and feedback process for the REACT. Although specific details of the current study are discussed in the methods chapter, theories of instruction and teacher change are discussed here to provide clarity.

**Theory of Instruction: Content**

The theory of instruction relates primarily to the nature of the content in PD. There are many effective instructional strategies supported in the literature. It is the belief of the author that these strategies should serve as the framework for effective PD. In the case of student ratings, nearly every item should align with strategies for instruction supported by empirical research. It follows that feedback derived from these items would be tied to evidence-based strategies for improvement and that teachers who use these strategies would likely observe improvements in students' perceptions of the classroom environment. For example, the use of targeted, specific, praise for appropriate academic and social behaviors is consistently supported in the literature on effective instruction (e.g., Brophy, 1980; Cameron & Pierce, 1994; Lewis et al., 2004). Thus, it certainly makes sense to ask students to report on the rate of praise in the classroom. Teachers who observe negative responses on these items would be referred to strategies intended to enhance positive reinforcement in the classroom. It is important to note that unlike other measures, the theory of instruction presented in the current study is not bound by a specific pedagogical orientation (e.g., constructivism or behaviorism)—it is bound by the belief that the classroom environment should be characterized by instructional strategies with sufficient empirical support.

**Theory of Instructional Change: Process**

At the most fundamental level, it is expected that simply viewing student responses on items connected to evidence-based teaching strategies should promote changes in teacher behavior. For many teachers, students' views of the classroom teaching environment uncover a discrepancy between what the teacher believes to be occurring and what students believe to be occurring. This discrepancy, in conjunction with the provision of specific strategies for improvement, is likely to stimulate teacher behavior. In addition to the natural process of reflection associated with viewing students' perceptions, special attention to guidelines for effective professional development is also warranted.

The field of professional development in education is large and there are a myriad of strategies and practices considered to fit the bill for "professional development." Nevertheless, all professional development opportunities are not created equal. Despite some disagreements, there is consistency among teachers and researchers regarding which features of PD are effective and which are not (Desimone, 2009; Garet et al., 2001; Guskey, 2003). Among other things, PD initiatives are improved by (a) an explicit focus on teachers' own classrooms, (b) the alignment of items with evidence-based teaching practices, and (c) the promotion of active teacher participation (Garet et al., 2001). As will be evident in the Methods chapter, the impact of the REACT leans heavily on these characteristics of professional development.

**Purpose**

Despite extensive research on which components of instruction warrant measurement in the classroom (as well as the development of a variety of assessment instruments) it is far less common for researchers to use these assessment tools to improve student outcomes. This is somewhat of a paradox given the common notion among school psychologists that assessment should be closely tied to intervention. If the educational community—both applied and academic—continues to call for an explicit link between assessment and intervention, why do researchers habitually stop short of testing that link? Recommendations for linking assessment and intervention are far more meaningful in the presence of empirical support for such a relationship.

The absence of an established link to intervention is particularly true for research on students' perceptions of the classroom environment. Years of educational research have established the potential utility of students' perceptions as predictors of student outcomes and as a reference for PD, but there is a need to embed these measurements in a working model of PD. The literature on effective instruction and effective PD offer a valuable resource for such endeavors.

The current project explicitly addresses the noted gaps in research on student perceptions of the classroom environment—particularly those related to teachers' use of student perceptions as a guide for instruction. More specifically, a student-report measure (the REACT) was developed to (a) measure students' perceptions of evidence-based components of the classroom environment and (b) provide specific and actionable feedback for teachers in a manner commensurate with effective professional development. Given the intended use of the REACT as a formative measurement tool, the primary research questions are as follows:

1. To what extent do teachers find the REACT to be useful as a tool to improve student perceptions of the classroom teaching environment?
2. To what extent do teachers find information obtained using the REACT to be clear for interpretation?
3. To what extent do teachers anticipate carrying out a "reaction plan" following feedback from REACT?
4. To what extent do teachers report carrying out their reaction plan?
5. To what extent do student ratings improve in classrooms using the REACT feedback relative student ratings in classrooms not using REACT feedback?

# Chapter 2

## Literature Review

Interest in factors associated with improved student academic achievement persists throughout the history of education (Bloom, 1968; Brophy, 1986; Fraser, 1994; Hattie, 2009). Within the last 30 years, educators and lawmakers have increased this focus with intensive efforts to improve the academic achievement of America's school children (Gardner, 1983; No Child Left Behind [NCLB], 2001; Race to the Top [RTTP], 2009). Because the classroom environment is proximal to student learning and largely under the control of teachers, many scholars have focused intently on the social and instructional supports available to students while at school.

Improving the classroom experience of students can be conceptualized as a three-fold process in which educational professionals are tasked with determining (a) what components of classroom environments improve student outcomes (b) how to measure those components, and (c) how to use assessments of those components to improve the status quo. Currently, the former two goals are markedly more established than the latter, with much knowledge on what is important for teaching and far less on how to change the way students interact with their environment (Garet et al., 2010).

The purpose of this review is to introduce various issues in classroom environment research. The first subsection notes the theoretical orientation of those who focus intently on components of the instructional environment and how this view differs from alternative opinions on the source of student success (or lack thereof). This is followed by an annotated account of specific instructional supports grounded in empirical research. The way in which these instructional supports are typically assessed in the classroom is also discussed. The remaining sections deal exclusively with the measurement and use of students' perceptions of the classroom environment. Evidence for the validity of this assessment approach is presented first, followed by a discussion of notable gaps in the student perception literature. Considered together, these subsections should orient the reader to the motivation and theoretical orientation that underlie the creation of the REACT while also highlighting the importance of the current study.

**Attribution of Academic Difficulties**

Although academic and behavioral problems in schools are widely recognized, beliefs about causation differ greatly. Problem attribution, a concept originally explored in attribution theory, may dictate approaches to improving student achievement (Heider, 1958). Although the source of students' problems can be conceptualized in a variety of ways, within-student factors, home-environmental factors, and instructional factors are three of the primary areas of problem attribution in education (Quay, 1973; Taylor & Ysseldyke, 2007). Those who attribute low level performance to within student factors attempt to identify and remediate inherent or acquired deficits that delay or completely inhibit a student's ability to learn (Christenson, Ysseldyke, Wang, & Algozzine, 1983).

A commonly observed example of within-student problem attribution in education may involve labeling a student as academically unmotivated, although within-student problem attribution is common in the context of behavior management as well (Bibou-nakou, Kiosseoglou, Stogiannidou, 2000). Finally, endogenous factors are often referenced in the context of psychopathology (American Psychiatric Association, 2000).

Although internal factors may help explain why a student does not respond to instruction proven to be effective for his or her same-aged peers (e.g., Massetti et al., 2008), such claims are often overused in schools and may detract from more constructive problem-solving (Carnine, Silbert, Kame'enui, & Tarver, 2010). Indeed, teachers who focus on the individual nature of behavior problems are more likely to respond negatively to students (Dobbs & Arnold, 2009; Espinosa & Laffey, 2003). In addition, causal statements concerning within-student deficits may be irrelevant or incorrect (Reschly, 2008). Finally, a strict endogenous perspective on student problems negates many of the environmental supports schools are capable of providing—if student disorder is viewed as something a student either has or does not have, there is little point to considering environmental supports (Sroufe, 1997).

Home-environment factors are also an area of concern in education (Mavropoulou & Padeliadu, 2002) and primarily relate to students' home setting and overall personal experiences (Crane, 1996)**.** Despite some variation in opinion about the importance of the home environment, heavily studied factors such as poverty are likely to play a significant role in students' classroom achievement (Sirin, 2005; White, 1982). It is important to note, however, that the exact relationship between socioeconomic status (SES) and student achievement depends on a variety of factors (e.g., geographic location, school-family relationship) and may in fact be negligible for some groups of people (Sirin, 2005). Further, it is often assumed that SES is highly correlated with contextual variables related to student achievement, but this is not always true. For example, Sui Chui and Willms (1996) employed a multi-level model in an effort to explore the relationship between four dimensions of parental engagement and student achievement. Among other things, the authors observed no relationship between SES and parental engagement. Last, the very definition of low SES—particularly the use of free or reduced lunch status—may mitigate its utility as a predictor of student achievement (Harwell, & LeBeau, 2010).

Parental education level is an additional environmental factor that may play a role in student achievement. In one large-scale study, kindergarten children demonstrated differing performance in letter recognition when grouped by their mother's education level (West, Denton, & Reaney, 2000). More specifically, the children whose mothers had attained at least a bachelor's degree scored nearly 50 percent better than children whose mothers had not completed high school. Along similar lines, the exposure to a diversity of words (Hoff, 2003) and the home literacy environment (Payne, 1994) may also contribute to student achievement.

Despite the recognized importance of the home environment, intervention efforts that address home-based factors are often under less control of school employees as home intervention efforts target factors that occur outside of the school environment and sometimes address deeply-rooted familial characteristics (SES, parenting style, etc). Given the relatively controlled atmosphere of schools, it may be easier for professionals in education to manipulate salient characteristics of the instructional environment. Indeed, a sharpened focus on those pedagogical characteristics that are both effective and alterable may prove more valuable for increasing student achievement (Brophy, 1986; Pianta & Hamre, 2009).

The third category of problem attribution focuses extensively on instructional factors and is marked by the belief that that academic and behavioral problems result primarily from instructional variables. Those who espouse this view attempt to identify ways in which instruction is breaking down for struggling students and to design interventions that match students' instructional needs (Jimerson, Burns & VenDerHeyden, 2006; Tessmer, 1990). Current assessment efforts build on the early work of psychologists who called for shifts in focus from internal student factors to the interaction between individuals and their instructional environment (e.g., Bijou, 1968; Bronfenbrenner, 1979).

Before considering the classroom environment in greater depth, it is important to recognize that although the factors noted above are sometimes considered to be unrelated, this is not how individual, home, or school factors are best conceptualized. Student problems are likely to manifest for a myriad of reasons, some of which are derived from each of the aforementioned factors (internal, home environment, instructional environment) (Bronfenbrenner, 1979)**.** It is highly unlikely that student behavior can be traced to a specific endogenous source, home source, or instructional source. Many scholars now adopt a systems perspective, thereby recognizing the complex interplay of these factors in the development of normal or deviant behavior (Pianta & Walsh, 1998; Sameroff, 1995). Thus, it is important to note many of those who adopt an instructional perspective, while recognizing the proximity and manipulability of the classroom environment, do not assume unilateral causality.

**Characteristics of Effective Instruction**

Due to the heterogeneity of the literature on effective teaching, the definition of a “qualified teacher” is often restricted to superficial descriptives such as prior education, experience, and academic credits earned (e.g.,, NCLB, 2002). While these characteristics are easily measured, there is little research to support their connection to retention (Guarino et al., 2006) or student outcomes (Druva & Anderson, 1983; Early, et al. 2007; Palardy & Rumberger, 2008). Nevertheless, much of the research on effective instruction hinges on the assumption that teachers do indeed play a critical role in student outcomes. Beyond public opinion that individuals who explicitly interact and instruct students on a daily basis are important (Hart & Teeter, 2002), a large body of research provides an empirical foundation for the assumption that teachers account for unique variance in student outcomes (Darling-Hammond, 2000; Nye, Konstantopoulos, & Hedges, 2004; Rockoff, 2004).

For example, Nye et al. (2004) used Project STAR data to assess the effect of teacher characteristics. The primary methodological difference between Nye et al.'s study and other research of the same nature stems from (a) the randomized nature in which students and teachers were assigned to classrooms, and (b) the size and representativeness of the sample. Although the initial aim of Project STAR (class size) differed from research questions pertaining to teacher characteristics, Nye and colleagues posited that by restricting analyses to classrooms of the same size, any confounding effects of class size could be disregarded.

Using a standardized achievement test in reading and mathematics as the primary outcome variable, the authors present several poignant conclusions for those interested in teacher effects. First, achievement differences differed significantly between classes, but not within specific classrooms. Second, although there were some cases in which superficial teacher characteristics—in this case, dichotomized levels of experience and education—had a significant impact on achievement, these never accounted for more than 5% of the variance in achievement. This finding is in agreement with earlier research of a similar nature (e.g., Druva & Anderson, 1983). Third, Nye et al. (2004) found that between-school differences tended to account for less variance than between-classroom (teacher) factors and observed larger teacher effects in low SES schools. The increased variation observed between classrooms in low SES schools aligns with the regular call for quality teaching in high-needs school districts (Berry, 2004; Darling-Hammond, 2010).

The results presented by Nye et al. (2004) leave educational researchers in a difficult position. That is, teaching matters, but what distinguishes an effective teacher from his or her colleagues? The answer may lie, in part, with research on highly specific teaching strategies. A rich history of educational research has helped identify a wide variety of effective instructional strategies (Hattie, 2009). These factors include, but are not limited to: modifying instruction to incorporate small group work (Ruhl, Hughes, & Schloss, 1987), changing the instructional arrangement of the classroom to fit student needs (Bonwell, 1991), peer-assisted learning strategies (PALS) (Fuchs, Fuchs, Mathes, & Simmons, 1997), high levels of positive student-to-teacher interaction (Posamentier & Jaye, 2006), clear expectations (Schunk, 1996), and teacher coaching (Taylor, Pearson, Peterson, Rodriguez, 2004). In some cases, researchers have drawn attention to the concept of instructional match and other individual student characteristics (Ysseldyke & Burns, 2008; Posamentier & Jaye, 2006).

Although each strategy for instruction uniquely contributes to an effective classroom teaching environment, they are often grouped into larger components for assessment (e.g., Anderson & Wahlberg, 1968; Moos & Tricket, 1974; Pianta & Hamre, 2009; Ysseldyke & Christenson, 2002). For organizational purposes, the following section outlines four broad categories of effective instruction and identifies multiple instructional strategies that comprise those components. These categories are not intended to satisfy all views on effective instruction, but serve as one way to organize effective supports in the classroom environment.

**Instructional match.** Although not the first to recognize the important role of the instructional environment, early work by Bloom (1964; 1971) focused on issues of instructional match still prevalent in current empirical and descriptive work today (e.g., Burns, 2008; Cooper, 2003; Ysseldyke & Burns, 2008). More specifically, Bloom's work presented a mastery model in which (among other things) teachers modify their instruction to meet the needs of individual students. This approach to instruction stands in contrast to the traditional model in which teachers provide uniform instruction and instructional materials to students. A consequence of the traditional approach to instruction is that students who respond well to particular methods and materials excel, while students who struggle may falter throughout the school year.

In a broad sense, mastery learning is driven by the philosophical belief that "under appropriate instructional conditions, virtually all students can learn well" (Block & Burns, 1976, pg. 4). Bloom argued that by shifting focus to criterion measures and adapting instruction to student need, educators could increase the number of students who reach mastery (Bloom, 1971). While the effectiveness of this approach is not immune to debate (e.g., Kulik, Kulik and Bangert-Drowns, 1990; Slavin, 1990), educational researchers continue to advocate for many of the components integral to mastery learning (Kyriakides & Creemers, 2007). The issue of instructional match in particular was identified early on as paramount to effective instruction (Block, 1971); however, the concept of instructional match has differed in its conceptualization and application to classroom settings. For clarification, some of these differences are addressed below.

***Matching instruction, assessment, and curriculum*.** One interpretation of "match" is the notion that instruction should align with intended outcomes and assessment instruments (Tyler, 1949; Cohen, 1984). Instructional alignment, although differing slightly from the concept of matching instruction to student need, still requires *at a minimum,* a match between instruction and assessment. When expanded, the concept of alignment may include the curriculum and state-standards as well (Anderson, 2002). In this model, curricula should align with the state-standards, classroom assessments should align with curricula, and as mentioned, instruction should align with classroom assessments (Porter, 2002). Although assessing multi-level congruence is important, even minor misalignments at the classroom level may directly impact student performance (Koczor, 1984). Thus, from this viewpoint, establishing instructional goals and assessments are just as important as instructional practices (Cohen, 1987); however, the likelihood of an appropriate match between instruction and assessment decreases if only a minority of teachers are trained in programs with an explicit focus on effective assessment practices (Stiggins, 2002). This latter point is particularly important given the discrepancies in how teachers are trained (Darling-Hammond, Chung, & Frelow, 2002).

***Matching to academic competence.*** Many of the principles discussed by Bloom in his seminal works are similar to other early research examining the concept of instructional match. For example, early work in task analysis (Resnick, Wang, & Kaplan, 1973) and Haring and Eaton's (1978) instructional hierarchy highlighted the potential benefits of adaptive instruction. Haring and Eaton (1978) suggested that students fall into one of four hierarchically arranged learning stages (acquisition, fluency, generalization, or adaptation). In each stage, students may be reliant on some aspects of instruction more than others (e.g., modeling, feedback, reinforcement). The learning hierarchy serves as the foundation for more recent lines of research indicating that specific academic outcomes improve when instruction is matched to students' level of learning (Burns, Codding, Boice, & Lukito, 2010; Codding et al., 2007; Daly & Martens, 1994; Daly et al., 1999; Connor et al., 2009). As an example, consider the differences between students in the acquisition stage of instruction and those who may be in the fluency stage. Among other things, students in the acquisition stage may require explicit modeling of the academic task and students who are in the fluency stage may benefit most from repeated practice (Haring & Eaton, 1978). In line with this logic, students at the acquisition level may respond better to interventions at the acquisition level (e.g., Incremental Rehearsal; Tucker, 1989), and students at the fluency level may respond more favorably to interventions at the fluency level (e.g., Taped Problems; Samuels, 1979) (Burns et al., 2010).

The concept of instructional match may also include work that emphasizes the significance of performance deficits versus skill deficits (i.e., "can't do" versus "won't do" assessments) (Daly, Witt, Martens, & Dool, 1997; VanDerHeyden & Witt, 2008). Scholars in this area warn against umbrella interventions (Gresham et al., 2001) and seek to evaluate the degree to which students are unable to complete a task relative to the degree to which students lack the motivation to complete a task (VanDerHeyden & Witt, 2008). The result is a framework in which interventions are matched to students' motivational (in the case of a performance deficit) or instructional (in the case of a skill deficit) needs (Daly et al., 1997; Duhon et al., 2004).

***Learning styles versus differentiated instruction.*** Although the concept of instructional match discussed thus far appears to be synonymous with arguments for multiple learning styles (e.g., Dunn, 1990); this is not the case. For the purposes of this paper, it is important to differentiate between the two.

Many proponents of multiple learning styles assert that students' preferred mode of learning (i.e., visual, auditory, kinesthetic) must be considered when differentiating instruction under the assumption that students will learn best when instruction is matched to their personal learning style. Despite research touting this approach (e.g., Sternberg, Grigorenko, & Zhang, 2008), and the plethora of learning style models available to schools (Coffield, Moseley, Hall, & Ecclestone, 2004), most articles written in support of multiple learning styles produce insufficient evidence for a "crossover effect" in which students perform differently according to their respective learning style (Kirschner, Sweller, & Clark, 2010; Pashler, McDaniel, Rohrer, & Bjork, 2009).

Alternative explanations for many of the gains observed in learning styles research have focused on the diversity of instructional methods students are exposed to (e.g., Mayer, 2008). Here, an important distinction is made between individual predisposition for instructional methods and the effectiveness of a diverse approach to instruction. More specifically, using multiple (e.g., lecture, group work, individual work) and appropriate (e.g., varied levels of demand) strategies for instruction should not be interpreted as evidence that individual students will systematically respond to a particular style of instruction (Cooper, 2003; Mayer, 2008). Carefully considering the "who," "what," "where," and "how" of teaching differs from matching instruction to students' alleged "learning profile" (Cooper, 2003). Nevertheless, it bears mentioning that expecting a positive student response to uniform and fixed instructional approaches is equally foolish.

**Classroom management.** In some ways, effective classroom management serves as a catch-all for effective instruction. That is, if effective instruction is occurring, one might expect very few behavior problems to occur in the classroom (Darch and Kame'enui, 2004; Sutherland, Alder, & Gunter, 2003; Wang, Haertel, & Wahlberg, 1993) and when students are given assignments at their instructional level they tend to spend less time engaging in off-task behaviors (Beck, Burns, & Lau, 2009; Gicking & Armstrong, 1978; Treptow, Burns, & McComas, 2007). Thus, several instructional strategies fit well under the larger umbrella of classroom management.

Most classroom management strategies can be partitioned into proactive and reactive categories. Proactive strategies encompass activities occurring *before* problems arise in the classroom. Among other things, this might include adjusting the physical structure of the room, reducing the time between activities, consistently reinforcing positive behavior, or explicitly teaching rules and procedures (Arlin, 1979; Marzano, Marzano, Pickering, 2003). Conversely, reactive management strategies occur after student misbehavior and typically involve how teachers respond to the behavior (e.g., warnings, threats, redirection, ignoring) (Clunies-Ross, Little, & Kienhuis, 2008).

Despite guidelines regarding the effective use of reactive strategies (Emmer, Evertson, & Anderson, 1980; Brophy, 1986), overreliance on reactive discipline is associated with increased levels of stress for teachers and decreased levels of on-task behaviors for students (Clunies-Ross, et al. 2008). Thus, proactive strategies continue to be an area of emphasis in classroom management research and training (Marzano et al. 2003; Sutherland et al., 2003; Wehby & Lane, 2009). In their seminal study of the classroom environment, Emmer et al. (1980) observed 27 third grade teachers up to 10 times during the first three weeks of school. Emmer and his colleagues provided an interesting look into the typology of effective instruction during the first 3-weeks of school. Teachers identified as "effective classroom managers" devoted a considerable amount of time during the first week to proactive strategies (e.g., spending time teaching classroom rules and procedures). In addition, these teachers tended to use a variety of rewards, developed signals for appropriate behavior, and clearly articulated desired behaviors.

Indeed, positive teacher-student interactions have emerged as strong predictors of academic and social success (Pianta, Steinberg, & Rollins, 1995; Hamre & Pianta, 2005). Although building a strong teacher-student relationship is an ongoing and multi-faceted process, targeted praise continues to be a useful instructional tool (Brophy, 1980; Cameron & Pierce, 1994; Lewis et al., 2004**).** When used correctly, praise also tends to be preferred as a reward by most students. After surveying 747 students between the ages of 8 and 12 on preferences for praise and feedback, Burnett (2001) observed a large majority of students (greater than 90%) who reported a desire to be praised for academic achievement and behavior. Nevertheless, students also reported a preference for praise type, with only 31% preferring praise to be a public affair. Similar results were observed in an earlier study by Merrett & Tang (1994).

The flow of a lesson from one activity to the next is also a worthwhile focal point for proactive classroom management. The importance of lesson continuity—or according to Gnagey (1975), the absence of “thrusts” and “dangles,” was recognized early on in research examining teacher behavior (Kounin, 1970). Early studies sought to identify specific behaviors associated with disruptions in "activity flow" and "time flow" within the classroom (Arlin, 1979). Although the idea of lesson momentum, smoothness, and/or signal continuity was originally presented by Kounin in 1970, Arlin (1979) hypothesized that unstructured, vague, or hurried transitions between activities in the classroom would increase the likelihood of a disruption in continuity.

Indeed, data from his 1979 study provided support for this idea. Further, qualitative interviews with teachers who lead the most orderly classrooms revealed techniques that are nearly identical to those identified in more current approaches (e.g., Darch et al., 2003). Strategies for effective transitions and procedures can be considered in isolation (e.g., providing instruction on proper procedures, rewarding smooth transitions); however, specific interventions comprised of several components occurring in concert (e.g., instruction, explicit timing) have also received support (Campbell & Skinner, 2004; Yarbrough, Skinner, Lee, & Lemmons, 2004).

Large-scale applications of proactive behavior management have also gained momentum in schools. Drawing on the conceptual and empirical framework underlying applied behavior analysis (e.g., Baer, Wolf, Risley, 1968; Carr et al., 2002) the Positive Behavior Interventions and Supports framework (PBIS; Sugai et al., 2000) is an example of a research-based, school-wide approach to preventative behavior management (Horner et al., 2009). In line with its applied behavior analysis roots, PBIS includes an explicit focus on (a) teaching appropriate behavioral routines and (b) reinforcing appropriate behaviors (Sugai et al., 2000). In addition, less emphasis is placed on commonly used consequences such as suspension and time-out, something that may be difficult for schools to adjust to without explicit training (Bradshaw et al., 2008). Collectively, strategies for classroom management comprise one of the most important components of effective instruction.

**Instructional presentation.** The way in which instruction is presented is undoubtedly a critical component of effective instruction. In regard to the general format for instructional communication, many educational researchers advocate for a direct approach to instruction (Carnine et al., 2010; Hattie, 2009). Although the title "Direct Instruction" often evokes images of monotonous and unilateral interactions between the teacher and students, it is more accurately conceptualized as a diverse instructional model in which a strong emphasis is placed on relevant contextual variables (Adams & Engelmann, 1996; Carnine et al., 2010). Among the guiding principles of Direct Instruction, is the structured manner in which teachers are expected to interact with students. More specifically, lessons that progress through a sequence of modeling, group work with guidance, and independent work are in line with a direct approach to instruction (Carnine et al., 2010). Conversely, there are a number of student-oriented approaches to instruction relying heavily on self-discovery and exploration that align with a constructivist approach to instructional presentation (e.g., Steffe & Gale, 1995); however, these approaches continue to receive conceptual and empirical criticism (Kirschner et al., 2006; Mayer, 2004).

Many of the specific teaching strategies promoted in the literature on effective instruction are aligned with direct approaches to instruction. For example, the degree to which teachers articulate academic expectations and clearly present instruction is often noted as a component of effective classroom environments (Hattie, 2009; Marzano, 2007) and is aligned with “modeling” as it is discussed in the Direct Instruction literature (Adams & Engelmann, 1996; Carnine et al., 2010). Early work in the area of teacher clarity offered a meaningful distinction between "vague" and "clear" teachers and provided numerous specific behaviors exhibited by teachers who were perceived as clear instructors by students (Hines, Cruickshank, & Kennedy, 1985; Kennedy et al., 1978). More specifically, clear teachers tend to stress important aspects of the content, assess deficiencies in understanding, coach rather than “tell,” and provide frequent and diverse examples (Hines et al., 1985; Taylor et al., 2004). Although the construct of teacher clarity can be somewhat nebulous, early guidelines are similar to later literature-based definitions that focus on verbal clarity and lesson structure, and there is general support for the argument that organizing and presenting lessons coherently is associated with gains in student achievement (Chesebro, 2003; Hattie, 2009; Titsworth, 2001).

**Formative feedback.** When discussing the potential benefits of feedback, it is important to consider whether or not students receive feedback at all, as well as the manner in which that feedback is given. Evidence from several meta-analyses and reviews of feedback suggest that an emphasis on the latter is considerably more important (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Black & William, 1998;

Hattie & Timperly, 2008; Shute, 2008). Although the feedback students receive may be considered separate from instruction, it is more likely that the two are intertwined such that feedback itself serves as a form of effective instruction. Effective feedback typically requires more detailed information than is usually associated with the evaluative grades students receive on classroom tests (Black & William, 1998). In this way, the concept of feedback is congruent with Bloom's conceptualization of Mastery Learning in which students receive formative feedback until a content criterion is reached (Bloom, 1964). Thus, formative feedback can be thought of as feedback intended to improve student performance in a content area rather than feedback intended to simply inform students of their performance (Black & William, 1998).

Shute (2008) outlines various models for delivering feedback and suggests that highly specific recommendations for feedback may have variable effects across contexts. As an alternative, to overly specific rules, he outlines general guidelines for feedback. Among these guidelines are (a) providing specific, objective, feedback related to the task, (b) establishing clear criterion goals and communicating student progress toward those goals, and (c) providing elaborated feedback in simple and manageable units. Likewise, in their meta-analysis of feedback in the classroom, Hattie & Timperly (2008) emphasize a process approach to feedback, indicating that effective feedback (a) adequately articulates the goal at hand, (b) provides information on how students are doing in relation to the goal, and (c) communicates what is needed to reach the goal. While differences exist across specific feedback models, there is general consensus regarding the ineffective nature of relying solely on evaluative forms of feedback—a commonly observed characteristic of the traditional grading system employed in most in K-12 classrooms.

**Assessing the Instructional Environment**

Although achievement tests administered in the classroom may provide helpful instructional information (e.g., Helman, 2005) assessing the environment in which students learn is paramount. There is a gap between evidence-based teaching strategies and teaching strategies used in classrooms (Burns & Ysseldyke, 2008; Scheeler et al., 2004). Thus, simply identifying effective instructional strategies is a necessary but insufficient step toward meeting students' needs in the classroom. Loup, Garland, Ellet, and Rugutt (1996) write, "it seems that any serious effort to improve instruction will need an assessment component to identify teachers who need help" (pg. 129). In addition, Ysseldyke and Christenson (2002) argue that any effort to assess students should not exist in a vacuum—it must include information on the nature of the instructional environment.

In their overview of assessment methodology, Christ, Riley-Tillman, & Chafouleas (2009) suggest that system-level data (e.g., attendance records, academic records, office discipline referrals, and test results) may help to evaluate the effectiveness of broad-based plans, but are potentially less useful at the level of small-groups and individuals, and could in fact be misleading (Koedel & Betts, 2007**.** It is rare for system level data to offer insight into many of the research-based supports discussed in the previous section. Nevertheless, it is common for professionals in education to make decisions using superficial data, a practice that does very little to improve the instructional environment for students (Medley, Coker, & Soar, 1984; NCLB, 2002).

It should be noted that surface-level outcomes—typically in the form of test scores—are in fact important. Ultimately, the goal is for all students to achieve at a high level and in accordance with this goal, the effectiveness of specific instructional strategies or teaching methods are often evaluated using student outcomes. For example, the use of peer-assisted learning was validated by assessing its effect on student outcomes (Fuchs et al., 1997). Indeed, much of the research cited in the previous subsection relies heavily on outcome variables (e.g., achievement) that considered in isolation, are not particularly helpful for meeting student needs. Nevertheless, when research demonstrates that a particular instructional strategy is connected to meaningful outcomes, measurement of the strategy itself may be much more useful in future research. That is, when measuring the extent to which effective instruction is occurring for students, it may be more appropriate to assess the use of previously validated methods. For example, providing frequent and specific feedback for students is consistently associated with student gains (Gersten & Baker, 2001; Hattie, 2009). Given the strong literature base for feedback, educators interested in assessing the instructional environment might do well to operate under the assumption that certain modes of feedback are effective, and make a systematic effort to measure their application in the classroom.

To explore specific components of instruction, educators might include indirect or direct data (Shapiro, 2010). The difference between indirect and direct data collection is synonymous with early distinctions in the classroom environment literature between alpha press (observed phenomena) and beta press (perceived phenomena) (Murray, 1938). Indirect data collection (e.g., interviews and rating scales) typically occurs outside of the target time frame and may rely more heavily on subjective perceptions (Rosenshine, 1970). Given that rating scales are often comprised of specific questions *and* answer choices, they may be considered to be heavily structured interviews (e.g., an item might read "John's behavior in class today was typical of how he usually behaves" and the teacher would be given a rating scale of "strongly agree" to "strongly disagree") (Christ & Boice, 2009).

Direct observation procedures differ from the aforementioned indirect procedures as data collection occurs at the time and place of the behavior in question (Shapiro, 2010). Accordingly, the focus of these procedures is on highly specific and observable components of instruction that require less inference (Rosenshine, 1970). Information gathered in this manner is less subject to the retrospective constraints of interviews and rating scales, but direct observation methods are not immune to criticism. The necessity of an external observer, resource requirements, and the piecemeal approach of direct methods may limit their use in schools.

Direct methods of classroom observation are useful to identify teaching behaviors associated with academic and social competency (Evertson, & Emmer, 1980; Howes et al., 2008; Pianta & Hamre, 2009); however, indirect methods of data collection that rely on students' perceptions of the classroom environment continue to offer a meaningful and unique view into the student experience (Fouts & Meyers, 1992; Fraser, 1998; Goodenow, 1993; Wang & Holcombe, 2010).

Both direct and indirect methods of assessment have the potential to provide meaningful information to teachers; however, perhaps it is most important to consider how assessment data are used. Despite the prevalence of methods for assessing the instructional environment, teachers often improve instruction based solely on experience (VanTassel-Baska, Quek, & Feng, 2007). Given the commonplace nature of this "trial by fire" approach to instruction, it is not surprising to find that inadequate support systems and ineffective professional development are often noted as barriers for dealing with challenging behavior in the classroom (Hoy & Spero, 2005; Westling, 2010). Using instructional assessments that are supportive, rather than purely evaluative and obligatory, may (a) improve the effectiveness of classroom instruction and (b) divert negative attitudes toward "teacher evaluation" held by teachers. The classroom environment is composed of many complex interactions between the teacher and students. Assessments that offer little insight into these interactions, or are used in ways that do not recognize the difficulty associated with developing effective learning environments, are destined to fail.

The remaining subsections discuss students' perceptions of the classroom environment as a potential avenue for classroom assessment and highlight some of the key issues in this area of research. As noted, there are many ways to help teachers meet the needs of students and student feedback is only one. Nevertheless, decades of research in this area provide support for the use of instruments that consider students' perceptions of various social and instructional supports in the classroom.

**Validity of Student Perceptions of the Classroom Teaching Environment**

Before discussing specific findings in research exploring students' perceptions of the classroom environment, it is important to note that "classroom environment" is often conceptualized differently depending on the theoretical orientation of the author. Similarly, it is typical for researchers to examine only one component of the instructional environment. For example, there is an expansive literature on achievement goals—which are one aspect of the classroom teaching environment—that continues to link students' perceptions of classroom goal structure with various cognitive and social outcomes (e.g., Ames & Archer, 1988; Kaplan & Maehr, 1999; Murayama & Elliot, 2009; Wang & Holcombe, 2010). Students' perceptions of another specific component—teacher support—have also received substantial attention, with much research pointing to the critical importance of creating a learning environment in which students feel supported by their teacher (Brewster & Bowen, 2004; Reddy, Rhodes, & Mulhall; Skinner & Belmont, 1993).  Despite the specific measurement focus of some classroom environment studies, the overall theme is similar—the way students *perceive* the classroom environment is an accessible and meaningful source of information.

Broadly speaking, the classroom experience of students holds strong predictive validity for a variety of student outcomes (Burnett, 2002; Dorman, 2002; Greene, Miller, Crowson, Duke, & Akey, 2004; Patrick, Kaplan, & Ryan, 2007). As mentioned, student perceptions of environmental variables such as teacher support may affect their engagement, self-efficacy, and motivation to learn (Friedel, Cortina, Turner, & Midgley, 2010; Nolen, 2003; Ryan et al., 1990; Goodenow, 1993; Wang & Holcombe, 2010). For example, Goodenow (1993) observed a strong positive correlation between students' sense of belonging and their level of achievement and effort. Teacher support—one component of Goodenow's "belonging" construct—explained over a third of the variance in students' perceptions of the value of academic work in their class. More recently, Hughes (2011) observed a statistically significant relationship between students' perceptions of teacher support and self-efficacy. Overall perception of teacher support also predicted students' sense of self belonging. A small relationship was observed between students' perceptions of support and math achievement and there was a small non-significant relationship between students' perceptions and reading achievement. The results of this study suggest that students' perceptions of the classroom environment may not always have a direct effect on achievement. Similar results were observed in a later report by the same author (Hughes et al., 2012).

In one of the earlier classroom environment projects, Fraser & Fisher (1982) examined the predictive validity of two commonly used measures of the classroom environment: the Individualized Classroom Questionnaire (ICEQ) and the Classroom Environment Scale (CES) (Fraser, 1980;Moos & Tricket, 1974; Tricket & Moos, 1973). These measures are outlined in more detail below, but as the name suggests, the ICEQ purports to differentiate between classrooms with individually focused environments and those without. Conversely, the CES aims to provide a broader perspective on the classroom environment.

Both the ICEQ and the CES were administered during the middle of the school year and information on the outcome variables (student cognition and ability) was collected in a pre- post-test format spread across one full school year. Positive correlations between select subscales of the ICEQ and CES and measures of cognition and ability were observed. In addition, evidence is presented suggesting that each scale explained unique variance in the outcome variables. Nevertheless, it is important to note that Fraser & Fisher pay much attention to students' raw post-test scores without including important covariates (e.g., prior levels of student performance).

The middle school environment in particular has served as an area of special interest in classroom environment research, primarily because it is distinct from the elementary classroom environment in many important ways. Among other things, a stronger emphasis is placed on normative comparison in middle school classrooms, with classroom scores taking on new meaning for some students (Anderman, Maehr, & Midgley, 1999). The marked contextual shift in middle school makes it a particularly important time to create and maintain healthy social and instructional environments for students (Goodenow, 1993).

Other scholars have argued that the changes in students' learning environments that coincide with middle school exert a causal influence on student attitude (Feldlaufer, Midgley, & Eccles, 1988; Eccles et al., 1993; Way, Reddy, & Rhodes, 2007; Wigfield, Eccles, Mac Iver, Reuman, & Midgley, 1991). For example, in a longitudinal study, Wigfield et al., (1991) collected data on a group of students before and after entering middle school. On average, students reported lower self-esteem as well as lower attitudes toward school after entering the middle school environment; however, an increase in self-esteem and attitude toward school was observed as the school year progressed. This increase is consistent with later work revealing increases in peer efficacy for students moving from the $7^{th}$ to the $8^{th}$ grade (i.e., students moving "up the chain" in middle school) (Ryan & Patrick, 2001).

Despite the promising results from early research on the predictive validity of student perceptions of various components of the classroom environment, it is worth noting that much of this work did not address the clustered nature of student data, making arguments for statistical conclusion validity somewhat difficult. In some cases, authors have attempted to address the nested nature of classroom data by aggregating to the classroom level (e.g., Fraser & Fisher, 1982). The use of aggregated ratings is appropriate when the results are discussed at the group level, but it is equally (if not more) interesting to examine variation in individual student ratings and account for differences among classrooms in the same statistical model. More advanced techniques for data analysis such as hierarchical linear modeling allow for simultaneous student level and classroom level interpretations by accounting for the nested nature of such data (Marks, 2000; Wheldall, Beaman, & Mok 1999).

On the most fundamental level, it is intuitive to conceptualize variance in student reports of the classroom environment as attributable to three sources: the student, the classroom, and the school. Instruments for assessment that purport to capture aspects of the classroom environment should provide evidence that there is sufficient variance between classrooms. That is, beyond the variance attributable to student characteristics and school characteristics, there should be substantial variability between classrooms if one goal of an assessment is to differentiate between classrooms. In an effort to address the limitations associated with single-level analyses, Wheldall, Beaman, & Mok (1999) explored the variance among ICEQ scores using hierarchical linear modeling (HLM).

To examine the variability in ICEQ scores between classrooms, Wheldall et al. (1999) fit three models: a fully unconditional two-level model, an unconditional three-level model, and a two-level model conditional on grade level. In each model, a significant amount of variance was observed between classrooms, with the two-level conditional model fitting the data best. The three-level model showed varying effects of schools depending on which subscale of the ICEQ was in question (i.e., personalization, participation, independence, investigation, differentiation); however, none of the school level variance components were significant. The authors' findings demonstrate substantial variability in students' perceptions of the instructional environment across schools and provide further evidence for the utility of collecting such information.

Finally, students' perceptions of the classroom environment may have stronger predictive validity relative to other sources of information such as teachers' perceptions. That is, predictions using student perceptions of classroom teaching strategies are often better indicators of future performance relative to teacher perceptions of the same teaching strategies (Bernaus & Gardner, 2008). At the very least, students and teachers tend to differ in their reports of the classroom environment (Desimone, Smith, & Frisvold, 2010; Fraser & Fisher, 1983). Using 2000 NAEP data for 8th grade mathematics, Desimone et al. (2010) examined student and teacher report on the same seven questions about instruction. The questions asked students and teachers to report the frequency with which textbook problems, partner activities, measurement tools, writing problems, discussion activities, computers, and calculators were used in the classroom. The authors observed small but statistically significant differences between student and teacher responses for many of the questions, with the largest differences observed for the use of partner activities, discussion, and writing problems (these were reported to be used more frequently by teachers).

In Desimone et al.'s study, the differences between instruction as perceived by teachers and instruction as perceived by students were largest for students in lower achieving classes and students from low SES backgrounds (SES was measured by parental education). This incongruity between teacher and student perceptions of the classroom is particularly interesting when considering research highlighting teachers' tendency to underestimate the ability of students in low SES or low performing classrooms (Ready & Wright, 2010).

Finally, in a later study by Hughes (2011) comparing teacher and student perceptions of support and conflict in the classroom, mixed results were observed. More specifically, while teachers and students tended to be similar in their views of classroom conflict, the same was not true for support. In this example, it is important to note that many of the items assessing conflict were directly observable in the classroom. This contrasts from items measuring teacher support, which tended to focus on many unobservable factors. Thus, when the overtness of the factor in question was reduced, teacher and student perceptions were no longer related to one another.

**Dimensionality in Students' Perceptions of the Classroom Environment**

As might be clear at this point, the use of students' perceptions to explore the classroom environment is not particularly novel. The earliest explorations include work by Moos & Tricket (1974) and Fraser, Anderson, and Wahlberg (1982). Early work in students' perceptions resulted in the Learning Environment Inventory (LEI) and the Classroom Environment Scale (CES)—measures that are still in use today. Likewise, James Ysseldyke and Sandra Christenson at the University of Minnesota have long pointed out the pitfalls of overreliance on traditional observations and have thus incorporated student perceptions into several instruments that assess students' instructional environment (e.g., The Instructional Environment Scale [TIES] and the Functional Assessment of Academic Behavior [FAAB]). These specific measures, along with many others, are noted in Table 1 and are detailed later in this paper;

however, they are important to mention here as they demonstrate three early lines of research that called attention to students' perceptions of the classroom environment.

Despite the longstanding interest in students' perceptions of the classroom environment, it follows from Table 1 that all instruments are not created equal. As with any research project, researchers develop assessment tools with a particular measurement construct in mind. The classroom environment is a broad construct and the theoretical orientation of authors plays a critical role in the types of items that survive the development process. For example, the Constructivist Learning Environment Survey (CLES; Taylor, Fraser, & Fisher, 1997) is unlikely to include questions regarding teachers' use of rewards as this strategy is incompatible with the ideals underlying constructivist theory.

Table 1.

*Instruments Measuring Student Perceptions of the Classroom Environment*

| Instrument | Type | Level | Items | Item Format | Scales | Scales/Dimension Name | Reference |
|---|---|---|---|---|---|---|---|
| Functional Assessment of Academic Behavior (FAAB) | Multi Source and Method | K-12 | n/a | Multi-Method | 12 | Instructional Match, Instructional Presentation, Instructional Expectations, Relevant Practice, Adequate Feedback, Progress Monitoring, Cognitive Emphasis, Classroom Management, Adaptive Instruction, Curricular Match, Motivational Strategies, and Academic Engagement. | Ysseldyke & Christenson, 2002 |
| Learning Environment Inventory (LEI) | Student Report | Secondary | 91 | 4-point Rating Scale | 13 | Cohesiveness, friction, favoritism, cliqueness satisfaction, apathy, speed, difficulty, competitiveness, diversity, formality, and material environment | Fraser, Anderson, & Wahlberg, 1982 |
| Classroom Environment Scale (CES) | Student Report | Secondary | 90 | True/False | 9 | Involvement, Affiliation, Teacher Support, Task Orientation, Competition, Order and Organization, Rule Clarity, Teacher Control, Innovation | Moos & Trickett, 1974 |
| My Class Inventory (MCI; simplified version of LEI) | Student Report | Elementary | 38 | Yes/No | 5 | Cohesiveness, Friction, Satisfaction, Difficulty, and Competitiveness | Fraser, Anderson, & Wahlberg, 1982 |

| Instrument | Type | Level | Items | Item Format | Scales | Scales/Dimension Name | Reference |
|---|---|---|---|---|---|---|---|
| Individualized Classroom Questionnaire (ICEQ) | Student Report | Secondary | 50 | 5-point Rating Scale | 5 | Personalization, Participation, Independence, Investigation, Differentiation | Fraser, 1990 |
| Questionnaire on Teacher Interaction (QTI) | Student Report | All ages | 77 | 5-Point Rating Scale | 8 | Leadership, Helping/Friendly, Understanding, Student responsibility/freedom, Uncertain, Dissatisfied, Admonishing, Strict | Wubbels & Levy, 1993 |
| Constructivist Learning Environment Survey (CLES) | Student Report | Secondary | 30 | 5-Point Rating Scale | 5 | Personal Relevance, Uncertainty, Critical Voice Scale, Shared Control, Student Negotiation | Taylor, Fraser, & Fisher, 1997 |
| College and University Classroom Environment Inventory (CUCEI) | Student Report | College | 49 | 5-point Rating Scale | 8 | Personalization, Involvement, Student Cohesiveness, Satisfaction, Task Orientation, Innovation, Individualization | Fraser & Treagust, 1986 |
| What is Happening in this Classroom (WIHIC) | Student Report | Secondary | 56 | 5-point Rating Scale | 7 | Student Cohesiveness, Teacher Support, Involvement, Investigation, Task Orientation, Cooperation | Aldridge & Fraser, 2000 |
| Science Laboratory Environment Inventory (SLEI) | Student Report | Secondary and College | 35 | 5-point Rating Scale | 5 | Student Cohesiveness, Open-Endedness, Integration, Rule Clarity, Material Environment | Fraser, Giddings, & McRobbie (1995) |

| Instrument | Type | Level | Items | Item Format | Scales | Scales/Dimension Name | Reference |
|---|---|---|---|---|---|---|---|
| ClassMaps Survey (CMS) | Student Report | K-8 | 55 | 4-Point Rating Scale | 8 | Teacher-Student Relationship, Peer Friendships, Peer Conflict, Worries about Peer Aggression, Home-School Relationships | Doll, Champion, & Kurien, 2008 |
| Our Class and Its Work (OCIW) | Student Report | Grades 3-12 | 40 | 4-Point Rating Scale | 8 | Didactic Instruction, Enthusiasm, Feedback, Use of Instructional Time, Provision of Opportunity to Learn, Pacing, Structuring Components, Task Orientation | Waxman & Eash, 1982 |
| Rochester Assessment Package for Schools -- Student Self-Report (RAPS-S) | Student Report | n/a | 80 | Rating Scale | 5 | Student Rated Engagement, Perceived Teacher Context, Perceived Competence, Perceived Autonomy, Perceived Relatedness | Research Assessment Package For Schools, 1998 |
| The Classroom Life Instrument (CLI) | Student Report | Grades 5-9 | 59 | 5-Point Rating Scale | 12 | Cooperative learning, positive goal interdependence, resource interdependence, teacher academic support, teacher personal support, student academic support, student personal support, class cohesion, fairness of grading, achieving for social approval, academic self-esteem, alienation | Johnson, Johnson, & Anderson, 1983 |

| Instrument | Type | Level | Items | Item Format | Scales | Scales/Dimension Name | Reference |
|---|---|---|---|---|---|---|---|
| My Classroom Scale (MCI) | Student Report | Grades 6-9 | 10 | 5-Point Rating Scale | 2 | Satisfaction with Classroom Environment and Relationship with Teacher | Burnett, 2002 |
| Student Perceptions of Classroom Quality (SPOQ) | Student Report | High School | 38 | 5-Point Rating Scale | 5 | Meaningfulness, Challenge, Choice, Self-Efficacy, & Appeal | Gentry & Owen (2004) |

Although construct differences account for some of the variability in the available measures of the classroom environment, it is clear from Table 1 that the number of factors posited by the authors of conceptually similar classroom assessment instruments varies widely. For example, Fraser et al.'s (1982) LEI purports to measure thirteen distinct factors in the classroom environment while the CES (Moos & Tricket, 1974) purports to measure nine factors. From instrument to instrument, the number of scales varies.

As mentioned, some of these differences are related to the items that comprise the instruments themselves. In the case of the ClassMaps Survey (CMS; Doll, Champion, & Kurien, 2008), a strong case can be made for distinct factors—"Peer Aggression" is certainly different than "Home-School Relationships." Yet the wide differences in the number of final scales, especially when the scales cover factors related to the teacher, is certainly influenced by the analyses used to provide structural evidence for the instrument itself. That is, the dimensionality of these instruments is highly dependent on the statistical methods employed by the authors. As will be clear, much of the evidence provided by previous authors, while often demonstrating interdependence among items, falls short of demonstrating dimensionality.

Theories regarding the dimensionality of classroom environment instruments are perhaps most troublesome when internal consistency estimates (i.e., coefficient alpha) are the only evidence offered in support of ostensibly distinct scales. For example, in their adaptation of the QTI to elementary school settings, Goh & Fraser (1996) report coefficient alpha for eight scales, but omit critical evidence that would support the notion of scale independence. At first glance, high internal consistency within scales seems to offer good evidence of independent factors; however, it is important to note that while coefficient alpha provides an estimate of the interrelatedness of items (i.e., how related a group of items are to one another), it does not connote homogeneity of those items (i.e., evidence that items are only related to the hypothesized factor) (Netemeyer, et al., 2003). Consider a situation in which there are two factors: A and B. Factor A is composed of five items and the internal reliability estimate is .90. Factor B, while referencing five different items, has an identical estimate of internal reliability. In this scenario, it is appropriate to refer to the items in either Factor A or B as interrelated, but there is no evidence to suggest that items in Factor A are not *also* related to items in Factor B. That is, the five items in Factor A (or Factor B) are not homogenous. In the above situation, it is possible that Factor A and B are distinct factors, but it is also entirely possible that the two factors are indistinguishable from one another. Without the use of more appropriate measures of dimensionality (e.g., EFA or CFA), there is insufficient evidence for either claim.

The distinction between interrelatedness and homogeneity is an important one to make because confusion of the two can lead to misleading interpretations of subscale scores. As noted, although items within the same scale are often interrelated, they may also be related to items in other scales. Relationships of individual items (or groups of items) across scales are often evident in correlations between scales. If this is the case, it may be inappropriate to interpret changes in the composite score for any given scale to be indicative of a specific change to the construct of interest. In the literature on students' perceptions of the classroom environment, correlated scales are fairly common (Dorman, 2002; Fraser, 1998).

In their presentation of short forms for three classroom environment instruments, Fraser & Fisher (1983) report average correlations of the scales within each instrument. Although correlations varied by instrument and by the specific scale in question, the observed values were large enough (often higher than .30) to warrant reservations regarding multidimensionality; however, the authors only mention these correlations in one sentence, asserting that each instrument measures distinct "but somewhat overlapping" components of the classroom environment (pg. 10). In a later—and often cited—review of instruments measuring the classroom environment, Fraser (1998) reports correlations among the scales of different instruments but does not discuss the implications of large values.

Relative to coefficient alpha, exploratory factor analysis (EFA) is a more appropriate assessment when considering the factor structure of rating scales (Netemeyer et al., 2003). Many recent studies have used EFA as a tool for providing structural evidence for a particular measure of students' perceptions (e.g., Doll et al., 2010; Johnson & McClure, 2004). In some cases, this allows authors to highlight the integrity of the hypothesized factor structure. Evidence for multidimensionality is most evident when there are large item loadings on one factor, but not others. For example, with a few exceptions, Doll et al. (2010) demonstrate that items intended to measure a particular construct (e.g., teacher variables) are not related to items that load heavily on other factors (e.g., student worry). Nevertheless, it is fairly common to observe items that relate to more than one construct (Johnson & McClure, 2004). It is also worth noting that it is sometimes difficult to discern the magnitude of cross loadings given that many authors omit this information.

A large majority of the early research on students' perceptions of the classroom environment has relied on estimates of internal consistency and EFA; however, others have employed the use of CFA, which allows for stronger claims regarding dimensionality. A recent analysis of the CES by Boren, Callahan, & Peugh (2010) serves as an excellent illustration of the potential advantages associated with CFA. Boren and her colleagues collected new data and used CFA to test the hypothesized factor structure of the CES. A brief reference to Table 1 shows that the CES is comprised of nine factors; however, the authors posited that a more parsimonious model would fit the data better. CFA was used to test the nine factor solution and every reported fit index suggested bad model fit. Next, treating each of the 90 items as independent, EFA was used on half of the data to gauge the possibility of a simpler model. After eliminating 33 items, the authors relied on evidence from the EFA to test a three factor solution on the second half of the data. As predicted, the fit indices for the three factor solution were better than the original nine factor solution, albeit short of objectively acceptable values.

If nothing else, the results presented by Boren et al. (2010) highlight the potential for alternative perspectives on the dimensionality of tools used to measure students' perceptions of the classroom environment. This is particularly true when the content that is covered by the assessment is highly related. While one might expect students' perceptions of clearly distinct factors to differ (e.g., perceptions of home environment vs. classroom environment), it is a far longer stretch to assume students' perceptions of different instructional factors are entirely independent of one another. While some researchers allow factors to correlate, thereby recognizing some level of interdependence, the correlations are rarely discussed in depth and the factors are often interpreted as distinct from one another. Boren et al.'s reconceptualization of the CES marks an important step in classroom environment research, but even in the three factor solution presented by the authors, large correlations were observed between factors. It may prove fruitful for researchers to consider these correlations more closely when conceptualizing the structure of instruments measuring students' perceptions of the classroom environment.

Gentry & Owen (2004), although rejecting the notion of a second-order construct, do well by explicitly discussing correlations among apparently distinct constructs of the classroom environment. In their 2004 article, the authors discuss an instrument (Student Perceptions of Classroom Quality [SPOCQ]), intended to capture "classroom climate" and outline five constructs: meaningfulness, challenge, choice, self-efficacy, and appeal. The authors present the results of a CFA in which groups of items were parceled to form composite scores. These scores served as the manifest variables and were hypothesized to load onto one of the five constructs listed previously. The authors allowed the five latent constructs to correlate with one another. The result of this factor structure—five distinct, but related constructs—is not uncommon, but unlike previous work, the authors draw attention to the concept of interdependence as it relates to the classroom environment.

Despite offering an interesting look at the dimensionality of students' perceptions of classroom climate, there are potential measurement flaws in Gentry &

Owen's structural model. Perhaps most important, is the use of item parcels. Due in part to the parsimonious nature of structural models that use parcels, the aggregation of items is more likely to lead to replicable results (Little, Cunningham, Shahar, & Widaman, 2002); however, item parceling is often viewed as a problematic approach to representing the dimensionality of a set of items (Bandalos & Finney, 2001). That is, if observed data are parceled, the model no longer represents individual responses.

Ryan & Patrick's (2001) study is another example of research exploring the presence of second-order classroom factor. In this case, the authors observed a strong relationship between four first order factors and one higher order factor deemed the "classroom social environment." Although the second-order factor structure was not the primary focus of the study, the authors' observations provide some evidence that when several related factors are under study, it may be more appropriate to consider the possibility of an overarching construct rather than simply allowing factors to correlate while withholding interpretation of the observed correlations.

Despite the intuitive appeal of a higher order construct, statistical hypothesis testing of the relationship between lower order factors and a higher order construct is rare in classroom environment research.

**Specific Indirect Measurement Tools for the Classroom Environment**

Although many measurement tools have been discussed in depth and the list of available instruments in Table 1 has been previously referenced, it is worthwhile to provide additional information on some of the more prominent measures of the classroom environment. In accordance with the broad nature in which the classroom environment can be conceptualized, there are a variety of tools available to assess the classroom environment in both a specific and comprehensive manner. In the most recent *Handbook of School Psychology*, Gettinger et al. (2011) provide an overview of commonly used assessment tools, many of which are multidimensional and well documented in the educational literature. The assessment tools listed below include many of those highlighted by Gettinger and her colleagues; however, many others are included as well[1]. Table 1 is not an exhaustive list, but includes a wide variety of tools used in previous research. Additional information for a select group of these instruments is provided below.

All of the assessments listed below (and in Table 1) are multidimensional; however, they differ from one another in regard to which components of the classroom environment are of the most interest, the degree to which empirical support is provided for their dimensional structure, and their use in schools. Each assessment may address a different number of components and many of these components are substantively different. In some cases, such as the ICEQ, the difference is explicitly stated. In other cases, the discrepancies are more nuanced.

**Learning Environment Inventory (LEI).** The LEI is designed for use with high school students and is one of the earliest measures of the classroom environment (Anderson & Wahlberg, 1968). The LEI is comprised of 15 dimensions of the classroom environment with seven items factoring into each dimension. The 15 dimensions include Cohesiveness, Diversity, Formality, Speed, Material Environment, Friction, Goal Direction, Favoritism, Difficulty, Apathy, Democracy, Cliqueness, Satisfaction, Disorganization, and Competiveness. Each item is phrased so that students respond along a four-option continuum of agreement (Strongly Disagree-Strongly Agree). Answers may be considered at the individual or classroom level.

**Classroom Environment Scale (CES).** Like the LEI, the CES (Moos & Tricket, 1974; Tricket & Moos, 1973) marks one of the earliest attempts to capture multiple dimensions of the classroom environment. The CES is comprised of 90 true-false items organized into nine dimensions including involvement, Affiliation, Teacher Support, Task Orientation, Competition, Order and Organization, Rule Clarity, Teacher Control, and Innovation. The primary target population includes high

---

[1] Although not included in the overview below, the tool used for the current study is outlined in detail in the Methods chapter.

school students. Research on the CES has contributed to the empirical foundation for claims regarding the association between student perceptions of the classroom environment and numerous relevant outcomes (e.g., achievement, motivation, grades, etc) (Fisher & Fraser, 1983; McMahon & Wernsman, 2009; Myers & Fouts, 1992)

**Individualized Classroom Environment Questionnaire (ICEQ).** The ICEQ (Wheldall, Beaman, &  Mok, 1999) differs from other assessments of the classroom environment in its explicit focus on the "individualized" nature of environments. Five scales are defined (Personalization, Participation, Independence, Investigation, and Differentiation), each with 10 items. In addition to information on student perceptions of the classroom environment, the ICEQ also asks student respond to the same items using an "ideal" environment as a reference.

**What is Happening in this Class Questionnaire (WIHIC).** The WIHIC (Fraser, Fisher, & McRobbie, 1996) consists of seven scales: Student Cohesiveness, Teacher Support, Involvement, Investigation, Task Orientation, Cooperation, and Equity.  Each scale consists of eight items and the primary population is restricted to high school students. Nevertheless, the WIHIC is currently used in several settings across multiple countries (Gettinger et al. 2011).

**The ClassMaps Survey.** The most recent ClassMaps survey (Doll et al. 2009) consists of 47 items factoring into seven (in some cases, eight) subscales. The authors describe four of these subscales as relating to relational aspects of the classroom environment, these include My Teacher, My Classmates, Kids in this Class, and Talking with My Parents. The remaining three subscales are designed to measure self-regulation and include Believing in Me, Taking Charge, and Following Class Rules. In some cases, an eighth scale is included that assesses bullying in the classroom. The wording for items is similar within each subscale. For example, items in the Kids in this Class category all begin with "kids in this class." Across all subscales, students answer using a 4-point rating scale assessing frequency (Never, Sometimes, Often, Almost Always).

**Functional Assessment of Academic Behavior (FAAB).** Using Bronfenbrenner's (1979) definition of the instructional environment, FAAB (Ysseldyke & Christenson, 2002) incorporates a multi-source and multi-method approach to assess three broad supporting areas that relate to a student's "total learning environment": instructional support, home support, and home-school support. Due to its comprehensive approach to assessment, FAAB is recommended for individual use rather than as a class-wide assessment tool. The instructional support domain of FAAB is comprised of 12 components: Instructional Match, Instructional Presentation, Instructional Expectations, Relevant Practice, Adequate Feedback, Progress Monitoring, Cognitive Emphasis, Classroom Management, Adaptive Instruction, Curricular Match, Motivational Strategies, and Academic Engagement.

**Student Feedback as a Catalyst for Instructional Change**

While the existing research calls to the development and predictive value of student perceptions of the classroom environment, very few researchers have addressed the utility of using student perceptions as means for instructional change. Currently, teachers differ widely in their sense of preparedness for classroom teaching (Darling-Hammond, Chung, & Frelow, 2002), and rarely receive adequate feedback or support after entering the classroom (Westling, 2010). Further, teachers receiving little support in the classroom are less likely to develop a sense of self-efficacy (Hoy & Spero, 2005), which in turn, may decrease instructional experimentation in the classroom (Cousins & Walker, 2000). Thus, efforts to increase and improve the feedback teachers receive are critical for improving the instructional experience of students and teachers alike (Hoy & Spero, 2005; Freiberg & Waxman, 1988; Ross, 1992).

In their summary of current feedback practices for teachers, Scheeler et al. (2004) include 10 experimental and quasi-experimental studies in which a directly observable teaching behavior was identified as the dependent variable. The paucity of well-designed research on teacher feedback as well as discrepant conclusions among the available research resulted in few strong recommendations; however, strong support was observed for immediate feedback. In addition, studies including positive, specific, and corrective feedback also produced promising results. Unfortunately, Scheeler and colleagues restricted their focus to relatively intensive feedback, and research using student-report data was ignored entirely. Thus, more work is needed to (a) develop research-based items that provide specific and actionable feedback to teachers, (b) assess these new items using empirically rigorous analytic strategies (e.g., confirmatory factor analysis, SEM, HLM), and (c) explore the utility of these items in changing the way teachers develop and maintain classroom environments

The literature on using student perceptions to facilitate improvements in the classroom environment is sparse. Nevertheless, promising examples exist (Thorpe, Burden, & Fraser, 1993; Fraser et al., 1982). In one small-scale case study, Fraser et al. (1982) used student perceptions of the classroom environment in consultation meetings to develop a teacher improvement plan. In this particular case, students were asked to rate several dimensions of the classroom environment on the ICEQ in regard to “actual” and “ideal” conditions. The two components with the largest discrepancy between actual and ideal ratings were identified as areas for improvement and strategies for improving student perceptions were discussed among the researchers and the teacher. After implementing the identified strategies, a significant reduction in the discrepancy between perceived and ideal classroom conditions was observed for the domains identified for intervention. Despite the methodological disadvantages associated with convenience sampling and a non-experimental design (Campbell & Stanley, 1963), Fraser et al. (1982) provide preliminary evidence for the utility of using student perceptions of the instructional environment as an impetus for instructional change.

Subsequent work using a similar methodology supports the argument for using student perceptions of the environment to inform intervention, but ambiguous or resource-intensive approaches to intervention and a lack of experimental control consistently undermine research in this area (Blose & Fisher, 2003; Thorp, 1994;

Waldrip, Reene, Fisher, Dorman, 2008; Yarrow, Millwater, & Fraser, 1997). Nevertheless, previous approaches to using student perceptions of the classroom environment as tools for feedback provide a procedural framework for intervention. This procedure typically includes 5 components: an initial assessment, provision of feedback to teachers, reflection and discussion, intervention, and reassessment (Fraser, 1999). Thus, additional work using student perceptions of the environment—particularly within an experimental design—may benefit from the procedural framework currently available. Further, the way in which this feedback is presented to teachers may depend largely on the prior experience and background knowledge of teachers (Helman, 2006).

The extant literature on teacher-student feedback (Black & William, 1998; Hattie & Timperly 2007; Shute, 2008) may also serve as a template for future approaches to providing feedback to teachers on student-reports of the classroom environment. Despite the obvious differences between teachers and students, many of the feedback guidelines for students could be expected to work equally well when applied to teachers (e.g., specific, goal-oriented feedback). Additional research aligning research on effective feedback with the literature base on quality instructional environments is warranted.

**Moving Forward**

Most researchers agree that academic and behavioral outcomes among students are partially influenced by teachers. Accordingly, a great deal of research seeks to identify specific components of classroom environments associated with positive student outcomes. As a natural byproduct of such inquiries, the ways in which information can be collected on relevant instructional variables is well established.

Although research on which instructional variables matter and how they should be measured is necessary for educational change, it is not sufficient. Given the abundance of research on effective instructional strategies, as well as ways in which information can be collected on those strategies, research addressing the proper use of assessment information is critical. Guidelines for effective consultation and intervention implementation may serve as useful references for how student ratings can best be used to guide professional development. More specifically, this might include decreasing the complexity and jargon of assessment information, increasing the usability of information, allowing teachers to take an active role in the feedback process, and sustained support (Forman, & Burke, 2008; Supovitz & Turner, 2000; Taylor et al., 2004). Further, professional development research stands to benefit from research questions that address (a) organic implementation procedures and (b) how much professional development is necessary to observe changes in student outcomes (Wayne, Yoon, Zhu, Cronen, & Garet, 2008).

Few researchers have addressed how student-ratings of the instructional environment can be used as catalysts for instructional change. Even fewer projects have done so while guarding against potential threats to internal and external validity. Regardless of the observed effect, such studies are practically relevant and will significantly contribute to the empirical literature on student-teacher feedback. In an applied sense, educators are in need of research identifying ways in which assessments of the classroom environment can be used to improve instruction for students. Studies offering empirical insight into these questions will assist educators in determining how classroom assessments can be included as a tool for professional development.

# Chapter 3

# Methods

The following subsections provide detailed information on the materials, procedure, participants, and analyses used for the current study. To provide sufficient background for the procedures and analyses, information on participants and measures is presented first. The description for the REACT in the materials section is particularly detailed given the central role of this instrument in the current study. Following a description of the measures, the procedures—from participant recruitment to final data entry— are described in detail. Finally, the analyses used to address each research question are discussed.

## Participants

**Students.** Seven hundred ninety seven students from three public middle schools participated in the current study. All three schools were suburbs of the Minneapolis metro area and were located in two different districts. Students were nearly evenly distributed between District A (46.7%) and District B (53.3%). Two schools from District B participated in the study. Demographic information for the current study, including gender and ethnicity, are displayed in Table 2. The distribution of male (50.4%) and female (49.3%) participants was nearly equivalent. A larger percentage of students participating in the current study were White (83.3%) and a lower percentage of students were minorities relative to the Minnesota state average (NCES, 2011). A small percentage (4.9%) of students selected “other” when indicating ethnicity. All three schools included in the study served students in grades six through eight. The distribution of participating students was such that more eighth graders were included (43.6%) than sixth (29.9) or seventh (26.5) graders.

Table 2.

*Student and Teacher Demographics Aggregated Across Group Assignment.*

| Variable (n) | Percent |
|---|---|
| **Students (797)** | |
| Gender (795) | |
| Male | 50.4 |
| Female | 49.3 |
| Ethnicity (797) | |
| Hispanic/Latino | 3.4 |
| American Indian or Alaskan Native | 0.3 |
| Asian | 6 |
| Black or African-American | 1.9 |
| Native Hawaiian or Pacific Islander | 0.3 |
| White | 83.3 |
| Other | 4.9 |
| Grade Level | |
| Six | 29.9 |
| Seven | 26.5 |
| Eight | 43.6 |
| **Teachers (30**) | |
| Gender (30) | |
| Male | 30 |
| Female | 70 |
| Experience (30) | |
| 1-3 Years | 23.3 |
| 3+ Years | 76.7 |

**Teachers.** Thirty one teachers participated in the study (see Table 2 for descriptive statistics). All teachers were White and a larger percentage of teachers were female (70%) than male (30%). As mentioned, roughly half (46.7%) of teachers were employed in District A, which included one participating school. The remaining teachers were employed in one of two schools in District B. Teachers varied in experience, with 23.3% reporting one to three years of teaching experience and 76.7% reporting four or more years. Teachers were responsible for a wide variety of content areas, including English (33.3%), Math (10%), Science (16.7%), Social Studies (26.7%), Language (10%) and Special Education-Reading (3.3%).

## Materials

**The Responsive Environmental Assessment for Classroom Teaching (REACT).** Given the central importance of the REACT for the current study, the following section provides an overview of the intended measurement construct, the process by which items on the current form were selected, and a brief discussion of the available evidence for its use as a measure of the classroom environment.

The REACT is a 30-item survey that captures students' perceptions of the classroom teaching environment in a manner intended to be useful for classroom teachers. Students respond to statements by selecting one of four answer options: Yes, Mostly Yes, Mostly No, or No. While it is an assessment of the classroom teaching environment, the REACT is intended to be used formatively. In other words, the REACT is not intended to evaluate teacher performance per say—it is intended to guide teachers in their efforts to improve instruction. In the most basic sense, the REACT allows teachers to (a) gather information on how students perceive the learning environment, and (b) respond to those perceptions. While previous work on the REACT has focused extensively on item development and factor structure, the version used for the current study connects the items to several research-based strategies for improvement. This addition is in line with the intent of the current study.

Available data on the REACT are limited, but it can be used at the classroom, small group, or individual level. If administered at the classroom level, teachers are encouraged to keep names anonymous to promote student honesty. After collecting data with the REACT, teachers can view the distribution of scores for each item. As mentioned, there are also instructional strategies provided along with the REACT to facilitate problem solving. Students' responses on the REACT should help teachers decide which strategies are most likely to improve students' perceptions in the future.

Note that the REACT is not a compendium of effective instructional strategies. It is beyond the scope of the tool to provide a comprehensive list of research-based strategies for instruction. The strategies included in REACT feedback are intended to serve as a preliminary reference for teachers—the primary catalyst for action is the summary of students' responses. Nevertheless, the general strategies provided on the feedback form are grounded in research and should serve as an initial guide for teachers. For example, if a large number of students answered "no" to the item "my teacher tells me when I do a good job," it is appropriate for the teacher to consider increasing the amount of targeted verbal praise. After consistently implementing the chosen strategies over a period of several weeks, teachers may ask students to complete the REACT once again to gauge improvement. In practice, the REACT is intended to be used by the teacher throughout the school year as a formative measure of the classroom teaching environment.

*Construct definition.* The classroom teaching environment is a broad construct. As reviewed, many classroom measures focus intently on a specific aspect of the classroom environment (e.g., use of achievement goals, student autonomy, classroom management, use of technology, etc.). Relative to these instruments, the REACT is an omnibus measure of the classroom teaching environment. Nevertheless, there are aspects of the "classroom environment" not captured by the 30 items that comprise the instrument. For example, the REACT does not explicitly address peer relationships or the physical classroom layout. Instead, many of the items focus intently on the behavioral and instructional supports provided by the teacher.

The initial measurement model was designed to align with Ysseldyke and Christenson's (2002) 12-part conceptualization of "instructional support" outlined in FAAB. Given the intentions of the instrument, Ysseldyke and Christenson's framework was a suitable starting point; however, additional review of the measurement implications associated with FAAB definitions resulted in 11 components, which are available in Appendix A. These 11 components served as the blueprint for the REACT's measurement construct.

*Item development.* The 30 items included on the REACT were selected through an iterative process in which items were developed, field tested, and refined according to psychometric and substantive criteria. As might be expected, the process for inclusion became more stringent after pilot data were available. Item development for the REACT followed three distinctive phases: initial item writing, field testing, and final review. These phases are outlined below. Note that throughout the process, items were evaluated according to the degree to which they (a) tapped into malleable teacher behaviors and (b) were supported by empirical research.

*Initial item writing.* A majority of the REACT items were developed through a series of research meetings in which items were matched to the revised categories for “instructional support” provided by Ysseldyke & Christenson (2002). In addition, all items contained words appropriate for students in grades 6-9. Each item was evaluated by according to its (a) relation to the overarching construct, (b) relation to the 11 sub-categories in Appendix A, (c) appropriateness for students in grades 6-9, and (d) connection to malleable teacher behaviors. All developed items were presented to a group of graduate students in the Educational Psychology department at the University of Minnesota. In addition, a group of teachers were asked to evaluate the items in a similar fashion. Given the importance of face validity in classrooms, the degree to which the items adequately measured the classroom environment was perhaps the most important component of teacher review. Several items were refined and added to the REACT following the focus group and teacher review. The 100+ items were once again reviewed by the research team according to the aforementioned criteria, resulting in an assessment comprised of 100 total items connected to 11 ostensibly distinct categories.

*Field-Testing.* There were two primary questions to be addressed during field-testing. First, despite Ysseldyke & Christenson’s (2002) compelling substantive argument for multidimensionality in the “instructional support” domain, no factor analytic data were available. Thus, one goal of field-testing included an evaluation of the hypothesized factor structure of the REACT. The dimensionality of the REACT is discussed more extensively in the following section, but it is important to note that original hypotheses regarding the measure’s underlying factor structure were not supported. Instead, exploratory data analyses supported the presence of a single factor, with some items loading more heavily than others. The factor loadings from the first data collection served as a reference for retaining or rejecting items (high factor loadings [>.50] were preferred over low factor loadings).

In addition to the factor loadings, non-parametric item analyses were used to create option characteristic curves (OCC) (see Figure 1). OCC plots allow students' expected total score (expressed as $\theta$) to be plotted along with the probability of selecting the four possible answer options (yes, mostly yes, mostly no, and no). The expected $\theta$ is a representation of the predicted perception of the classroom environment overall. Thus, as the predicted $\theta$ increases (in this case, indicating an increasingly negative view of the classroom environment), the probability of responding "no" for any given item should also increase.

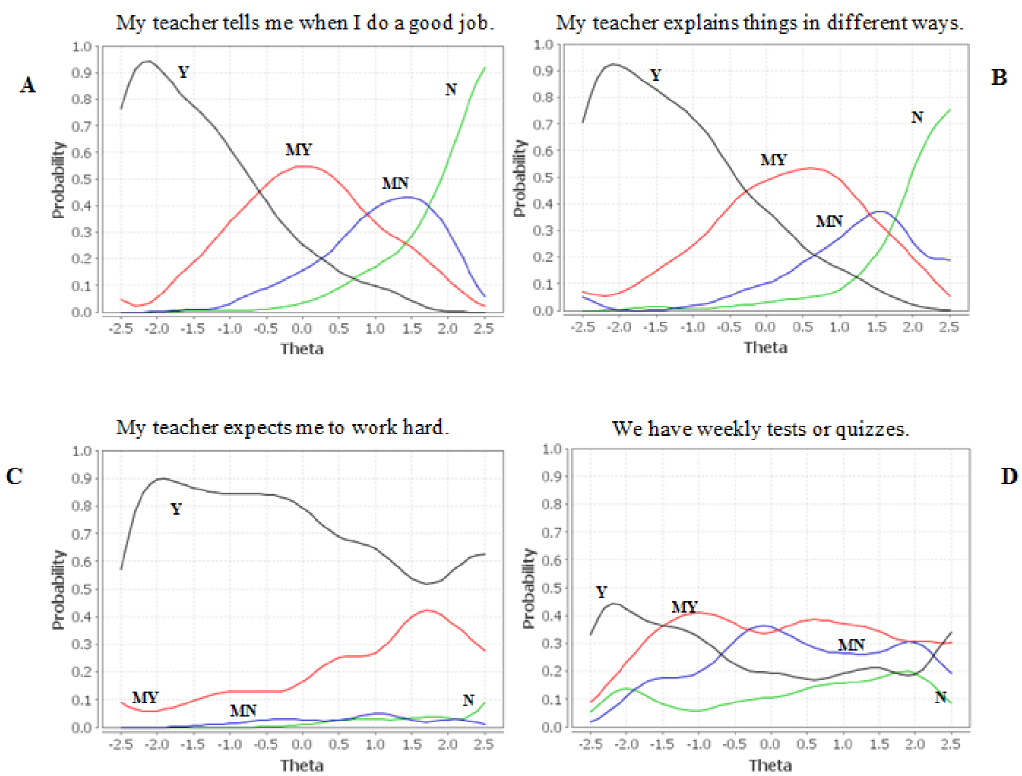



*Figure 1.*

Select Option Characteristic Curves (OCC) from 100 piloted REACT items (Y = Yes, MY = Mostly Yes, MN = Mostly No, N = No)

OCC curves for each item were visually analyzed by members of the research team for (a) adequate representation of each answer option and (b) correct ordering of each answer option. For example, for any given item, a high $\theta$ should not be associated with a high probability for the answer option “yes.” Four examples of OCC curves are provided in Figure 1. Items in panels A (“my teacher tells me when I do a good job”) and B (“my teacher explains things in different ways”) were retained for final review. Note that all answer options are represented and the probability of each answer option is aligned with $\theta$. The items in panel C and D were eliminated. The item in panel C, “my teacher expects me to work hard,” is an example of a situation in which most students responded “yes” regardless of their overall orientation toward the classroom environment. Equally inappropriate, are the responses for the item “we have weekly tests or quizzes” in panel D. In this case, the probabilities are similar across all $\theta$ values, suggesting that this particular item will not be useful in practice.

Last, it is important to note that four new items were piloted during the last data collection. These items were added for field-testing in an effort to measure the degree to which teachers emphasize formative assessment and mastery learning in the classroom. Two of the four new items were retained for final review.

*Final Review.* Similar to the first phase of item development, the final review involved a more substantive approach to assessment. First, the degree to which items could be expected to provide actionable information to teachers was carefully evaluated in conjunction with field-testing data and preference was given to unambiguous items. In cases where two similar items were retained throughout the item review process (e.g., “my teacher explains things in different ways” and “my teacher tries to explain things in different ways”), these items were carefully evaluated for wording and the potential to instigate teacher action until a retention decision could be made. Nevertheless, four broad items (e.g., “I like this class”) were retained in light of favorable item characteristics and their potential use as indicators of change and broad perceptions of the classroom environment. Items were also evaluated (as a whole) for range in content. Given that the REACT purports to measure several aspects of the classroom teaching environment, including items across a wide range of content was critical. Following the final review, all 11 of the original components (see Appendix A) were represented across 30 items. The final items and corresponding components are available in Appendix B. Note that items in Appendix B are separated by the strategies teachers could be expected to employ given low scores in that particular area.

**REACT Dimensionality.** As mentioned, the current project began with the assumption that the 30 items which comprise the REACT are unified by one construct. The argument for unidimensionality is rooted in preliminary analyses with pilot data for 1,510 students collected over the course of two years. Original hypotheses for the measure's underlying factor structure were based on a diverse range of content across items; however, these substantive hypotheses were not supported. Instead, early data analyses supported the presence of one large factor, with some items loading on this factor more heavily than others. Following the item editing process outlined above, the factor structure of the 30 proposed REACT items was evaluated using confirmatory factor analysis (CFA). Using all available student data, the presence of a unified construct (i.e., "classroom environment") was supported by multiple indices; however, there were several items with correlations unexplained by the unidimensional construct and four error variances were allowed to correlate. Thus, at the time of this project, the available evidence for the factor structure of the REACT pointed toward a single construct.

Interpretation of unidimensionality, or even large correlations between classroom factors, is not intuitive at first. In part, this is because previous research has taken special care to present aspects of the classroom environment as independent constructs; however, the observed interdependence should not be surprising. In fact, if different components of the classroom environment (e.g., use of praise, instructional match, clear directions) are considered as a community of actions, interdependence is predictable. Consider the term interdependence in its most common context—biology. For many, it is tempting to view the organisms that comprise a given community as independent. From this interpretation, it follows that efforts to improve or degrade the quality of one component would only focus on factors that directly affect that species; however, most biologists would find this to be too restricting. In reality, the quality of one component depends on direct effects that are often easily identifiable, but it also depends on countless indirect factors. To consider any one component as independent, then, is inherently flawed.

Perhaps attempts to partition aspects of the classroom teaching environment into independent categories are equally flawed. For example, if a teacher were to increase the use of clear directions, students in the classroom would likely report an increase in the clarity of directions; however, the use of clear directions might also create more time for the teacher to provide feedback and praise on student work, changing perceptions of those factors as well. Thus, to consider the clarity of directions as independent from other components of the classroom teaching environment may represent a limited view.

To summarize, the dimensionality argument presented here does not imply that the classroom environment is overly simplistic. Quite the contrary, the argument is that the classroom environment is remarkably complex, with changes to one substantively unique component producing smaller changes in many other components of the environment.

**REACT summary.** Following data collection, a classroom summary is generated for teachers. The REACT summary provides teachers with a platform for interpretation of REACT data. There are two primary components of the REACT summary: student data and evidence-based strategies. The frequencies for each item option are arranged in 30 color-coded bar plots (one plot per item). More specifically, items in which students respond most often with “no” or “mostly no” are identifiable by more darkness in the bar plot, while items for which students respond most often with “yes” or “mostly yes” are identifiable by more white area in the bar plot. The horizontal arrangement of item plots allows teachers to scan items for areas of improvement (see Appendix C).

Once teachers have identified which items are lower than expected, there are six groups of evidence-based strategies available for reference. Despite the unidimensional nature of the REACT, items are separated into seven strategy groups to facilitate problem solving (see Appendix D for an example of one strategy group).

**Teacher survey.** After receiving the classroom summary, teachers complete a short survey that explores the clarity and usefulness of the REACT for classroom practice (see Appendix D for items). Survey items were developed solely for the purpose of the current study, as no existing tools were directly applicable. Guidelines for survey and assessment item writing were used to develop survey items (e.g., Haladyna, Downing, & Rodriguez, 2002; Rodriguez, in press). In two cases, items originally designed for the Intervention Rating Profile (IRP; Witt & Martens, 1983) were adapted to align with the current project. All items conform to a four-point scale ranging from "strongly agree" to "strongly disagree".

**Procedures**

Before providing substantial detail on study procedures, a brief overview of the study design is needed for clarity. Although two of the primary research questions for the current study can be answered descriptively, a pre- post-test design was used to address the impact of the REACT feedback on student perceptions of the classroom environment. Thus, the primary dependent variable was students' overall REACT score and the primary independent variable of interest was group assignment. Teachers in the experimental group were provided access to student feedback immediately and teachers in the control group did not have access to students' ratings until after the second data collection. REACT data were collected at three time points (Time 1, Time 2, and Time 3) with approximately three weeks between each data collection. Note that teachers who were initially assigned to the control group received feedback following the *second* data collection. Although no differences were predicted between the control and experimental group at Time 3, this period of data collection provides a proper outcome for teachers in the control group and the opportunity to assess maintenance for the experimental group (refer to Figure 2 for a visual representation of the study design). The following subsections provide details for all procedures used before, during, and after the study. Descriptions are arranged sequentially.

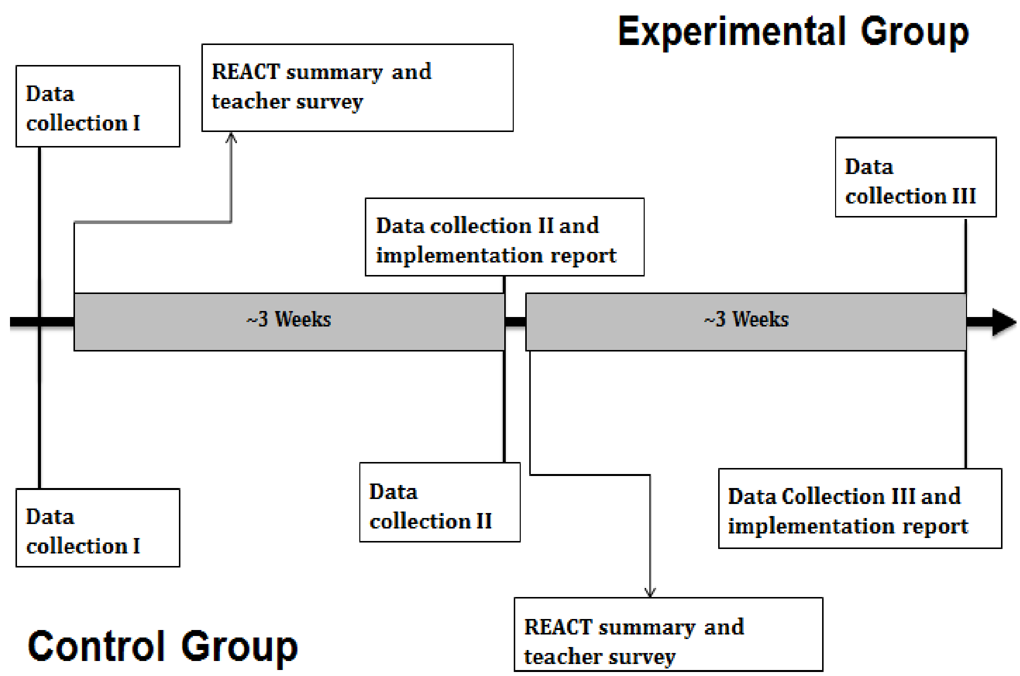



*Figure 2.* Overview of project design.

**Teacher recruitment.** Teachers in District A and B were recruited in an identical manner. Following superintendent and principal approval, a meeting was scheduled in which the researcher provided a 20-minute description of the study to all teachers. The meeting included a description of the REACT, the purpose of the study, logistical requirements for participation, and a discussion of incentives (each teacher received a twenty-dollar gift card at the completion of the study). Teachers were also informed that specialist classes, such as physical education, art, or music were ineligible for participation. At the conclusion of the meeting, interested teachers were asked to sign up using their name and email address. A total of 33 teachers expressed interest in the study. Two of the teachers were excluded from participating in the study due to the inappropriate nature of their content area (one teacher was a band director and the other oversaw a small group of student leaders).

**Group assignment and selection of class periods.** Following teacher recruitment, the 31 participating teachers were randomly assigned to one of two groups using the statistical program *R* (R Core Team, 2012). Random assignment was preferred over alternative methods (i.e., teacher preference) as it protects against many threats to internal validity. 16 teachers were assigned to the experimental group (described in study design section) and 15 teachers were assigned to the control group.

Because the current study focused on the middle school environment, there were a number of logistical obstacles. Relative to elementary schools, students in middle schools move between classrooms much more often, which results in students interacting with multiple teachers and teachers interacting with a very large number of students. Unfortunately, this system of operation makes controlled research somewhat complicated. First, every teacher had several class periods that met criteria for inclusion. Including every possible class would have far exceeded the practical constraints for data collection. Thus, only one class period per teacher was selected for participation. Before discussing how these class periods were chosen, it is necessary to introduce the second large obstacle for data collection—overlapping students. In this case, students who were enrolled in multiple classrooms selected for participation are considered overlapping students. Ideally, each student included in the study would only complete the REACT once; however, if the data collection schedule was sufficiently disorganized, a student could theoretically complete the REACT in every period of the day. Thus, class periods were selected to reduce the number of overlapping students. That is, classes with little to no overlap in students were preferred over others.

**Data collection.** There were three total data collections that occurred throughout the fall semester. These administrations will be referred to as Time 1, Time 2, and Time 3. Due to concerns regarding student confidentiality, the REACT was administered by the researcher or a graduate level school psychology student. The number of additional data collectors varied from six to seven and the total time spent in each classroom per data collection was approximately 15 minutes. Before beginning, students received instructions on the format of the REACT and were told that a researcher from the University of Minnesota was interested in learning more about students' perceptions of the classroom environment. In addition, students were informed that their responses were anonymous.

Each student received a plastic sleeve with the REACT, a response form, a passive consent form, and a blank piece of paper. The response form and the blank paper were also labeled with an identification number. Students were asked to record their initials and a number on the blank paper. These markings would serve as a reference for locating materials in future data collections. After completing the REACT, students returned the materials to the plastic sleeve. During the administration, teachers were in the room and discouraged from viewing the questions; however, there were two exceptions. Prior to administration at Time 1, all teachers were asked if any students would require accommodations. Two teachers indicated that several students would need the questions read aloud, resulting in a decision to read the REACT for all students in these classrooms. Both teachers were assigned to the experimental group, making the change in protocol far less problematic (all teachers in the experimental group saw the questions shortly after the first administration).

As mentioned, data collections were evenly spaced by three weeks. Between data collections, response sheets were removed from the plastic sleeves while the form with students' initials was left in place. During the second data collection, students obtained the appropriate response form by locating the plastic sleeve with the correct initials. A shortened version of the directions was read aloud to students, although special emphasis was still placed on student anonymity. The third data collection was identical to the second.

**Feedback meetings.** Two days after the first data collection, teachers assigned to the experimental group met with the researcher before the school day began. If a teacher was unable to attend the feedback meeting, an individual meeting was arranged during the teacher's planning period.

All teachers were presented with a brief presentation that covered the format and interpretation of the REACT summary. Following the presentation, teachers were provided with the REACT summary for their classroom. As mentioned, the summary form provided color-coded frequency distributions for each item. Items were partitioned into substantively meaningful categories (see Appendix C). Teachers were asked to create an improvement plan by selecting as little or as many categories as desired for improvement. For example, if a large number of students responded "mostly no" or "no" to items in the "positive reinforcement" category, a teacher might have identified that category for improvement. For each category, teachers were encouraged to view a short list of research-based strategies that would be expected to improve students' responses on items associated with that category. Next, teachers used a short form to indicate (a) the strategy group(s) intended for implementation, and (b) the degree to which they intended to use the strategies selected for implementation. The latter question was answered using a four point rating scale (Yes, Mostly Yes, Mostly No, and No). Finally, teachers were asked to complete a brief survey about the REACT (see Appendix D).

Before the conclusion of the feedback meeting, teachers were encouraged by the researcher to inform students that their responses had been seen (as a class) and that there was a plan in place to address students' concerns. Teachers were specifically advised not to inform students which strategy group was being targeted. Teachers were also instructed not to share their classroom summary or the content discussed in the feedback meeting with other teachers, students, or administrators.

Following the second data collection, teachers assigned to the experimental group were contacted and asked to report the degree to which they implemented the strategies selected at Time 1 (“Yes”, “Mostly Yes”, “Mostly No”, or “No”). This question served as a measure of teachers’ reported implementation integrity. Note that teachers in the experimental group did not receive any feedback beyond what was provided after the first data collection. All teachers assigned to the control group participated in an identical feedback meeting following the second data collection.

**Follow up interviews**. Three teachers were randomly selected to participate in a brief interview about their experience with the study as well as the REACT more generally. All interviews were semi-structured and completed individually. One meeting occurred in person and two other interviews were conducted remotely. In addition to several structured questions about REACT questions, strategies, and administration procedures, teachers were encouraged to provide more general feedback about their experience.

**Data entry and coding errors.** During each of the three phases of data collection, all data were collected, entered, and compiled for teacher feedback within a span of one week. Given the expedited nature of this schedule, automated methods of data entry (e.g., the use of a scanning machine) were impractical. All variables were coded numerically and data were entered manually by the lead researcher and volunteers. Following the last data entry period, ten percent of the data were randomly selected and checked for accuracy. During this process, an inconsistency in coding was uncovered for two variables: self-reported grade and self-reported expected grade. This inconsistency was fixed for all data and no other systematic errors were observed for other variables. Of the ten percent of data selected for evaluation, less than one percent were entered incorrectly. Given the negligible amount of error observed in the random sample, all remaining data were assumed to be entered correctly.

**Post-hoc teacher exclusion.** After data collection, two participating teachers were excluded from the analysis. The first teacher was excluded due to the inappropriate nature of her content area. The teacher worked with five students receiving special education services and the classroom functioned as a study hall. The second teacher operated on a unique yearly schedule, and could not participate in the first data collection. This resulted in cluster-level missing data and the teacher was removed from the final analysis. Nevertheless, this teacher did participate in data collection at Time 2 and Time 3.

**Analyses**

The type of analysis used in the current study was contingent on the proposed research questions. For clarity, the research questions that guided the current study are reprinted here:

1. To what extent do teachers find the REACT to be useful as a tool to improve student perceptions of the classroom environment?
2. To what extent do teachers find information obtained using the REACT to be clear for interpretation?
3. Which strategies do teachers select when creating a reaction plan?
4. To what extent do teachers carry out the "reaction plan" developed after receiving feedback from REACT?
5. To what extent do student ratings differ in classrooms using the REACT feedback relative to student ratings in classrooms not using REACT feedback after a three week period of implementation?

Because the first three research questions are descriptive in nature, the accompanying analyses are also descriptive. Basic analyses for these research questions were conducted in SPSS and figures were created in Microsoft Excel.

Although not explicitly stated in the above research questions, structural evidence for the REACT is requisite for subsequent interpretations of students' ratings. Thus, before considering the final research question, the statistical program LISREL (version 8.1 for Windows) was used to conduct a CFA on the hypothesized structure of the REACT. Reporting these results is critical for two primary reasons. First, it is important to draw attention to the two items introduced later in the development of the REACT. These items may alter CFA results when assessed with a larger sample of students. Second, the partitioned structure of the REACT feedback, although initially created to aid teachers in developing a reaction plan, represents a novel conceptualization of the REACT. Although this alternative model (described in detail below) is similar to the initially proposed unidimensional model, a direct comparison between the two models is warranted.

The last topic for analysis—the degree to which student ratings differ as a function of group assignment—was assessed using hierarchical (or multi-level) linear modeling. Again, students' average rating across all 30 REACT items was used as the dependent variable. The statistical program HLM (Raudenbush, Bryk, Cheong, & Congdon, 2000) was used for all multi-level analyses. A-priori power analysis was conducted using Optimal Design (Raudenbush, et al., 2011).

In statistics, power refers to the probability of detecting an effect (i.e., rejecting the null hypothesis when it is false). Thus, higher values are preferred over lower values. In multi-level modeling, power analysis is often conducted to determine the number of clusters required to attain a minimal level of power. Traditionally, 0.8 is considered an acceptable level of power in educational research (Hedges & Rhoads, 2010); however, the amount of power required is highly dependent on the consequences of committing a Type II error (failing to reject when the null hypothesis is false). Because power analysis is typically conducted *a priori*, the parameters of the sample are unknown and must be predicted. Some of these parameters are selected based on tradition (e.g., an alpha level of 0.05), while others are informed by previous research in the field of interest (e.g., effect size, intra-class coefficient). The current study assumed an alpha level of .05, a cluster size of 25, an effect size of .5, and an intra-class coefficient (ICC) of .15. The cluster size was based on the average class size for high school students in Minnesota and the ICC was predicted using previous research with the REACT. Under these parameters, a sample of 30 clusters (teachers) was associated with 0.86 power.

# Chapter 4

# Results

Results for the current study are organized according to the specific research question addressed. Descriptive analyses for the teacher survey and REACT are discussed first. Results from the HLM analyses are presented last; however, before any results are detailed, evidence for the underlying structure of the REACT is explored.

## Confirmatory Factor Analysis

As discussed in the Methods section, preliminary work with the REACT provided evidence for unidimensionality. That is, previous work suggested that each of the 30 items on the REACT loaded on one construct: the Classroom Teaching Environment. Nevertheless, when teachers received feedback in the current study, the 30 items were partitioned into qualitatively different categories to aid with interpretation. Although the purpose of the categories was practical in nature, it is possible that the grouping used for the current study represents a better fitting model relative to the previously supported, unidimensional model. Examining the categories presented in Appendix B, it follows that the alternative model contains 30 items loading on seven factors. Nevertheless, previous work with the REACT—as well as the notion of classroom interdependence—would predict the presence of a second-order factor (i.e., the classroom teaching environment) accounting for the shared relationship between the seven first-order factors.

Although both potential models provide evidence for unidimensionality, the second-order model explicitly recognizes the presence of first-order factors. The difference between the two models for interpretation of an overall average is negligible—both provide evidence for interpreting a mean score—but evidence for the second-order model supports the use of subfactors as well.

**Model fit.** Correlations between the 30 REACT items are provided in Figure 3. Although data were obtained on three occasions, student ratings at Time 1 were used for the CFA. Before moving the data to LISREL, a correlation matrix was created using SPSS. Pairwise deletion resulted in a sample size ranging from 709 to 716. In total, 729 students participated in the first data collection. In this case, missing data are not problematic as less than 3% of the data were observed to be missing. Both a priori models were tested using the correlation matrix obtained from data collection at Time 1.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| **2** | .32 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| **3** | .33 | .39 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| **4** | .13 | .16 | .22 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| **5** | .20 | .21 | .23 | .21 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | |
| **6** | .37 | .22 | .26 | .19 | .23 | 1 | | | | | | | | | | | | | | | | | | | | | | | | |
| **7** | .31 | .26 | .31 | .26 | .19 | .20 | 1 | | | | | | | | | | | | | | | | | | | | | | | |
| **8** | .26 | .32 | .28 | .22 | .19 | .22 | .37 | 1 | | | | | | | | | | | | | | | | | | | | | | |
| **9** | .52 | .28 | .38 | .19 | .23 | .34 | .35 | .37 | 1 | | | | | | | | | | | | | | | | | | | | | |
| **10** | .55 | .29 | .33 | .21 | .23 | .38 | .32 | .30 | .54 | 1 | | | | | | | | | | | | | | | | | | | | |
| **11** | .16 | .24 | .25 | .14 | .18 | .26 | .22 | .29 | .27 | .29 | 1 | | | | | | | | | | | | | | | | | | | |
| **12** | .21 | .20 | .25 | .17 | .18 | .26 | .23 | .32 | .26 | .19 | .31 | 1 | | | | | | | | | | | | | | | | | | |
| **13** | .27 | .29 | .39 | .22 | .22 | .21 | .31 | .29 | .33 | .30 | .24 | .31 | 1 | | | | | | | | | | | | | | | | | |
| **14** | .29 | .21 | .32 | .28 | .22 | .20 | .46 | .38 | .42 | .32 | .16 | .20 | .28 | 1 | | | | | | | | | | | | | | | | |
| **15** | .25 | .26 | .33 | .19 | .24 | .30 | .30 | .36 | .32 | .35 | .37 | .32 | .32 | .32 | 1 | | | | | | | | | | | | | | | |
| **16** | .27 | .32 | .30 | .15 | .20 | .13 | .31 | .27 | .22 | .26 | .18 | .26 | .26 | .28 | .31 | 1 | | | | | | | | | | | | | | |
| **17** | .20 | .34 | .28 | .18 | .25 | .22 | .28 | .29 | .22 | .25 | .22 | .24 | .27 | .30 | .30 | .39 | 1 | | | | | | | | | | | | | |
| **18** | .29 | .26 | .26 | .12 | .23 | .31 | .27 | .29 | .31 | .32 | .34 | .34 | .28 | .28 | .45 | .28 | .31 | 1 | | | | | | | | | | | | |
| **19** | .37 | .23 | .35 | .30 | .21 | .27 | .46 | .40 | .44 | .37 | .17 | .19 | .23 | .63 | .31 | .26 | .26 | .31 | 1 | | | | | | | | | | | |
| **20** | .16 | .13 | .20 | .18 | .14 | .14 | .25 | .22 | .23 | .24 | .20 | .21 | .29 | .26 | .24 | .25 | .22 | .22 | .24 | 1 | | | | | | | | | | |
| **21** | .21 | .14 | .17 | .15 | .14 | .20 | .19 | .23 | .24 | .25 | .17 | .23 | .19 | .21 | .18 | .15 | .16 | .24 | .19 | .22 | 1 | | | | | | | | | |
| **22** | .35 | .25 | .31 | .12 | .16 | .25 | .27 | .24 | .34 | .35 | .26 | .22 | .29 | .25 | .31 | .26 | .26 | .29 | .26 | .29 | .22 | 1 | | | | | | | | |
| **23** | .26 | .29 | .34 | .30 | .19 | .17 | .37 | .27 | .32 | .27 | .21 | .17 | .25 | .42 | .27 | .32 | .28 | .23 | .47 | .24 | .15 | .23 | 1 | | | | | | | |
| **24** | .35 | .22 | .34 | .27 | .20 | .30 | .44 | .32 | .42 | .38 | .17 | .20 | .23 | .65 | .30 | .29 | .27 | .28 | .76 | .21 | .20 | .26 | .44 | 1 | | | | | | |
| **25** | .20 | .27 | .34 | .17 | .14 | .28 | .23 | .30 | .33 | .32 | .33 | .27 | .25 | .27 | .36 | .20 | .21 | .38 | .31 | .18 | .25 | .24 | .29 | .32 | 1 | | | | | |
| **26** | .37 | .28 | .32 | .23 | .19 | .34 | .29 | .29 | .41 | .43 | .31 | .31 | .29 | .34 | .40 | .24 | .24 | .39 | .31 | .22 | .33 | .33 | .29 | .34 | .47 | 1 | | | | |
| **27** | .16 | .34 | .26 | .07 | .15 | .19 | .20 | .30 | .22 | .18 | .20 | .26 | .27 | .24 | .27 | .28 | .33 | .25 | .20 | .11 | .09 | .24 | .25 | .16 | .21 | .26 | 1 | | | |
| **28** | .33 | .25 | .27 | .13 | .19 | .53 | .18 | .22 | .32 | .33 | .24 | .25 | .17 | .23 | .29 | .19 | .27 | .30 | .31 | .14 | .18 | .23 | .21 | .38 | .30 | .36 | .20 | 1 | | |
| **29** | .24 | .22 | .28 | .17 | .19 | .21 | .23 | .27 | .25 | .23 | .15 | .18 | .23 | .28 | .24 | .23 | .18 | .21 | .33 | .22 | .09 | .18 | .25 | .26 | .30 | .27 | .12 | .25 | 1 | |
| **30** | .30 | .21 | .38 | .22 | .22 | .27 | .32 | .28 | .39 | .33 | .31 | .16 | .26 | .37 | .31 | .26 | .23 | .31 | .38 | .24 | .22 | .33 | .31 | .40 | .32 | .31 | .16 | .28 | .37 | 1 |

*Figure 3.* Bivariate correlations for all REACT items at Time 1.

Fit indices for both models—as well as benchmark criteria for each fit index—are reported in Table 3. Measures of model fit can be partitioned into three categories: absolute, parsimony-adjusted, and incremental (Bandalos & Finney, 2010). Given the unique contribution of each type of fit index, researchers typically report one or more fit index from each category; however, there is a great deal of discrepancy in which fit indices are reported and there are a wide range of values reported for any given fit index (Schreiber, et al., 2006). Nevertheless, specific, and commonly accepted guidelines around fit indices exist (Bandalos & Finney, 2010; Hu & Bentler, 1999; Yu, 2002) and are referenced in Table 3.

Table 3.

*Comparison of Unidimensional and Second-Order Model Fit Statistics*

| Statistic | Unidimensional | Second-order | Acceptable Values |
|---|---|---|---|
| Degrees of Freedom | 407 | 398 | |
| Absolute Indices | | | |
| Chi-Square | 2180.15 | 1139.65($p$<.001) | $p$>.05 |
| Standardized Root Mean Square Residual (SRMR) | .11 | .05 | ≤ .08 |
| Parsimonious Fit Indices | | | |
| Root Mean Square Error of Approximation | .09 | .05 | ≤ .05 = Good<br>≤ .08 = Acceptable |
| Incremental Indices (Fit relative to a baseline model) | | | |
| Non-Normed Fit Index | .91 | .95 | ≥ .95 |
| Comparative Fit Index | .92 | .97 | ≥ .95 |

Absolute fit indices compare the observed correlation matrix to the hypothesized model and include the chi-square test and the standardized root mean square residual (SRMR). The chi-square test for both hypothesized models was statistically significant (*p* <.001) suggesting bad model fit (i.e., the proposed model deviates from the population of inference); however, the chi-square test is considered by many to be an overly stringent criterion for CFA (Bandalos & Finney, 2010; Maruyama, 1997). Because the chi-square test is highly sensitive to sample size—which tends to be very large in CFA research—the null hypothesis is rarely retained. Thus, the SRMR may be a more appropriate assessment of absolute model fit. The SRMR for the current study was much lower for the second-order factor model (.05) relative to the unidimensional model (.11), and provides evidence for the second-order factor model.

Relative to absolute fit indices, parsimony-adjusted fit indices penalize researchers for overly complex models. The most commonly reported parsimony-adjusted fit index is the root mean square error of approximation (RMSEA), with lower values indicating better fit. Similar to the absolute indices, the RMSEA for the second-order factor model (.05) indicated a good fit to the data.

Finally, incremental, or comparative, fit indices compare the fit of the model to a "null model" in which no correlations among variables are specified.  The two most commonly reported incremental fit indices include the comparative fit index (CFI) and the Tucker-Lewis Index (TLI). In both cases, values of .95 and above indicate an acceptable fit. Although the CFI (.91) and TLI (.92) values for the unidimensional model were nearly acceptable, the observed fit indices for the second-order model were stronger (CFI=.97, TLI=.96).

Excluding the chi-square test, all fit indices supported the tenability of the second-order factor model. The final structural model for the REACT is presented in Figure 4. Although factor loadings differed in magnitude, all standardized values for the second-order factor model were positive and significant at $p < .05$. The overall fit indices for the proposed factor model provide sufficient evidence for using an overall REACT score to assess the classroom teaching environment.

It is important to note that the second-order factor model also specifies seven first-order factors (see Figure 4); however, computing and analyzing these scores is beyond the scope of the current study. Further, although each factor loading was statistically significant, low factor loadings for some of the variables indicate localized model misfit.

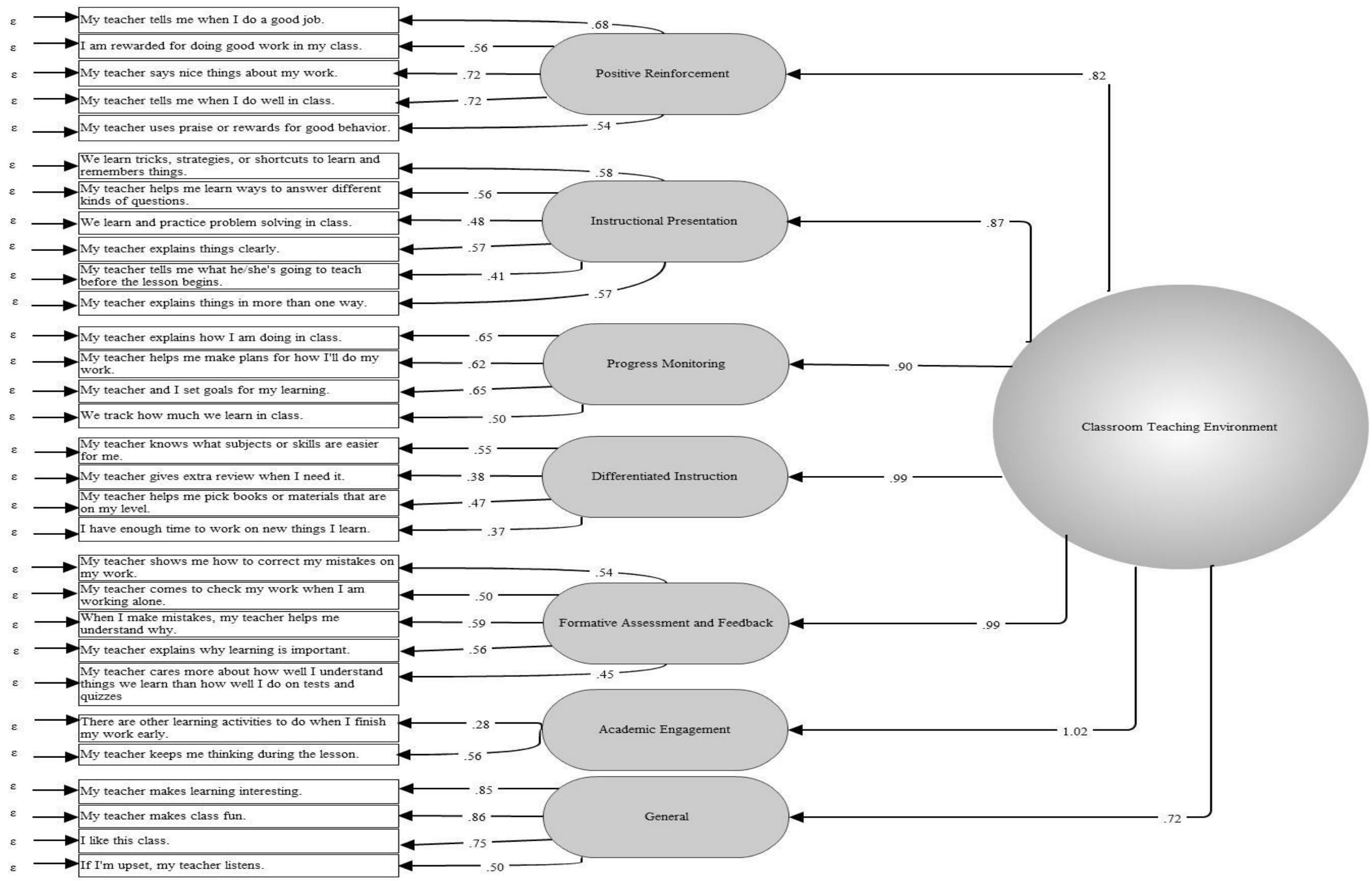
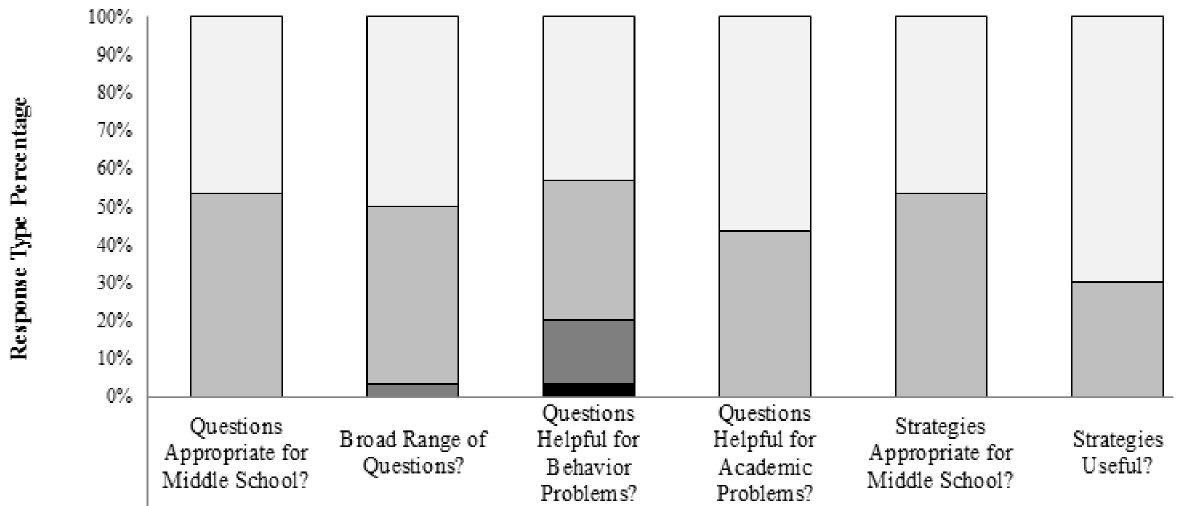



*Figure 4*. Final second-order structural model with completely standardized factor loadings.

**Teacher Survey**

The research questions associated with the teacher survey were primarily descriptive (i.e., to what degree do teachers find the REACT to be useful and clear). Although the inferences associated with descriptive analysis are confined to the sample, such analyses offer critical insight into the experience of those who are expected to use the REACT in practice.

The questions included on the teacher survey can be partitioned into two broad groups relating to the (a) utility or (b) practicality of the REACT. There were six questions on the teacher survey (see Appendix C) relating to utility, and four questions relating to practical aspects of the REACT. All teachers responded to each question and there were no missing data. The frequency distribution for the first six questions is available in Figure 5 and data for the last four questions are available in Figure 6. From these figures, it is clear that most teachers responded favorably (i.e., “Mostly Agree” or “Agree”) to nearly all questions on the survey. In regard to the utility of the REACT, teachers tended to perceive the questions and strategies to be useful; however, 20% of teachers found the questions on the REACT to be inappropriate for addressing behavior problems in the class. All teachers included in the sample agreed that the REACT questions were helpful for addressing academic problems. Likewise, all teachers considered the REACT questions and strategies to be appropriate for students in middle school.

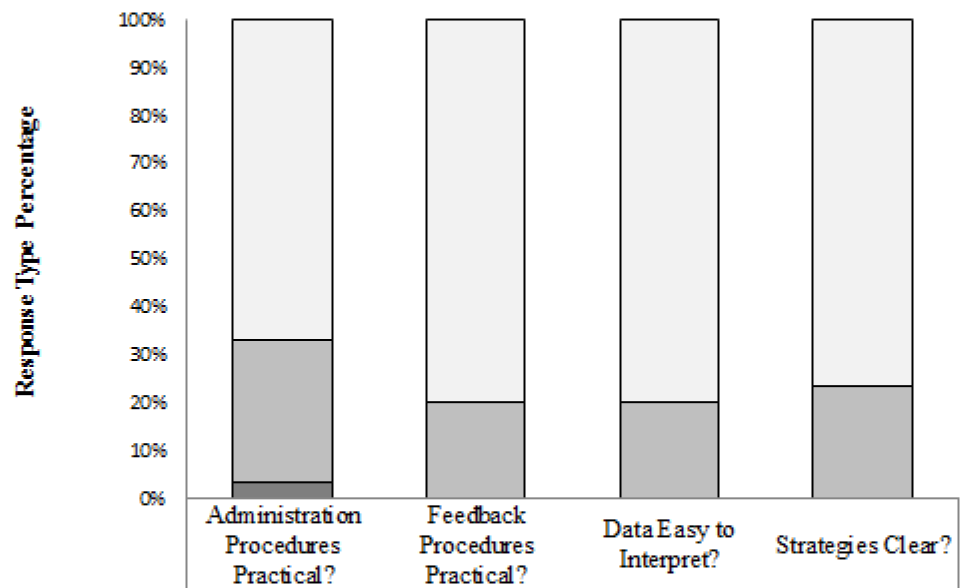



| | Questions Appropriate for Middle School? | Broad Range of Questions? | Questions Helpful for Behavior Problems? | Questions Helpful for Academic Problems? | Strategies Appropriate for Middle School? | Strategies Useful? |
|---|---|---|---|---|---|---|
| □ Agree | 14 | 15 | 13 | 17 | 14 | 21 |
| ■ Mostly Agree | 16 | 14 | 11 | 13 | 16 | 9 |
| ■ Mostly Disagree | 0 | 1 | 5 | 0 | 0 | 0 |
| ■ Disagree | 0 | 0 | 1 | 0 | 0 | 0 |

*Figure 5.* Teachers' responses to survey questions related to the REACT items and associated strategies.

Teachers found the REACT administration procedures to be practical for applied implementation, with only one teacher indicating that data collection was not practical. All teachers responded "Agree" (80%) or "Mostly Agree" (20%) when asked whether the feedback meeting procedures were practical. In regard to interpretation, an overwhelming percentage of teachers found the REACT strategies to be clear (77% responded "Agree" and 23% responded "Mostly Agree") and easy to interpret ( 80% responded "Agree" and 20% responded "Mostly Agree").

**Reported strategies.** The last two descriptive research questions relate to the strategies selected by teachers and the degree to which teachers implemented said strategies. As outlined in the procedures subsection, all teachers created a reaction plan during the feedback meeting. Teachers reflected on the classroom summary and selected one or more strategy groups for implementation. In addition, teachers indicated whether they planned to implement a strategy of their own, a strategy included on the REACT summary, or both. On average, teachers selected 2.23 strategies (SD=1.55).




| | Administration Procedures Practical? | Feedback Procedures Practical? | Data Easy to Interpret? | Strategies Clear? |
|---|---|---|---|---|
| Agree | 20 | 24 | 24 | 23 |
| Mostly Agree | 9 | 6 | 6 | 7 |
| Mostly Disagree | 1 | 0 | 0 | 0 |
| Disagree | 0 | 0 | 0 | 0 |

*Figure 6.* Teachers' responses to survey questions relating to REACT administration procedures and clarity of strategies.

Overall, teachers tended to select progress monitoring (30%) and positive reinforcement (25%) strategies groups most often. Conversely, the strategies relating to instructional presentation and differentiated instruction were selected less often (5% and 8% respectively). Across all strategy groups, teachers tended to select REACT strategies for implementation as rather than electing their own. This was most evident in the positive reinforcement group, in which all but one teacher selected a strategy provided in the REACT summary. Despite the overall trend of selecting REACT strategies, teachers who implemented a progress monitoring strategy tended to select a strategy of their own more often.

After implementing the chosen strategies for 3-weeks, each teacher was asked to respond to the statement "I used the strategy I selected for implementation" by selecting "Agree", "Mostly Agree", "Mostly Disagree", or "Disagree." Teachers answered this question for each chosen strategy group. For example, if a teacher planned to implement strategies that related to positive reinforcement and progress monitoring, he or she would provide information on implementation integrity twice (one for each group). Responses were coded sequentially (Agree=4 to Disagree=1). Teachers' reported implementation ranged from one to four. The average implementation rating was 3.09 (*SD*=0.79), with most teachers responding Agree (31%) or Mostly Agree (51%).

Finally, all teachers were asked whether or not they would use the REACT in practice. All teachers expressed interest in using the questionnaire in the future; however, 67% of teachers indicated a need for support with data collection and analysis.

**Descriptive Analyses for Student-Level Explanatory Variables**

Before exploring the effect of group assignment on students' REACT scores, it is worthwhile to parse out some of the descriptive statistics of interest by group assignment. Table 4 includes students' average REACT scores, reported grades, expected grades, and reported frequency of getting into trouble for Time 1, 2, and 3, separated by group. Note that students' responses were coded sequentially, with "Yes" corresponding to 4 and "No" corresponding to 1.

Table 4.

*Descriptive Statistics for Student-Level Variables Partitioned by Group Assignment.*

| Variable | Control (N=362-408) Mean | Control (N=362-408) SD | Experimental (N=298-343) Mean | Experimental (N=298-343) SD |
|---|---|---|---|---|
| REACT Score: Time 1 | 2.93 | 0.50 | 2.99 | 0.45 |
| REACT Score: Time 2 | 2.77 | 0.56 | 2.93 | 0.54 |
| REACT Score: Time 3 | 2.79 | 0.64 | 2.94 | 0.58 |
| Reported Grade: Time 1 | 4.30 | 0.95 | 4.35 | 0.79 |
| Reported Grade: Time 2 | 4.49 | 0.80 | 4.44 | 0.78 |
| Reported Grade: Time 3 | 4.30 | 0.89 | 4.19 | 0.93 |
| Expected Grade: Time 1 | 4.64 | 0.65 | 4.72 | 0.48 |
| Expected Grade: Time 2 | 4.75 | 0.49 | 4.69 | 0.53 |
| Expected Grade: Time 3 | 4.74 | 0.51 | 4.72 | 0.53 |
| Reported Trouble in Class: Time 1 | 1.26 | 0.53 | 1.33 | 0.62 |
| Reported Trouble in Class: Time 2 | 1.32 | 0.61 | 1.36 | 0.68 |
| Reported Trouble in Class: Time 3 | 1.34 | 0.60 | 1.34 | 0.60 |

*Note: REACT Score*=Average Item Score. *Reported/Expected Letter Grade:* 5=A; 4=B; 3=C; 2=D; 1=F. *Trouble in Class*: 4=Yes, 3=Mostly Yes, 2=Mostly No, 1=No. *White:* 0=Non-White, 1=White.

**REACT scores.** Examining students' average REACT scores from the first data collection, it appears as though students in classrooms assigned to either group tended to respond in a similar fashion, with both averages falling just under three (which corresponds to "Mostly Yes"). Thus, on average, students responded favorably across all classrooms. Students' responses at Time 2 suggest a slight decrease in perceptions, although this decrease appears to be more pronounced for the students of teachers who did not receive REACT feedback. At Time 3, students' responses were nearly identical to Time 2. At the time of the third data collection all teachers had received REACT feedback; however, teachers assigned to the experimental group did not receive feedback after the second data collection. When all three data collections are considered together, descriptive analyses indicate an overall decrease in students' responses over time; however, responses were generally more positive than negative.

**Item response frequencies.** To provide further insight into students' responses on the REACT across all classrooms, frequency counts for each item after the first data collection are presented in Figure 7. This figure mirrors the classroom summary available to teachers in the feedback meeting. Note that positive and negative responses are dichotomized, which allows for clearer inspection of the response distribution. Visual analysis of Figure 7 suggests that items differed widely in the frequency of students responding positively or negatively. For example, the items "My teacher helps me pick books or materials that are on my level" and "I am rewarded for doing good work in my class" received more negative responses overall than other items such as "My teacher explains things clearly." Further, groups of items also appear to be rated differently across all classrooms, with items relating to progress monitoring receiving more negative responses relative to items relating to instructional presentation.

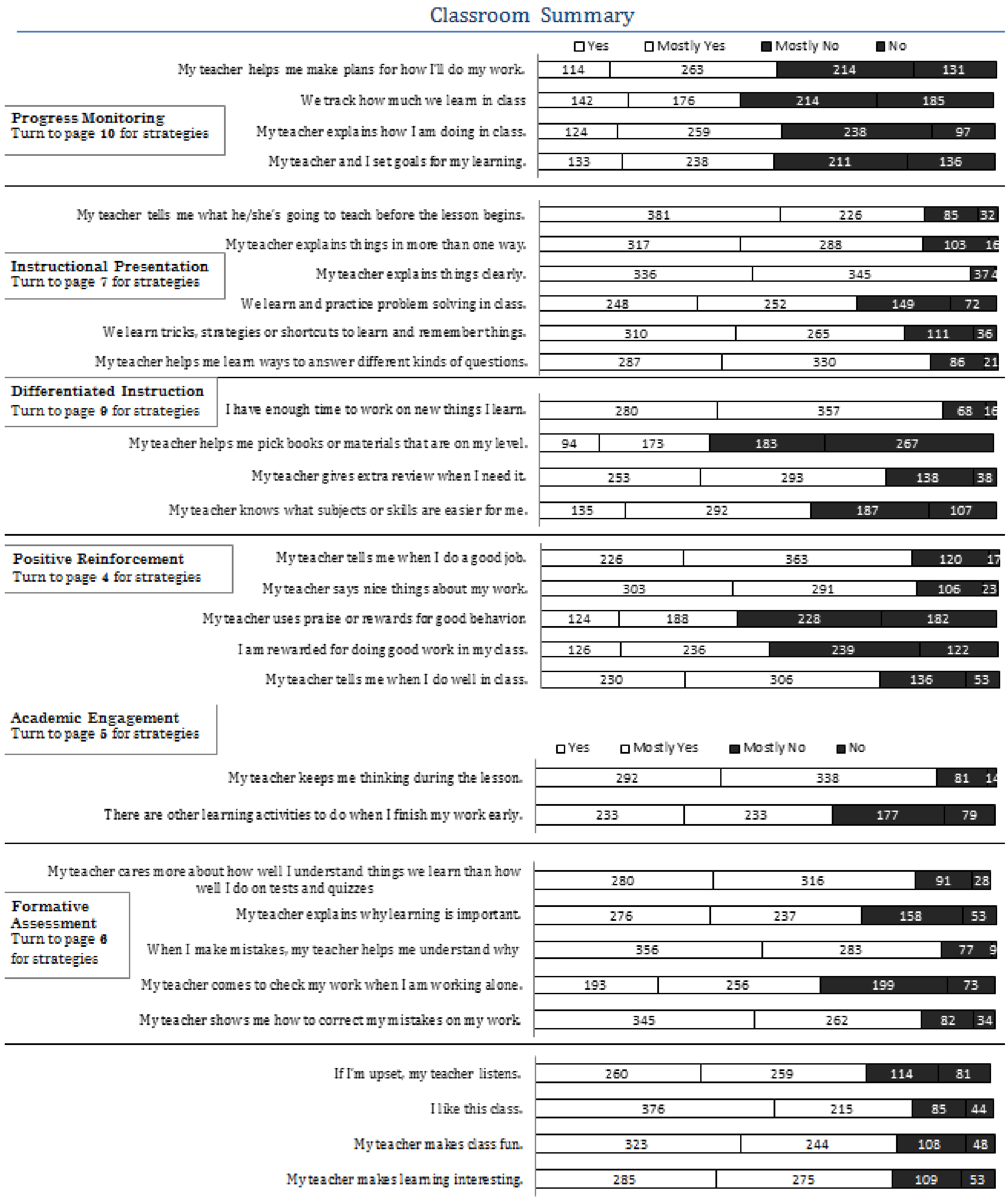



*Figure 7.* Distribution of students' responses on the REACT across all classrooms.

**Reported and expected grades.** Students' reported and expected grades were sequentially coded such that a reported grade of "A" was equal to 5 and a reported grade of "F" was equal to 1. An examination of students' reported and expected grades from the first data collection would suggest that students in classrooms assigned to either group tended to report relatively high grades, with students in both groups reporting estimated and expected grades between an A and B (see Table 4). This trend was similar for the second and third data collections as well.

**Reported trouble.** The degree to which students reported getting into trouble in class was coded such that "Yes" corresponded to 4 and "No" corresponded to 1. For both groups and across all data collections, students tended to respond negatively to this item, with averages falling between "No" and "Mostly No" (see Table 4).

**Variance in REACT scores across teachers.** Before formally assessing the variance among teachers, it is useful to examine average REACT scores by teacher. Means and standard deviations for each teacher included in the final analysis are available in Table 5. Note that teachers are ordered by the REACT average at Time 1. Examining teachers' classroom averages for any given data collection, there appears to be variability across classrooms. That is, classroom averages at Time 1, 2, and 3 show variability among the 30 teachers included in the study. Although there are some classrooms that showed signs of change between the three data collections, variability across occasions is less apparent. The general decrease in students' perceptions over time is also evident when examining averages for individual teachers.

Table 5.

*Average REACT Scores for all Teachers at Time 1, 2, and 3.*

| | Time 1 | | Time 2 | | Time 3 | | | Time 1 | | Time 2 | | Time 3 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | M | SD | M | SD | M | SD | | M | SD | M | SD | M | SD |
| **Tch. 1** | 3.30 | 0.22 | 3.20 | 0.18 | 3.26 | 0.27 | **Tch. 16** | 2.95 | 0.34 | 2.79 | 0.52 | 2.95 | 0.54 |
| **Tch. 2** | 3.26 | 0.26 | 3.22 | 0.40 | 3.19 | 0.42 | **Tch. 17** | 2.95 | 0.27 | 2.91 | 0.36 | 2.96 | 0.38 |
| **Tch. 3** | 3.23 | 0.26 | 3.30 | 0.23 | 3.39 | 0.21 | **Tch. 18** | 2.93 | 0.43 | 2.87 | 0.55 | 2.96 | 0.60 |
| **Tch. 4** | 3.21 | 0.53 | 3.00 | 0.68 | 3.17 | 0.66 | **Tch. 19** | 2.91 | 0.37 | 2.99 | 0.40 | 3.06 | 0.55 |
| **Tch. 5** | 3.19 | 0.46 | 2.97 | 0.53 | 3.12 | 0.56 | **Tch. 20** | 2.90 | 0.63 | 2.75 | 0.75 | 2.76 | 0.82 |
| **Tch. 6** | 3.17 | 0.37 | 3.21 | 0.42 | 3.18 | 0.43 | **Tch. 21** | 2.89 | 0.56 | 2.73 | 0.50 | 2.65 | 0.68 |
| **Tch. 7** | 3.14 | 0.39 | 2.98 | 0.49 | 3.13 | 0.51 | **Tch. 22** | 2.87 | 0.49 | 2.58 | 0.49 | 2.54 | 0.52 |
| **Tch. 8** | 3.12 | 0.40 | 2.98 | 0.40 | 2.95 | 0.45 | **Tch. 23** | 2.85 | 0.51 | 2.75 | 0.51 | 2.74 | 0.52 |
| **Tch. 9** | 3.12 | 0.47 | 2.87 | 0.34 | 2.79 | 0.43 | **Tch. 24** | 2.84 | 0.41 | 2.67 | 0.44 | 2.62 | 0.49 |
| **Tch. 10** | 3.10 | 0.32 | 3.10 | 0.44 | 3.10 | 0.58 | **Tch. 25** | 2.76 | 0.59 | 2.61 | 0.64 | 2.49 | 0.69 |
| **Tch. 11** | 3.06 | 0.51 | 3.03 | 0.58 | 2.98 | 0.65 | **Tch. 26** | 2.70 | 0.59 | 2.64 | 0.77 | 2.55 | 0.82 |
| **Tch. 12** | 3.05 | 0.39 | 3.00 | 0.52 | 3.14 | 0.51 | **Tch. 27** | 2.61 | 0.41 | 2.69 | 0.57 | 2.68 | 0.51 |
| **Tch. 13** | 3.04 | 0.37 | 2.99 | 0.44 | 3.22 | 0.50 | **Tch. 28** | 2.61 | 0.53 | 2.41 | 0.62 | 2.46 | 0.63 |
| **Tch. 14** | 2.99 | 0.40 | 2.71 | 0.54 | 2.68 | 0.40 | **Tch. 29** | 2.54 | 0.44 | 2.39 | 0.58 | 2.21 | 0.61 |
| **Tch. 15** | 2.98 | 0.39 | 2.90 | 0.38 | 2.88 | 0.46 | **Tch. 30** | | | 2.73 | 0.46 | 2.59 | 0.55 |

The descriptive data presented in Table 5 provide initial evidence for variability among teachers; however, examining the degree to which the observed variability is statistically significant, as well as whether this variability can be explained by covariates, requires a more rigorous approach to data analysis.

**Inferential Statistics: REACT Scores at Time 2**

Given that the current project sought to examine effects at the student level (for both student-level and teacher-level variables), a multi-level approach was predicted to be most suitable. Following a brief explanation of the model fitting process, two final models are presented. The first uses students' average REACT score at Time 2 as the dependent variable. As outlined in Figure 2, the first multi-level analysis represents the primary research question of interest as it directly assesses the impact of group assignment on REACT scores. The second model examines group differences at Time 3; however, differences between groups are less likely to emerge when REACT data from the third data collection are used because all teachers (control group included) had received student feedback prior to the third data collection (see Figure 2).

**Model fitting.** Multi-level models are typically constructed sequentially, with the "fully unconditional" model (explained in detail below) always serving as the first step. Put in more familiar terms, the fully unconditional model in HLM is a one-way random effects ANOVA with teachers included as the random effect. Thus, the fully unconditional model allows variance in the dependent variable to be partitioned into two levels—that of the student (level 1) and the teacher (level 2). If there is sufficient variance among teachers, subsequent models aim to explain variance at level 1 and 2. If there is no variance among teachers—and thus no clustering effect—it is typical for researchers to revert to a more parsimonious approach to analysis such as OLS regression.

As with any regression analysis, the process for model fitting should be theory-driven and researchers should recognize any limits imposed by the design or sample size of the study. In general, researchers caution against the temptation to include an unnecessarily large number of covariates in the model (McCoach, 2010; Raudenbush & Bryk, 2002). Raudenbush and Bryk (2002) liken this particular problem to focusing a projector—as the operator moves the projector farther and farther away, the image will invariably lose focus. Adding a large number of covariates, particularly when the sample size is small, often results in a blurred depiction of data. Nevertheless, it is best practice to examine the relationship between all explanatory variables included in the model as a failure to address collinearity may introduce substantial bias around effect sizes (McCoach, 2010; Raudenbush & Bryk, 2002).

Throughout the model building process for the current study, special attention was given to the (a) theoretical relevance of explanatory variables and (b) limits associated with sample size. In addition, the relationship among explanatory variables—as well as the relationship between explanatory variables and the dependent variable—was considered before decisions were made on the composition of the final model. In line with best practices in multi-level model fitting, a theory-driven and parsimonious approach was adopted for model fitting. This approach stands in contrast to more inclusive approaches; however, these approaches typically result in a saturated model with a moderate level of variance explained, but no significant explanatory variables due to an over-partitioning of variance.

**Fully unconditional model.** Before moving forward with the HLM analysis, it is standard practice to fit a fully unconditional model in which intercepts are allowed to vary across clusters (teachers). As noted, the equation for this fully unconditional model is equivalent to a one-way random effects ANOVA and is provided below.

$$Y_{ij} = \beta_{0j} + r_{ij} \qquad \textit{Level } 1$$

$$\beta_{0j} = \gamma_{00} + u_{0j} \qquad \textit{Level } 2$$

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij} \qquad \textit{Full Model}$$

The term of most interest in the above model is the random effect $u_{0j}$ at level 2. Given that teachers represent the second level of analysis, $u_{0j}$ represents the unique effect of the jth teacher. The $\gamma_{00}$ value represents the mean REACT score for all teachers. If a particular teacher's mean REACT score is equal to the grand mean, $u_{0j}$ would be equal to 0. The variance in $u_{0j}$ (overall deviation of cluster means from the grand mean) is represented by the term $\tau_{00}$. The null hypothesis for the fully unconditional model is Ho: $\tau_{00}=0$.

The results from the fully unconditional model are presented in Table 6. The variance among teacher intercepts ($\tau_{00}$) is included under the heading "Random Effect" and is equal to 0.04. The percentage of variance in the model attributable to teachers can be better represented by computing the intra-class coefficient (ICC), which is equal to the variance among teacher intercepts divided by the total variance ($\tau_{00}/\tau_{00}+\sigma^2$). In this case, the intra-class coefficient is equal to 0.13. This indicates that 13% of the variance in average REACT scores at Time 2 was at the cluster level. The observed chi-square statistic for the variance among teachers was statistically significant ($\chi^2$ (29, $N$ = 30) = 1182.86, $p$ <.001). Given the observed level of within-cluster dependence in the fully unconditional model, HLM was deemed to be the most appropriate approach for data analysis.

**Random intercept models.** The next steps in the model fitting process attempt to explain variation at the individual and teacher level. Note that "variance explained" connotes different meaning in multi-level models. Rather than an overall $R^2$ value, variance is separated by the level of analysis and does not represent the overall variance in the dependent variable.

Two random intercept models were fit to the data. In the first, eight variables at the student-level were added to the model. These included: gender, reported letter grade, expected letter grade, reported trouble in class, White/Non-White status, first REACT score, sixth grade status, and seventh grade status. Finally, group assignment—a level 2 explanatory variable—was added to create the final model. The equation for the final model is listed in Figure 8. Note that grade classifications were included as student level covariates because not all classrooms consisted of students in the same grade.

$$Y_{ij} = \beta_0 + \beta_1(\text{REACT.1} - \overline{\text{REACT.1}}_j) + \beta_2(\text{Gender}) + \beta_3(\text{Exp.Grade}) + \beta_4(\text{Rep.Grade}) + \beta_5(\text{White}) + \beta_6(\text{Trouble}) + \beta_7(\text{6th Grade}) + \beta_8(\text{7th Grade}) + \mu_{0j} + r_{ij}$$

$$\beta_0 = \gamma_{00} + \gamma_{01}(\text{Group Assignment}) + \gamma_{02}(\text{REACT.1}_j - \overline{\text{REACT.1}}) + \mu_{0j}$$

$$\beta_1 = \gamma_{10}$$

$$\beta_2 = \gamma_{20}$$

$$\beta_3 = \gamma_{30}$$

$$\beta_4 = \gamma_{40}$$

$$\beta_5 = \gamma_{50}$$

$$\beta_6 = \gamma_{60}$$

$$\beta_7 = \gamma_{70}$$

$$\beta_8 = \gamma_{80}$$

*Figure 8.* Full equation for final HLM model

***Student-level model.*** There are several important points to consider before interpreting the fixed and random effects associated with the student level model. First, it is customary to evaluate the overall fit of the model to the data. If two models are nested, a direct comparison can be made using *model deviance*, which is a measure of model fit. The deviance value for the new model is subtracted from the deviance associated with the previous model. The observed value can be linked to a chi square distribution with the degrees of freedom equal to the number of newly added parameters. All things being equal, if no statistically significant reduction in model deviance is observed, the new model may be disregarded as it does not improve fit. The deviance statistic of the student-level model, when compared to the fully unconditional model, was significantly lower, $\chi^2$ (8, $N$ = 28) = 997.57, $p < .001$, indicating a better fit to the data.

Second, level 1 slopes were not allowed to vary across schools (see Figure 8). That is, the variance of student-level slopes was constrained to 0, indicating that the effect for all predictors is assumed to be identical across clusters. Finally, it is important to note that students' first REACT score was centered at the group mean. The values for covariates centered at the group mean should be interpreted as the predicted change in Y for each unit increase above the cluster average of X. In this case, this represents the expected change in the REACT score at Time 2 if the first REACT score was one unit above the classroom average. Group mean centering was used to decrease the effect of the first REACT score—a student-level covariate—on cluster-level variance. One disadvantage of group mean centering is that cluster level explanatory variables are not adjusted for student level explanatory variables that have been centered at the group mean. This problem is remedied by creating a cluster level aggregate of the student level variable and including it in the final model.

Controlling for other covariates included in the student-level model, four variables were observed to be statistically significant. First, being male was associated with an average REACT score increase of 0.10, t(621)=3.40, $p<.01$, 95% CI[.16,.04]. The level 1 fixed effect of -0.10 for "Trouble in Class" was also statistically significant, t(621)=-3.26, $p<.01$, [-0.12,-0.04], and indicates that each unit of agreement with the statement "I get into trouble more often than other students," on average, was associated with a decrease in the average REACT score at Time 2. Relative to other grades, seventh grade status was associated with a -0.20 decrease in classroom perceptions at Time 2, t(621)=-3.04, $p<.01$, [-0.34,-0.06]. Finally, REACT score at Time 1 was a significant predictor of average REACT score at Time 2, with each unit increase in students' REACT score at Time 1 associated with an average increase of 0.83 in students' REACT score at Time 2, t(621)=-24.89, $p<.001$, [0.89,0.77].

In addition to the fixed effects, any reductions in variance at level-1 or level-2 are of interest. After entering all student-level predictors, there was a decrease in residual variance (0.126) and cluster variance (0.033). The amount of variance explained at each level of analysis is best represented by subtracting the variance component observed in the fully unconditional model by the variance component observed in the new model. The student level model accounted for 53.2% of the residual variance and 17.5% of the cluster level variance. Theoretically, none of the variables included in the student-level model should account for variance among teachers; however, if a student level variable differs somewhat systematically across clusters, it is possible. Despite the small reduction in variance at the cluster level, the significant chi-square statistic for the observed variance in average REACT scores among clusters (0.033) was still statistically significant, $\chi^2$ (28, $N$ = 29) = 185.09, $p$ <.001, and provides justification for additional model fitting.

***Final HLM model.*** The final model includes group assignment as well as students' first REACT scores aggregated to the teacher level. Although other cluster level variables were available to be included in the final model, these variables were omitted in the interest of clarity; however, it is important to note that none of the available cluster level variables were significantly correlated with group assignment or students' REACT scores at Time 2. Thus, it is highly unlikely that the omission of these variables introduced bias in fixed effect estimates.

As mentioned, it is important to consider any reductions in model deviance relative to previous models. Because the student level model is nested within the final model, it is acceptable to directly compare the deviance estimates between the two. The model deviance for the final model (481.80) was significantly lower than the deviance associated with the student level model (535.61), $\chi^2$ (2, $N$ = 28) = 53.81, $p$ < .001, indicating a better fit to the data. Given that a statistically significant reduction in variance was observed, interpretation of the fixed and random effects is appropriate.

Table 6.

*Fully Unconditional, Level-1, and Final HLM Models for Average REACT Score at Time 2.*

| | Fully Unconditional Model | | | Student-Level Model | | | Full Model | | |
|---|---|---|---|---|---|---|---|---|---|
| *Fixed Effect* | *Coefficient* | *se* | *t ratio* | *Coefficient* | *se* | *t ratio* | *Coefficient* | *se* | *t ratio* |
| REACT Score | 2.86 | 0.04 | 69.31*** | 2.96 | .18 | 16.70* | 2.84 | 0.17 | 16.99*** |
| Level 1 | | | | | | | | | |
| Gender | | | | 0.10 | 0.03 | 3.40** | 0.11 | 0.03 | 3.44** |
| Reported Letter Grade | | | | -0.01 | 0.02 | -0.48 | -0.01 | 0.03 | -0.27 |
| Expected Letter Grade | | | | 0.04 | 0.04 | 1.20 | 0.03 | 0.04 | 0.77 |
| Trouble in Class | | | | -0.08 | 0.02 | -3.26** | -0.07 | 0.02 | -3.89** |
| White/Non-White | | | | -0.05 | 0.04 | -1.27 | -0.03 | 0.04 | -0.84 |
| First REACT Score | | | | 0.83 | 0.03 | 24.89*** | 0.84 | 0.03 | 24.25*** |
| Sixth Grade | | | | -0.07 | 0.08 | -0.84 | -0.07 | 0.03 | -2.03* |
| Seventh Grade | | | | -0.20 | 0.07 | -3.04** | -0.05 | 0.04 | -1.38 |
| Level 2 | | | | | | | | | |
| First REACT Score (Aggregate) | | | | | | | 0.93 | 0.08 | 11.04*** |
| Group | | | | | | | 0.09 | 0.03 | 2.94** |

| *Random Effect* | *Variance Component* | *df* | $\chi^2$ | *p value* | *Variance Component* | *df* | $\chi^2$ | *p value* | *Variance Component* | *df* | $\chi^2$ | *p value* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Teachers | 0.040 | 29 | 1182.86 | 0.00 | 0.033 | 28 | 185.09 | 0.00 | 0.0001 | 26 | 30.28 | 0.256 |
| $\sigma^2$ | 0.269 | | | | 0.126 | | | | 0.126 | | | |
| *Additional Model Statistics* | | | | | | | | | | | | |
| Between-ClassVariance Explained | | | | | | | | 17.5 | | | | 99.8 |
| Residual Variance Explained | | | | | | | | 53.2 | | | | 53.2 |
| Deviance | | | | 1193.14 | | | | 535.61 | | | | 531.60 |

*Note:* Based on 29 of 30 classrooms. *First REACT Score* is centered at the group mean. *Gender:* 0 female; 1 male. *Reported/Expected Letter Grade:* 5=A; 4=B; 3=C; 2=D; 1=F. *Trouble in Class*: 4=Yes, 3=Mostly Yes, 2=Mostly No, 1=No. *White:* 0=Non-White, 1=White. *Group:* 0=Control, 1=Feedback.

* $p < .05$ ** $p < .01$ *** $p < .001$

Controlling for other covariates included in the final model, the fixed effects of gender, self-reported trouble, and REACT score at Time 1 were statistically significant and similar to the estimates associated with the student level model. The statistical significance associated with being in the seventh grade disappeared in the final model; however, being in the *sixth* grade variable was, on average, associated with a the slightly negative effect equal to 0.07[-0.10,-0.04], t(619)=-2.03, *p*<.05.

When aggregated to the teacher level, students' perceptions at Time 1 were associated with an average increase of 0.93 in perceptions at Time 2, t(26)=11.03, *p*<.001, [1.10,0.85]. Again, the aggregate of students' first REACT scores was included at the teacher level because the effect of group assignment would have otherwise remained unadjusted for students' prior perceptions due to the group centering of REACT scores at the student level. Note that the aggregate variable was centered at the grand mean. Thus, each unit increase above the grand mean for cluster REACT scores at Time 1 was associated with a 0.93 increase in students' REACT scores at Time 2.

Controlling for all other explanatory variables in the model, group assignment—whether teachers received feedback after the first data collection or not—was a statistically significant predictor of students' perceptions of the classroom teaching environment at the second data collection, t(26)=2.94, *p*<.01. More specifically, group assignment was associated with an average REACT increase of 0.09 [0.12, 0.06]. That is, students of teachers who received feedback on the classroom teaching environment tended to view the same environment more favorably three weeks later relative to students of teachers who did not receive feedback. Further consideration of the unstandardized effect of group assignment suggests that the effect is small. The effect of teacher feedback on students' perceptions of the classroom teaching environment can be standardized by dividing the unstandardized effect (0.09) by the square root of the total variance($\delta = \frac{\widehat{\gamma_{01}}}{\sqrt{\sigma^2 + \tau_{00}}}$), which is equal to 0.16 [0.21, 0.11]. In other words, the REACT scores at Time 2 for teachers who received feedback at Time 1 were 0.16 standard deviations higher than the scores of teachers who did not receive feedback.

Examining the amount of variance explained between clusters provides another indication of the explanatory power associated with the two cluster level variables included in the final model. Considered together, students' aggregated REACT scores at Time 1 and group assignment explained 99.8% of the variance between classrooms for REACT scores at Time 2. Relative to the student level model, the final model accounted for an additional 82.3% of variance among clusters and provides further evidence for the importance of including these variables in the final model.

**Model checking.** While some of the statistical assumptions that underlie multi-level models are difficult to assess statistically (e.g., independence of observations), others can be directly addressed. A failure to meet statistical assumptions may bias estimates, making inferences from the fitted models inappropriate. Several plots were constructed to visually analyze the homogeneity and normality of level 1 and level 2 residuals for the final model (See Figure 9). Visual analysis of student level residuals across clusters (Panel A in Figure 9) provides evidence for homogeneity of these residuals. In addition, homogeneity of level 1 residuals can be formally assessed in HLM (i.e., $\boldsymbol{\sigma_j^2 = \sigma^2}$). The observed chi square statistic provides evidence that the assumption of homogeneity among model residuals is satisfied, $\chi^2$ (28, $N$ = 29) = 27.54, $p$ > .50. Next, the assumption of normality for the group residuals was evaluated by plotting the model implied values against the observed values (Panel B in Figure 9). If the observed values approximate model implied values, each point should fall on a 45 degree line drawn from the point of origin. Visual analysis of this plot provides evidence for normality of group residuals. Given the information provided in Figure 9, the final model presented in Table 6 appears to be defensible.

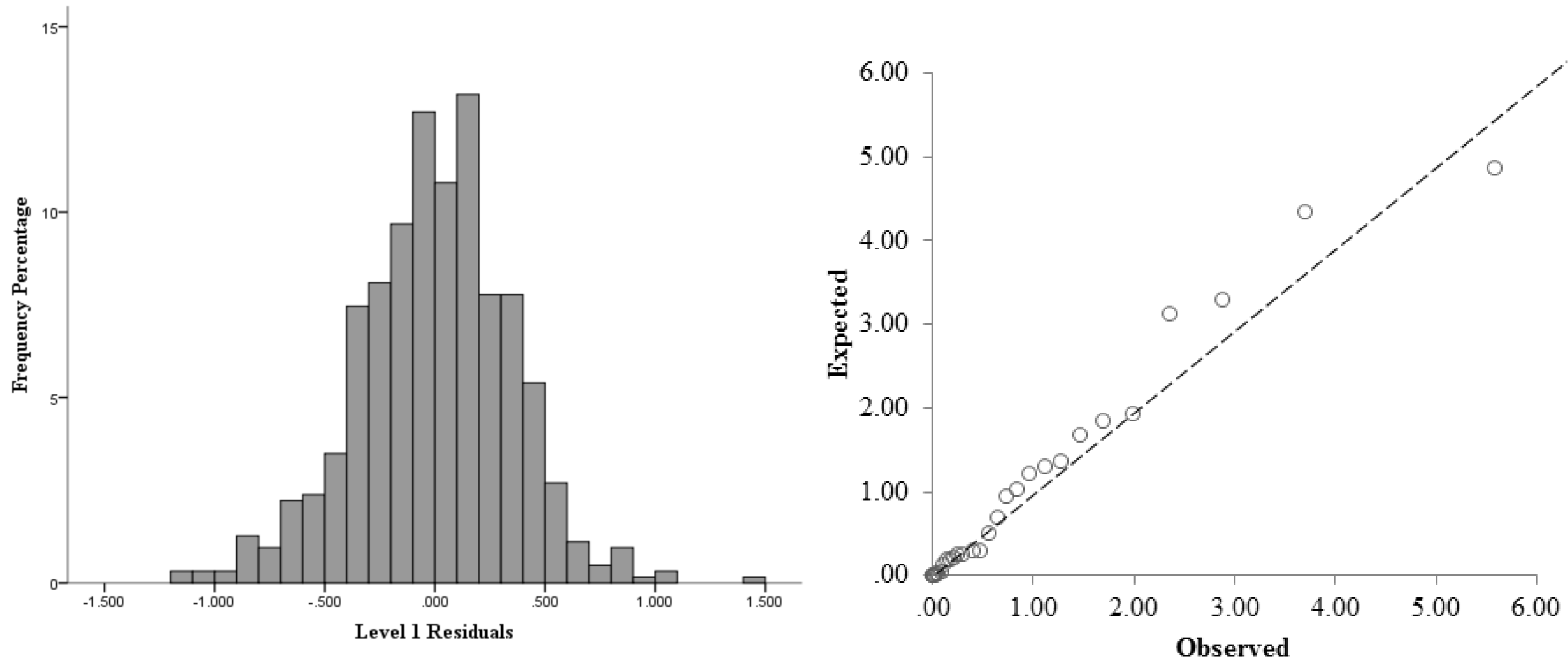



*Figure 9*. Model checking information for the final model using REACT scores at Time 2 as the outcome.

**Inferential Statistics: REACT Scores at Time 3**

The results from the fully unconditional model are presented in Table 7. The variance among school intercepts ($\tau_{00}$) was equal to 0.068, resulting in an ICC of .18. This indicates that 18% of the variance in students' REACT scores at Time 3 was at the classroom level. The observed chi-square statistic for the variance among teachers in the fully unconditional model was statistically significant ($\chi^2$ (29, $N$ = 30) = 190.20, $p$ <.001).

**Random intercept models.** As noted, the student level and full model for REACT scores at Time 3 are nearly identical to the models used previously; however, there are a few important differences regarding the hypothesized effect of group assignment as well as the composition of each model. Although group assignment was observed to be a significant predictor of REACT scores at Time 2, by Time 3 all teachers had received feedback on the REACT (teachers assigned to the control group received feedback after the second data collection). Further, teachers assigned to the experimental group did not receive feedback after the second data collection. Thus, the hypothesized effect of group assignment at Time 2 does not apply when considering students' responses at Time 3.

The variables included in the models are also slightly different. Rather than using students' first REACT score as a control for previous perceptions of the classroom teaching environment, student's second REACT score was used because it was more proximal to the outcome of interest. Likewise, information obtained about students' grades, expected grades, and perceived degree of getting into trouble was updated to reflect data obtained during the third data collection. No other changes were made to the model fitting sequence or the centering of variables. Results from each model are discussed below.

***Student-level model.*** The deviance statistic of the student-level model, when compared to the fully unconditional model, was significantly lower, $\chi^2$ (8, $N$ = 28) = 810.86, $p$ < .001, indicating a better fit to the data. Similar to the results observed when using Time 2 REACT scores as the outcome, statistically significant effects were observed for gender, reported trouble, and previous perceptions of the classroom teaching environment. Being male was associated with a 0.10 [.14, .16] average increase in perceptions at Time 3, t(572)=2.70, $p$<.01, while a one unit increase in the "Trouble in Class" variable was associated with an average decrease of 0.07 [0.10, 0.04] in perceptions at Time 3, t(572)=2.56, $p$<.01. Students' REACT scores at Time 2 were highly predictive of REACT scores at Time 3, with each average increase above the group mean associated with a 0.81 increase in REACT scores at Time 3, t(572)=18.66, $p$<.001. No other statistically significant fixed effects were observed.

Decreases in residual variance (0.113) and cluster variance (0.051) were observed after entering all student level predictors. 53.2% of the residual variance and 17.5% of the cluster level variance was explained by the student level model. The significant chi-square statistic for the observed variance in REACT scores among clusters was statistically significant, $\chi^2$ (28, $N$ = 29) = 282.07, $p$ <.001, providing justification for additional model fitting.

***Final HLM model.*** The observed deviance for the final model (381.85) was significantly lower than the deviance associated with the student level model (445.29), $\chi^2$ (2, $N$ = 28) = 63.44, $p$ < .001, indicating a better fit to the data.

Controlling for other covariates included in the final model, the fixed effects of gender, self-reported trouble, and REACT score at Time 1 were statistically significant and similar to the estimates associated with the student level model; however, after including group assignment and the aggregated REACT variable, being in the seventh grade was associated with a significant decrease in REACT scores at Time 3 relative to students in other grades, t(570)=-2.70, $p$<.01.

Table 7.

*Fully Unconditional, Level-1, and Final HLM Models for Average REACT Score at Time 3.*

| | Fully Unconditional Model | | | Student-Level Model | | | Full Model | | |
|---|---|---|---|---|---|---|---|---|---|
| *Fixed Effect* | *Coefficient* | *se* | *t ratio* | *Coefficient* | *se* | *t ratio* | *Coefficient* | *se* | *t ratio* |
| REACT Score | 2.88 | 0.05 | 55.05*** | 2.72 | 0.17 | 16.12*** | 2.64 | 0.13 | 20.37*** |
| Level 1 | | | | | | | | | |
| Gender | | | | 0.10 | 0.04 | 2.70** | 0.10 | 0.04 | 2.81** |
| Reported Letter Grade | | | | 0.02 | 0.02 | 1.02 | 0.02 | 0.02 | 1.05 |
| Expected Letter Grade | | | | 0.03 | 0.03 | 0.79 | 0.04 | 0.03 | 1.17 |
| Trouble in Class | | | | -0.07 | 0.03 | -2.56** | -0.07 | 0.03 | -2.43** |
| White/Non-White | | | | 0.03 | 0.04 | 0.82 | 0.04 | 0.03 | 1.14 |
| First REACT Score | | | | 0.81 | 0.04 | 18.66*** | 0.81 | 0.04 | 18.54*** |
| Sixth Grade | | | | 0.02 | 0.10 | 0.23 | 0.01 | 0.04 | 0.19 |
| Seventh Grade | | | | -0.25 | 0.13 | -1.91 | -0.11 | 0.04 | -2.70** |
| Level 2 | | | | | | | | | |
| First REACT Score (Aggregate) | | | | | | | 1.09 | 0.06 | 16.99*** |
| Group | | | | | | | -0.03 | 0.03 | -1.09 |

| *Random Effect* | *Variance Component* | *df* | $\chi^2$ | *p value* | *Variance Component* | *df* | $\chi^2$ | *p value* | *Variance Component* | *df* | $\chi^2$ | *p value* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Teachers | 0.068 | 29 | 190.20 | 0.00 | 0.051 | 28 | 282.07 | 0.00 | 0.001 | 26 | 30.28 | 0.256 |
| $\sigma^2$ | 0.312 | | | | 0.113 | | | | 0.112 | | | |

| *Additional Model Statistics* | Fully Unconditional Model | Student-Level Model | Full Model |
|---|---|---|---|
| Between-Class Variance Explained | | 25.0 | 98.0 |
| Residual Variance Explained | | 63.8 | 64.1 |
| Deviance | 1256.15 | 445.29 | 531.60 |

*Note:* Based on 29 of 30 classrooms. *First REACT Score* is centered at the group mean. *Gender:* 0 female; 1 male. *Reported/Expected Letter Grade:* 5=A; 4=B; 3=C; 2=D; 1=F. *Trouble in Class*: 4=Yes, 3=Mostly Yes, 2=Mostly No, 1=No. *White:* 0=Non-White, 1=White. *Group:* 0=Control, 1=Feedback.

* $p < .05$ ** $p < .01$ *** $p < .001$

When aggregated to the teacher level, each unit increase above the grand mean for students' perceptions at Time 2 was associated with an average increase in perceptions at Time 3 of 1.09 [1.15, 1.03], $t(26)=16.99$, $p<.001$. Controlling for all other explanatory variables in the model, group assignment was not a statistically significant predictor of students' perceptions of the classroom teaching environment at Time 3.

Similar to the model for REACT scores at Time 2, a substantial amount of variance in REACT scores among clusters was explained by group assignment and classroom level REACT scores at Time 2. Considered together, students' aggregated REACT scores at Time 2 and group assignment explained 98% of the variance between classrooms for REACT scores at Time 3. Relative to the student level model, the final model accounted for an additional 73% of variance among clusters and supports the inclusion of these variables in the final model; however, the non-significant fixed effect for group assignment indicates that the explanatory power of this variable decreased after all teachers had received feedback.

## Follow-Up Teacher Interviews

Following the study, one teacher from each school was randomly selected to participate in an informal interview about the REACT. The teacher survey (Appendix C) served as the primary guide for the interview and there were three broad areas of inquiry: REACT questions, REACT strategies, and study procedures.

**REACT questions.** All teachers found the REACT questions to be useful and appropriate overall. In particular, teachers expressed approval with the specific focus of many questions. The opportunity to learn about more specific aspects of the classroom teaching environment represented a novel opportunity for each of the three teachers. One teacher noted that many of the questions seemed geared toward upper elementary students instead of middle school students. There were also criticisms about the wording of specific items. For example, all three teachers suggested that the item "my teacher helps me pick books or materials that are on my level" be modified to include the words "assignments" instead of books.

**REACT strategies.** Overall, teachers found the strategies provided in the REACT summary to be highly appropriate and useful. The teachers all noted the ease with which they were able to identify a strategy that seemed appropriate for their classroom and there was general agreement on the effectiveness of the strategies. The use of rewards was a mildly controversial strategy, with one teacher viewing rewards as detrimental to intrinsic motivation.

**Study procedures.** All three teachers found the administration procedures to be feasible. If the REACT was available electronically, and summary reports were automatically generated, each teacher indicated that they would use the REACT in practice. Although teachers found the feedback meeting to be useful, there was unanimous agreement that the utility of the REACT could be drastically improved by allowing teachers to meet in small groups. In this format, teachers would receive feedback from students and then meet with other teachers who also received student feedback. All teachers also agreed that seeing students' responses at one time point was insufficient. If given the opportunity to view students' responses at Time 1, Time 2, and Time 3, teachers believed that students' responses would have improved more dramatically. This belief was grounded in the idea that changes to the classroom would have addressed the (possibly) evolving needs of students.

Finally, each teacher believed that the REACT could be effective when used with an individual student. Some of the items on the REACT were suggested to be rather difficult to address without knowing the identity of students (e.g., "I have enough time to work on new things I learn" or "My teacher and I set goals for my learning"). Thus, teachers believed that using the REACT with an individual student might increase the effect of changes to the classroom teaching environment.

# Chapter 5

# Discussion

The current study is most valuable because it provides a methodologically rigorous perspective on the impact of providing teachers with feedback on their classroom teaching environment. More specifically, results from the current study provide initial evidence that teachers who receive actionable feedback on students' perceptions of the classroom teaching environment observe more positive student perceptions three weeks later relative to teachers who do not receive such feedback.

The following subsections provide insight into the theoretical and practical significance of the reported results. The first subsections address findings related to dimensionality and several student level covariates included in the final model. The last sections provide a discussion around teachers' perceptions of the REACT and the significant effect observed for group assignment. Finally, limitations and suggestions for future work are discussed. Traditional conceptualizations of external and internal validity are addressed throughout.

## Dimensionality

Although some researchers have recognized the relationship among qualitatively different components of the classroom environment, this has not been the norm. Nearly all available instruments presume multidimensionality and no instruments explain the correlations among factors by using a general—or second-order—factor. The current study expands upon the existing literature on the classroom environment by providing sufficient evidence for a second-order factor structure. While the possible reasons for this structure have been discussed briefly (see description of materials), it bears repeating that the classroom environment may not be composed of entirely independent constructs. Thus, one plausible reason for the observed model fit is that the classroom teaching environment truly functions—at least through the eyes of students—as an interdependent construct. Indeed, the common use of the term "ecology" in educational research implies some level of interdependence. Previous research has observed moderate to strong correlations between factors (Fraser & Fisher, 1983; Fraser, 1998; Goh & Fraser, 1996; McMahon et al., 2009). The formation of a second-order factor simply makes the concept of interdependence more explicit.

The specific content included on the REACT is another plausible reason for the observed factor structure. In general, factor models reflect the structure of the items included in the final model. It is possible that if every aspect of the classroom environment were measured, a second-order factor structure would not be supported. The items included on the REACT dictate what is truly measured and a skeptic might suggest that those items are an incomplete reflection of the classroom teaching environment. This, of course, is a valid concern—it would be misleading to measure "x" and label it "y." As with any assessment, those who use the REACT would do well to consider the specific items that comprise the instrument and the resulting utility of the measure. Nevertheless, the development of the REACT—as well as the results reported here—offer support for the notion that the "classroom teaching environment" is indeed the measurement construct. More specifically, all items were mapped to 11 components of instructional support based on the longstanding work of experts in classroom environment research (Ysseldyke & Christenson, 2002). These items were evaluated by graduate students and practicing teachers for wording, connection to the 11 aforementioned components, and overall diversity in content. After substantial field testing, all 11 components were represented across the 30 REACT items. Finally, 30 of 31 teachers included in the current study agreed that the items on the REACT cover a broad range of instructional supports in the classroom. These facts offer support for the relationship between the items on the REACT and the overall classroom teaching environment. The REACT does not measure every aspect of the classroom teaching environment, but the procedures by which it was developed, the results from the CFA, and the teacher survey lend support to the notion that it does in fact measure the classroom teaching environment.

**Gender, Trouble, and Rating Stability**

**Gender.** Although it was not the central focus of this study, the observed gender effects are notable in light of previous research suggesting the opposite. That is, previous work examining students' perceptions of the classroom environment has reported males to have lower perceptions (on average) relative to female students (Byrne, et al., 1986; Doll, et al., 2010; Raviv, Raviv, & Reisel, 1990). The obvious conjecture regarding the results observed in the current study may draw attention to the age of participants—perhaps females begin to perceive the classroom environment more negatively upon entering middle school? While this is plausible, previous research observing more negative responses among male students has included middle school environments as well (Sinclair & Fraser, 2002). Nevertheless, the effect size associated with gender is often small, with some studies finding no differences (Eccles & Blumenfield, 1985; Goodenow, 1993).

Given that the correlations between gender and students' perceptions of the classroom environment are often quite small, it is possible that other covariates may negate or even reverse the observed effect. While the current findings differ from previous research, so do the covariates that comprise the final model. Perhaps of most interest in the case of gender, is the inclusion of self-reported trouble making. As most who have spent time in schools can attest, males tend to get into much more trouble than females in educational settings. Beyond anecdotal evidence, this disproportionality is readily apparent when partitioning behavioral referrals by gender (Kaufman, et al. 2010; Skiba, Peterson, & Williams, 1997; Sullivan, Klingbeil, & Van Norman, in press).

It is important to recall that in the current study, students who reported getting into trouble more often tended to view the classroom environment more negatively. These "trouble making" students tended to also be male. Given the results observed in the final model—the composition of which differs importantly from similar work—it is plausible to suggest that it is not *males* that perceive the classroom environment more negatively, but *trouble-makers*, who also tend to be male. If the degree to which students get into trouble is accounted for in statistical analyses, it may be possible to obtain a clearer estimate of gender effects. In other words, the seemingly unusual finding for the effect of gender may be informative for future work examining gender differences in classroom perceptions.

**Self-reported trouble.** The negative correlation between the average REACT score and student reports of getting into trouble represents a relatively new outcome for research using student ratings. Nevertheless, the notion that students who get into trouble often during class would also view the learning environment as less supportive is not itself a new idea. For example, in 33 interviews with students excluded from school, Pomeroy (1999) observed a strong consistency in students' reports of teacher support—or lack thereof. That is, many of the students interviewed by Pomeroy indicated that their teacher rarely provided the emotional or instructional support necessary for success.

Given the observed relationship between behavior problems and students' perceptions, it is tempting to ask whether inadequate supports in the classroom engender student misbehavior or misbehavior is changing the way teachers interact with their students. Perhaps the most likely explanation is that the relationship between student misbehavior and the classroom teaching environment is reciprocal in nature. Similar interactions have been hypothesized in discussions of student engagement (Altermatt, Jovanovic, & Perry, 1998). For example, when examining teacher-student interactions in a middle school, Altermatt et al. (1998) found that students who participated more in class often garnered the lion's share of teacher attention. It is possible, then, that students who are engaged in class receive more teacher attention and thus become more engaged in class. Conversely, students who do not exhibit high levels of academic engagement receive less teacher attention and become more withdrawn (Skinner, Kindermann, Connell, & Wellborn, 2009). A similar interaction may occur between student misbehavior and the classroom teaching environment. That is, misbehavior creates negative perceptions of classroom environment, which produce further misbehavior.

Regardless of the directionality in the relationship between trouble making and students' perceptions of the classroom environment, teachers have much to gain from collecting and using REACT data from students who get into trouble more often than others. These students may represent a classroom subgroup for which regular instructional practices are falling short of success. That is, if students who report getting into trouble more often tend to perceive the classroom teaching environment less positively, it may behoove teachers to use this information when determining a plan for instructional change.

**Stability in students' perceptions.** The primary reason for including at least two data collections was to provide a strong control for previous perceptions of the classroom environment. As with most studies controlling for prior levels of the dependent variable, students' perceptions at Time 1 were highly predictive of perceptions at Time 2. The large correlation between students' responses at Time 1 and Time 2 is consistent with research highlighting very small changes in students' perceptions of the classroom environment over time (Lawrenz, 1976). In fact, similar research has observed a slight decline in students' perceptions, regardless of initial ratings of classroom quality (Breckelmans, 1989; Mainhard, Brekelmans, Brok & Wubbels; 2011; Skinner & Belmont, 1993). While the REACT scores observed in the current study tended to differ *across* teachers, there was very little change over time. Indeed, the classroom profile (see Figure 7) for each teacher was remarkably similar on each occasion. Nevertheless, it is important to note that the stability observed in students' responses of the classroom environment may either reflect continuity at the student level or in the classroom environment itself.

Previous work (e.g., Mainhard et al., 2011) seems to adopt the view that *students* are stable; however, a more optimistic perspective may be that students' responses remain unchanged when *teachers* are stable. Although it would be short sighted to assume that students do not settle into their views of the classroom environment, the assumption that teachers do not settle into their own consistent routine is equally limited. From previous research, it follows that no changes—or slightly negative changes—in students' perceptions over the school year can be expected if the status quo remains intact; however, there is existing research lending support to the notion that changes in teaching behavior can alter students' perceptions of the classroom environment (Fraser, 1998; Sinclair & Fraser, 2002). Thus, while it is likely that students' perceptions are difficult to change, it is not an impossible task.

**Actionable Feedback as Catalyst for Instructional Change**

The results presented in the current study provide evidence for one method of using students' perceptions of the classroom environment as a tool for teacher feedback. Further, the use of a randomized group design improves the level of confidence associated with the observed effect of REACT feedback on students' perceptions of the classroom environment.

Hypotheses around the potential catalyst for the observed difference in students' responses can be framed within the professional development literature discussed in the Introduction section of this paper. More specifically, it is important to recall that professional development approaches are augmented when the content is aligned with evidence-based teaching practices, there is an explicit focus on teachers' own classrooms, and teachers are encouraged to actively participate in the process (Desimone, 2009; Garet et al., 2001).

In the context of the current study, teachers were able to view data on their own classroom and these data were connected to research-based ideas. Further, teachers were actively engaged in the interpretation of student data and the formulation of a reaction plan. It is likely that these characteristics were fundamental in the effectiveness of group assignment.

The most likely catalyst at work in the current study—simply viewing students' perceptions of the classroom teaching environment—is highly feasible in applied settings. In fact, the type of feedback given to teachers represents the tip of the iceberg in terms of professional development opportunities. Indeed, the feedback meeting was the primary opportunity for reflection and lasted a mere 20 minutes. In addition, teachers only saw classroom data on one occasion. This minimalist approach differs from similar work exploring how student perception data can guide instruction. For example, previous case studies (e.g., Sinclair & Fraser, 2002) detailing efforts to change students' ratings of the classroom environment used weekly meetings, and the window for changing students' responses spanned over a longer period of time.

The method of teacher support used in the current study can also be contextualized within the broader field of professional development, which continues to demonstrate the stubborn nature of genuine instructional change. In their 2000 study Supovitz & Turner examined survey data from over 3000 teachers who took part in a professional development initiative designed to increase the use of investigative teaching practices and foster a classroom culture of investigation. Involvement in targeted professional development was observed to have a positive impact on teaching practices and classroom environment; however, no positive effects were observed until teachers were exposed to more than 160 hours of professional development. Similar results were observed in later studies of a similar nature (Heck, Banilower, Weiss, & Rosenberg, 2008). Supovitz & Turner's (2000) study demonstrates the amount of resources that are often required to change deeply rooted instructional practices, and by extension, the broader culture of a classroom.

Despite the increasingly apparent need for intensive professional development, the parsimonious approach to feedback used in the current study was by design. It would make little sense to hire teacher coaches or provide daily feedback on implementation before exploring the impact of feedback alone. Identifying the minimum amount of support required to detect an effect is (perhaps) a more sensible approach. The amount of resources in schools is limited and assessments that can offer helpful information without a substantial need for additional resources are likely to be welcomed by teachers and administrators alike. Nevertheless, given the small effect size observed for group assignment, it will be necessary to explore aspects of implementation that can be expected to augment the impact of REACT feedback.

It is worthwhile to consider the input from teachers when considering the way in which the REACT can be used more effectively in practice. Although teachers found the process to be useful and productive, many indicated that access to students' ratings after each data collection would have greatly improved the effects associated with feedback. This was especially true for teachers who were assigned to the experimental group. These teachers were eager to view students' perceptions after the second data collection, but were disappointed to find that no additional data would be available until the third data collection. The recommendation to incorporate ongoing feedback was at odds with the purpose of the study, but not with the purpose of the REACT. In practice, teachers would be encouraged to use the REACT as often as needed.

The second most prominent point of feedback from teachers related to the way in which students' perceptions were used. Nearly all of the teachers found the procedures for the feedback meeting to be practical; however, many teachers indicated—sometimes without being prompted—that a group based approach to interpreting REACT data would be much more powerful. Given that many "teams" are already in place in schools (e.g., grade or content groups), teachers suggested that results from the REACT could be interpreted within these teams as part of existing professional development efforts. Alternatively, it might be useful for teachers to meet in groups based on their classroom profile. For example, if one teacher received particularly low ratings in progress monitoring, they could work with teachers with high ratings on these items. Regardless of the methods for grouping, it seems as though teachers stand to benefit from the opportunity to collaborate with their colleagues on the interpretation of REACT data.

**Limitations**

There are several limitations associated with the current study. First, internal validity issues relating to novelty and contamination are important to recognize. Teachers in the control group were aware that students were answering questions about the classroom environment. The novelty of the measure and the knowledge that data would be collected in the future may have produced a change in teacher behavior. To help weaken the possibility that teachers took special efforts to ensure high ratings, it was explicitly communicated that only the teacher and the researcher would have access to classroom data. In other words, it was made clear to teachers that classroom data would not be used for evaluative purposes.

In addition, the current design made it difficult to determine which components of the process were sufficient for producing changes in students' perceptions. For example, it would be acceptable to argue that changes were observed simply as a result of teachers having access to student data. The response to this argument is a practical one. It may be that viewing raw student data—or even the act of collecting data—is enough to promote changes students' perceptions, but the addition of strategies to the REACT summary requires virtually no additional resources and is aligned with the vision for the measure's use in schools. In short, the current study assesses the impact of a *process,* not necessarily a component of that process. Given that each component of the process should be present in practice (i.e., communicating intent to students, data collection, reflection, use of strategies) questions about the relative importance of the components are somewhat irrelevant. Teachers' use of the REACT in the current study constitutes the most minimal approach to improving students' perceptions.

In regard to external validity, there are several points to consider. Most importantly, teachers were not randomly selected for participation. Rather, all participants were volunteers who agreed to participate after hearing about the project at a short faculty meeting. The use of volunteers, while commonplace in educational research, restricts inferences to a broader population of teachers and students—especially those who work in urban school districts. Finally, as previously noted, the current study explores the impact of a process. Should teachers and students respond favorably to the REACT, deviations from the process outlined in the methodology section of this study may produce different results.

**Future Research**

The opportunities for future research in this area are abundant. The results from the current study, although interesting in their own right, are most interesting when framed as a reference for future work. The process by which teachers used REACT data in the context of this study can be situated at the beginning of a resource continuum. That is, at a minimum, teachers should view students' perceptions and have access to research-based strategies for instruction. Future work should explore the impact on students' perceptions as time and resources are added to the current approach. The amount of time and effort teachers can devote to classroom improvement efforts is limited, and methods for using students' perceptions of the classroom environment must improve the status quo without introducing undue burden.

Future work may also contribute to the field of classroom environment research by examining specific groups of individuals. The anonymity associated with the classroom level design used in the current study becomes increasingly difficult to maintain with a narrower unit of measurement; however, it would allow teachers to be more pointed in their efforts to improve classroom environment. Indeed, it may be frustrating for teachers to, for example, observe negative perceptions among 25% of the classroom without knowing which students comprise the 25%. Previous research exploring the usefulness of students' perceptions in a qualitative manner (e.g., Sinclair & Fraser, 2002; Waldrip et al. 2009) may serve as a reference for future work with individual students.

Finally, while the general body of classroom environment research supports a connection between students' perceptions of the classroom environment and several important outcomes, additional work of this nature is needed with the REACT. This research might compare the results of the REACT with students' achievement, engagement, self-efficacy, or attendance; however, it may be more interesting to compare the explanatory value of the REACT with existing measures of students' perceptions.

**Concluding Remarks**

Similar to previous work that explored students' perceptions of the classroom environment, the current study observed remarkable consistency in average REACT ratings over time—students appear to be reliable in their evaluations of the classroom teaching environment. Nevertheless, teachers who received feedback on their students' perceptions tended to observe more positive ratings of the classroom teaching environment relative to teachers who had not received feedback. This effect was present after controlling for several key variables, including students' prior REACT scores, gender, and self-reported behavior problems. The impact of REACT feedback is particularly relevant in light of the minimally invasive manner in which it was provided to teachers. When paired with information from follow up interviews, the small effect of group assignment is a promising beginning for researchers and applied professionals who, like many leaders in education, see value in measuring and improving the classroom experience of students.

**References**


Adams, G. L., & Engelmann, S. (1996). *Research on direct instruction: 25 years beyond DISTAR*. Seattle, WA: Educational Achievement System.

Altermatt, E. R., Jovanovic, J., & Perry, M. (1998). Bias or responsivity? Sex and achievement-level effects on teachers' classroom questioning practices. *Journal of Educational Psychology*, *90*(3), 516-527. 516-527.

American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders (4th ed., text rev.). doi:10.1176/appi.books.9780890423349

Anderman, E.M., Maehr, E.M., & Midgley, C. (1999). Declining motivation after the transition to middle school: Schools can make a difference. *Journal of Research & Development in Education, 32(3),* 131-147.

Anderson, L.W. (2002). Curricular alignment: A re-examination. *Theory into Practice, 41(4),* 255-260.

Arlin,M. (1979). Teacher transitions can disrupt time flow in classrooms. *American Educational Research Journal, 16(1),* 42-56. doi: 10.3102/00028312016001042

Baer, D.M., Wolf, M.M., & Risley, T.R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1(1),* 91-97.

Bandalos, D. L., & Finney, S. J. (2010). Factor analysis: Exploratory and confirmatory. In G.Hancock & R. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences,* 93-114. New York, NY: Routledge.

Beck, M., Burns, MK, & Lau, M.(2009). The effect of preteaching reading skills on the on-task behavior of children identified with behavioral disorders. Behavioral Disorders, 33, 91–99.

Bernaus, M. and Gardner, R. C. (2008). Teacher motivation strategies, student perceptions, student motivation, and English achievement. *The Modern Language Journal, 92,* 387–401. doi: 10.1111/j.1540-4781.2008.00753.x

Bibou-Nakou, I., Kiosseoglou, G., & Stogiannidou, A. (2000). Elementary teachers' perceptions regarding school behavior problems: Implications for school psychological services. *Psychology in the Schools, 37(2),* 123-134. doi: 10.1002/(SICI)1520-6807

Bijou, S.W. (1968). What psychology has to offer education—now. *Journal of Applied Behavior Analysis, 3(1),* 65-71. doi: 10.190/jaba.1970.3-65

Black, P., & William, D. (2006). Assessment and classroom learning. *Assessment in Education: Principles, Policy, & Practice, 5(1),* 7-74. doi: 10.1080/0969595980050102

Block, J.H., & Burns, R.B. (1976). Mastery learning. *Review of Research in Education, 4,* 3-49.

Bloom, B.S. (1968). Learning for mastery. (*UCLA-CSEIP) Evaluation Comment, 1,* 2.

Bloom, B.S. (1971). Mastery learning and its implications for curriculum development. In E.W. Eisner (Ed). *Confronting Curriculum Reform*. Boston: Little, Brown.

Blose, RJ, & Fisher, D. (2003, April). Effects of teachers' school-level environment perceptions on changing elementary mathematics classroom environments. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Bonwell, C.C. (1991). Activing learning: Creating excitement in the classroom. ERIC Clearing House on Higher Education Washington DC.

Boren, R., Callahan, C., & Peugh, J. (2011). Out With the Old, In With the New: Factor Analyses of a Classroom Environment Measure. *Journal of Psychoeducational Assessment*, *29*(3), 214-224.

Bradshaw, C.P., Koth, C.W., Bevans, K.B., Lalongo, N., & Leaf, P.J. (2008). The impact of school-wide positive behavioral interventions and supports (PBIS) on the organizational health of elementary schools. *School Psychology Quarterly, 23(4),* 462-473. doi: 10.1037/a0012883

Brewster, A. B., & Bowen, G. L. (2004). Teacher support and the school engagement of Latino middle and high school students at risk of school failure. *Child and Adolescent Social Work Journal*, *21*(1), 47-67.

Bronfrenbrenner, U. (1979). *The ecology of human development.* Cambridge, MA: Harvard University Press.

Brophy, J. (1986). Teacher influences on student achievement. *American Psychologist, 41(10),* 1069-1077. doi:10.1037/0003-066X.41.10.1069.

Burnett, P. C. (2002). Teacher praise and feedback and students' perceptions of the classroom environment. *Educational psychology*, *22*(1), 5-16.

Burns, M.K., Codding, R.S., Boice, C.H., & Lukito, G. (2010). Meta-Analysis of acquisition and fluency math interventions with instructional and frustration level skills: Evidence for a skill-by-treatment interaction. *School Psychology Review, 39,* 69-83.

Burroughs, S., Groce, E. and Webeck, M. L. (2005). Social studies education in the age of testing and accountability. *Educational Measurement: Issues and Practice, 24,* 13–20. doi: 10.1111/j.1745-3992.2005.00015

Byrne, D. B., Hattie, J. A., & Fraser, B. J. (1986). Student perceptions of preferred classroom learning environment. *The Journal of Educational Research*, 10-18.

Cameron, J., & Pierce, W. D. (1994). Reinforcement, reward, and intrinsic motivation: A meta-analysis. *Review of Educational research*, *64*(3), 363-423. doi: 10.3102/00346543064003363

Campbell, S. & Skinner, C.H. (2008). Combining explicit timing with an interdependent group contingency program to decrease transition times. *Journal of Applied School Psychology, 20(2),* 11-27. doi: 10.1300/J370v20n02_02

Carnine, D. W., Silbert, J., Kame'enui, E. J., *&* Tarver, S. G. *(*2010*). Direct instruction reading (*5th ed.*).* Upper Saddle River, NJ*: Pearson, Merrill Prentice Hall.*

Carr et al. (2002). Positive behavior support: Evolution of an applied science. *Journal of Positive Behavior Interventions, 4(1),* 4-16. doi: 10.1177/109830070200400102

Chesebro, J.L. (2003). Effects of teacher clarity and nonverbal immediacy on student learning, receiver apprehension, and affect. *Communication Education, 52(2),* 135-147. doi: 10.1080/03634520302471

Christ, T.J., & Boice, C. (2009). A brief review of nomenclature, components, and formatting to inform the development of Direct Behavior Rating (DBR). *Assessment for Effective Intervention, 34(4),* 242-250. doi: 10.1177/1534508409336182

Christ, T.J., Nelson, P.M., & Demers, J. A. (2012). *The Responsive Environmental Assessment for Classroom Teaching (REACT).* Unpublished instrument.

Christ, T.J., Riley-Tillman, T.C., Chafouleas, S.M. (2009). Foundation for the development and use of Direct Behavior Rating (DBR) to assess and evaluate student behavior. *Assessment for Effective Intervention, 34(4),* 201-213. doi: 10.1177/1534508409340390

Christenson, S., Ysseldyke, J.E., Wang, J.J. & Algozzine, B. (1983). Teachers' attributions for problems that result in referral for psychoeducational evaluation. *The Journal of Educational Research , 76(3),* 174-180.

Chu, S., & Wllms, J.D. (1996). Effects of parental involvement on eigth-grade achievement. *Sociology of Education, 69(2),* 126-141.

Clunies-Ross, P., Little, E., & Kienhuis, M. (2008). Self-reported and actual use of proactive and reactive classroom management strategies and their relationship with teacher stress and student behavior. *Educational Psychology: An International Journal of Experimental Educational Psychology, 28(6),* 693-710. doi: 10.1080/01443410802206700

Codding, R.S., et al. (2007). Comparing mathematics interventions: Does initial fluency predict intervention effectiveness? *Journal of School Psychology, 45(6),* 603-617.

Coffield, F., & Learning and Skills Research Centre. (2004). *Learning styles and pedagogy in post-16 learning*. London: Learning and Skills Research Centre.

Cohen, S.A. (1987). Instructional alignment: Searching for a magic bullet. *Educational Researcher, 16(8),* 16-20. doi: 10.3102/0013189X016008016

Connor, C.M., et al. (2009). Individualizing student instruction precisely: Effects of child x instruction interactions on first graders' literacy development. *Child Development, 80(1),* 77-100. doi: 10.1111/j.1467-8624.2008.01247.x

Cooper, J (Ed.). (2003). *Classroom teaching skills* (7th ed.). Boston: Houghton Mifflin Company.

Cousins, J.B., & Walker, C.A. (2000). Predictors of educators' valuing of systematic inquiry in schools. *Canadian Journal of Program Evaluation, Special Issue,* 25-53.

Crane, J. (1996). Effects of home environment, SES, and maternal test scores on mathematics achievement. The Journal of Educational Research, 89(5), 305-314. doi: 10.1080/00220671.1996.9941332

Daly, E. J., Witt, J. C., Martens, B. K., & Dool, E. J. (1997). A model for conducting a functional analysis of academic performance problems. *School Psychology Review, 26*, 554-574.

Daly, E.J., & Martens, B.K. (1994). A comparison of three interventions for increasing oral reading performance: Application of the instructional hierarchy. *Journal of Applied Behavior Analysis, 27(3),* 459-469.

Darch, C. B., Kame'enui, E. J., & Crichlow, J. M. (2003). *Instructional classroom management: A proactive approach to behavior management* (2nd ed.). Upper Saddle River, NJ: Prentice Hall

Darling-Hammond, L. (1999). *Teacher quality and student achievement: A review of state policy evidence*. Center for the Study of Teaching and Policy, Seattle, WA: University of Washington.

Darling-Hammond, L., Chung, R., & Frelow, F. (2002). Variation in teacher preparation: How well do different pathways prepare teachers to teach? *Journal of Teacher Education, 53(4),* 286-302.

Dee. T.S., Jacob, B.A., Hoxby, C.M., & Ladd, H.F. (2010). The impact of No Child Left Behind on students, teachers, and schools [with comments and discussion]. *Brookings Papers on Economic Activity,* 149-207.

Desimone, L.M., Smith, T.M., Frisvold, D.E. (2009). Survey measures of classroom instruction: Comparing student and teacher reports. *Educational Policy, 24(2),* 267-329. doi: 10.1177/0895904808330173

Dobbs, J., & Arnold, D. (2009). Relationship between preschool teachers' reports of children's behavior and their behavior toward those children. *School Psychology Quarterly, 24(2),* 95-105. doi: 10.1037/a00116157.

Doll, B., Kurien, S., LeClair, C., Spies, R., Champion, A., & Osborn, A. (2009). The ClassMaps Survey: A framework for promoting positive classroom environments. *Handbook of positive psychology in the schools*, 213-227.

Dorman, J. P. (2002). Classroom environment research: Progress and possibilities. *Queensland Journal of Educational Research, 18,* 112-140.

Douglas, K. (2009). Sharpening our focus in measuring classroom instruction. *Educational Researcher, 38(7),* 518-521. doi: 10.3102/0013189X09350881

Druva, C. A. and Anderson, R. D. (1983), Science teacher characteristics by teacher behavior and by student outcome: A meta-analysis of research. *Journal of Research in Science Teaching, 20(5), 467–479*. doi: 10.1002/tea.3660200509

Duhon, G..J. et al. (2004). Identifying academic skill and performance deficits: The experimental analysis of brief assessments of academic skills. *School Psychology Review, 33(3),* 429-443.

Dunn, R. (1990). Understanding the Dunn and Dunn Learning Styles Model and the need for individual diagnosis and prescription. *Journal of Reading, Writing, and Learning Disabilities, 6(3),* 223-247.

Early D. M., Maxwell K.L., Burchinal M., Alva S., Bender R.H., Bryant D., et al. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development, 78,* 558–580. doi: 10.1111/j.1467-8624.2007.01014.x

Eccles, J. S., & Blumenfeld, P. (1985). Classroom experiences and student gender: Are there differences and do they matter. *Gender influences in classroom interaction*, 79-114.

Eccles, J.S. et al. (1993). Development during adolescence: The impact of stage-environment fit on young adolescents' experiences in school and in families. *American Psychologist, 48(2),* 90-101. doi: 10.1037/0003-066X.48.2.90

Emmer, E.T., Evertson, C.M., & Anderson, L.M. (1980). Effective classroom management at the beginning of the school year. *The Elementary School Journal, 80(5),* 219-231.

*Espinosa, L.M., & Laffey, J.M. (2003). Urban primary teacher perceptions of children with challenging behaviors. Journal of Children and Poverty, 9(2), 135-156. doi:* 10.1080/10796120305436.

Fast, L.A., et al. (2010). Does math self-efficacy mediate the effect of the perceived classroom environment on standardized math test performance? *Journal of Educational Psychology, 102(3),* 729-740. doi: 10.1037/a0018863

Feldlaufer, H., Midgley, C., & Eccles, J.S. (1988). Student, teacher, and observer perceptions of the classroom environment before and after the transition to junior high school. *The Journal of Early Adolescence, 8(2),* 133-156. doi: 10.1177/0272431688082003

Forman, S. G., & Burke, C. R. (2008). Best Practices in Selecting and Implementing Evidence-Based School Interventions. In A. Thomas & J. P. Grimes (Eds.), *Best practices in school psychology V*. Bethesda, MD: National Association of School Psychologists.

Fouts, R.E. & Meyers, R.E. (1992). A cluster analysis of high school science classroom environments and attitude toward science. *Journal of Research in Science Teaching, 29(9),* 929-937. doi: 10.1002/tea.3660290904

Fraser, B. (1998). Classroom environment instruments: Development, validity, and applications. *Learning Environments Research, 1(1),* 7-34. doi: 10.1023/A:1009932514731

Fraser, B. (1999). Using learning environment assessments to improve classroom and school climates. In H. J. Freiberg (Ed.), *School climate: Measuring, improving and sustaining healthy learning environments* (pp. 65–83). London, UK: Falmer Press.

Fraser, B. J. (2002) Learning environments research: Yesterday, today and tomorrow. In S. W. Goh and M. S. Khine (Eds) *Studies in Educational Learning Environments: An International Perspective* (pp. 1-25). Singapore: World Scientific

Fraser, B. J., McRobbie, C. J., & Fisher, D. L. (1996). *Development, validation and use of personal and class forms of a new classroom environment instrument.* Paper presented at the annual meeting of the American Educational Research Association, New York.

Fraser, B. J., Seddon, T., Eagleson, J. (1982). Use of student perceptions in facilitating improvement in classroom environment. *Australian Journal of Teacher Education, 7,* 31-42.

Fraser, B.J. (1981). Predictive validity of an individualized classroom environment questionnaire. *Alberta Journal of Educational Research, 27,* 240-251.

Fraser, B.J. (1998). Classroom environment instruments: Development, validity, and applications. *Learning Environments Research, 1(1),* 7-34. doi: 10.1023/A:1009932514731

Fraser, B.J., & Fisher, D.L. (1982). Predicting students' outcomes from their perceptions of classroom psychosocial environment. *American Educational Research Journal, 19(4),* 498-518. doi: 10.3102/00028312019004498

Fraser, B.J., Anderson, G.J. & Walberg, H.J. (1982). *Assessment of learning environments: Manual for Learning Environment Inventory (LEI) and My Class Inventory (MCI)* (3rd version), Western Australian Institute of Technology, Perth.

Fredricks, J.A., Blumenfeld, P.C., & Paris, A.H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74(1),* 59-109. doi: 10.3102/00346543074001059

Freiberg, J.H., & Waxman, H.C. (1988). Alternative feedback approaches for improving student teachers' classroom instruction. *Journal of Teacher Education, 38(4),* 8-14. doi: 10.1177/002248718803900403

Friedel, J.M., Cortina, K.S., Turner, J.C., & Midgley, C. (2010). Changes in efficacy beliefs in mathematics across the transition to middle school: Examining the effects of perceived teacher and parent goal emphases. *Journal of Educational Psychology, 102(1),* 101-114. doi: 10.1037/a0017590

Fuchs, D., Fuchs, L., Mathes, P.G., & Simmons, D. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal, 34(1),* 174-206. doi: 10.3102/00028312034001174

Fuller, B., Wright, J., Gesicki, K., Kang, E. (2007). Gauging growth: How to judge no child left behind? *Educational Researcher, 36(5),* 268-278.

Gage, N. L. (1963). *Handbook of research on teaching*. Chicago: Rand Mcnally.

Gardner, D. (1983).*A Nation at Risk: The Imperative for Educational Reform.* Report of the National Commission on Excellence in Education. Washington, DC: Government Printing Office.

Garet, M. et al. (2010). *Middle School Mathematics Professional Development Impact Study: Findings After the First Year of Implementation* (NCEE 2010-4009). Washington DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Gersten, R., & Baker, S. (2001). Teaching expressive writing to students with learning disabilities: A meta-analysis. *Elementary School Journal, 101,* 251-272.

Gettinger, M., Schienebeck, C., Seigal, S., Vollmer, L. (2011). Assessment of classroom environments. . In B. Melissa & K. Thomas (Eds.), *The Oxford Handbook of School Psychology*: Oxford University Press, Inc.

Gicking, E.E., & Armstrong, D.L. (1978). Levels of instructional difficulty as related to on-task behavior, task completion, and comprhension. *Journal of Learning Disabilities, 11(9),* 559-566.

Gnagey, W. J. (1975). *Maintaining discipline in classroom instruction*. New York: Macmillan.

Goodenow, C. (2006). The psychological sense of school membership among adolescents: Scale development and educational correlates. *Psychology in the Schools, 30(1),* 79-90. doi: 10.1002/1520-6807(199301)30:1

Goodenow, C. (1993). Classroom belonging among early adolescent students relationships to motivation and achievement. *The Journal of Early Adolescence*, *13*(1), 21-43.

Goh, S. C., & Fraser, B.J. (1996). Validation of an elementary school version of the questionnaire on teacher interaction. *Psychological Reports*, *79*(2), 515-522.

Gore, J.M. and Gitlin, A.D. (2004). (Re)visioning the academic-teacher divide: Power and knowledge in the educational community. *Teachers and Teaching: Theory and Practice*, 10(1), 35-58.

Greene, B.A., Miller, R.B., Crowson, M.H., Duke, B.L., & Akey, K.L. (2004). Predicting high school students' cognitive engagement and achievement: Contributions of classroom perceptions and motivation. *Contemporary Educational Psychology, 29(4),* 462-482.

Greenwood, C. R., Carta, J. J., Kamps, D., Terry, & Delquadri, J. (1994). Development and validation of standard classroom observation systems for school practitioners: Ecobehavioral assessment systems software (EBASS). *Exceptional Children*, *61*, 197–210

Guarino, C.M., Santibañez, L., & Daley, G.A. (2006). Teacher recruitment and retention: A review of the recent empirical literature. *Review of Educational Research, 76*(2). doi: 10.3102/00346543076002173

Guisbond, L., & Neil, M. (2004). Failing our children: No Child Left Behind undermines quality and equity in education. *The Clearing House, 78(1),* 12-16.

Hair, J. F., Anderson, R. E., Tatham, R. L., & William, C. (1998). Multivariate data analysis. Prentice Hall, Upper Saddle River, New Jersey.

Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child development*, *76*(5), 949-967.

Haring, N. G., & Eaton, M. D. (1978). Systematic instructional procedures: An instructional hierarchy. In N. G. Haring, T. C. Lovitt, M. D. Eaton, & C. L. Hansen(Eds.), *The fourth R: Research in the classroom* (pp. 23–40). Columbus OH: Merrill.

Hart, P. D. & Teeter, R. M. (2002). *A national priority: Americans speak on teacher quality*. Princeton: Educational Testing Service

Harwell, M., & LeBeau, B. (2010). Student eligibility for a free lunch as an SES measure in education research. *Educational Researcher*, *39(2),* 120-131.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Oxon: Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77(1)*, 81-112. doi: 10.3102/003465430298487

Hedges, L. V., & Rhoads, C. *"Statistical Power Analysis in Educational Research (NCSER 2010–3006)." Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education, 2010.*

*Retrieved January 15th, 2013, from* [*http://ies.ed.gov/ncser/pubs/20103006/pdf/20103006.pdf*](http://ies.ed.gov/ncser/pubs/20103006/pdf/20103006.pdf).

Heider, F. (1958). *The psychology of interpersonal relations.* New York: Wiley.

Helman, L.A. (2006). What's in a conversation? Mentoring stances in coaching conferences and how they matter. In B. Achinstein & S. Z. Athanases (Eds.), *Mentors in the making: Developing new leaders for new teachers* (pp. 69–82). New York: Teachers College Press

Helman, L.A. (2005). Using literacy assessment results to improve teaching for English-language learners. *The Reading Teacher, 58(7),* 668-677.

Hines, C., Cruickshank, D.R., Kennedy, J.J. (1985). Teacher clarity and its relationship to student achievement and satisfaction. *American Educational Research Journal, 22(1),* 87-99. doi: 10.3102/00028312022001087

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development, 74(5),* 1368-1378. doi: 10.1111/1467-8624.00612

Horner, et al. (2009). A randomized, wait-list controlled effectiveness trial assessing school-wide positive behavior support in elementary schools. *Journal of Positive Behavior Interventions, 11(3),* 133-144. doi: 10.1177/1098300709332067

Hoy, A.W., & Spero, R.B. (2005). Changes in efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education, 21(4),* 343-356.

Hughes, J. N. (2011). Longitudinal effects of teacher and student perceptions of teacher-student relationship qualities on academic adjustment. *The Elementary school journal*, *112*(1), 38.

Hughes, J. N., Wu, J. Y., Kwok, O. M., Villarreal, V., & Johnson, A. Y. (2012). Indirect effects of child reports of teacher–student relationship on achievement. *Journal of Educational Psychology*, *104*(2), 350.

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

Jimerson, S.R., Burns, M.K., & VanDerHeyden, A.M. (2006). Response to intervention at school: The science and practice of assessment and intervention. *Handbook of Response to Intervention, 1,* 3-9.

Johnson, B., & McClure, R. (2004). Validity and reliability of a shortened, revised version of the Constructivist Learning Environment Survey (CLES). *Learning Environments Research*, *7*(1), 65-80.

Johnson, D. W., Johnson, R., & Anderson, D. (1983). Social interdependence and classroom climate. *The Journal of Psychology*, *114*(1), 135-142.

Kaufman, J. S., Jaser, S. S., Vaughan, E. L., Reynolds, J. S., Di Donato, J., Bernard, S. N., & Hernandez-Brereton, M. (2010). Patterns in office referral data by grade, race/ethnicity, and gender. *Journal of Positive Behavior Interventions*, *12*(1), 44-54.

Kennedy, J.J., Cruickshank, D.R., Bush, A.J. and Myers, B. (1978). Additional investigations into the nature of teacher clarity. *Journal of Educational Research, 72(1),* 3-10.

Kirschner, P., Sweller, J., & Clark, R.E. (2010). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41(2),* 2006. doi: 10.1207/s15326985ep4102_1

Kirschner, P.A., Sweller, J., & Clark, R.E. (2010). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41(2),* 75-86. doi: 10.1207/s15326985ep4102_1

Koczor, M. L. (1984). *Effects of varying degrees of instructional alignment in posttreatment tests on mastery learning tasks of fourth grade children*. Unpublished doctoral dissertation, University of San Francisco.

Koedel, C. & Betts, J.R. (2007) Re-examining the role of teacher quality in the educational production function. San Diego, CA: University of Missouri.

Kounin, J. S. (1977). *Discipline and group management in classrooms*. Huntington, N.Y: R.E. Krieger Pub. Co.

Kulik, C.C., Kulik, J.A., & Bangert-Drowns, R.L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research, 60(2),* 265-299. doi: 10.3102/00346543060002265

Kyriakides, L., Creemers, B.P.M, & Antoniou, P. (2009). Teacher behavior and student outcomes: Suggestions for research on teacher training and professional development. *Teacher and Teacher Education, 25(1),* 12-23.

Lawrenz, F. (1976). The stability of student perception of the classroom learning environment. *Journal of Research in Science Teaching, 14(1),* 77-81.

Levy, J., Wubbels, T., Den Brok, P., & Brekelmans, M. (2003). Students' perceptions of interpersonal aspects of the learning environment. *Learning Environments Research*, *6*(1), 5-36.

Liu, X. S., & Meyer, J. P. (2005). Teachers' perceptions of their jobs: A multilevel analysis of the teacher follow-up survey for 1994–95. *Teachers College Record*,107(5), 985–1003.

Loup, K.S., Garland, J.S., Ellet, C.D., & Rugutt, J.K. (1996). Ten years later: findings from a replication study of teacher evaluation practices in our 100 largest school districts. *Journal of Personnel Evaluation in Education, 10(3),* 203-226. doi: 10.1007/BF00124986

Mainhard, M., Brekelmans, M., Brok, P. D., & Wubbels, T. (2011). The development of the classroom social climate during the first months of the school year. *Contemporary Educational Psychology*, *36*(3), 190-200.

Marks, H.M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American Educational Research Journal, 37(1),* 152-184. doi: 10.3102/00028312037001153

Maruyama, G. M. (1997). *Basics of structural equation modeling*. Sage Publications, Incorporated.

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

Massetti, G.M., et al. (2008). Academic achievement over 8 years among children who met modified criteria for Attention-deficit/Hyperactivity disorder at 4-6 years of age. *Journal of Abnormal Child Psychology, 36(3),* 399-410. doi: 10.1007/s10802-007-9186-4

Mavropoulou, S., & Padeliadu, S. (2002). Teachers' causal attributions for behavior problems in relation to perceptions of control. Educational Psychology: An International Journal of Experimental Educational Psychology, 22(2), 191-202. doi: 10.1080/01443410120115256

Mayer, R.E. (2008). Incorporating individual differences into the science of learning: Commentary on Sternberg et al. (2008). *Perspectives on Psychological Science, 3(6),* 507-508. doi: 10.1111/j.1745-6924.2008.00093.x

McCoach, D. B. (2010). Hierarchical linear modeling. In G. R. Hancock, & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 123–140). Routledge.

McMahon, S.D. & Wernsman, J. (2009). The relation of classroom environment and school belonging to academic self-efficacy among urban fourth and fifth grade students. *The Elementary School Journal, 109(3),* 267-281. doi: 10.1086/592307

Medley, D. M., Coker, H., & Soar, R. S. (1984). *Measurement-based evaluation of teacher performance: An empirical approach*. New York: Longman.

Moos, R.H. & Trickett, E.J. (1974) *Classroom Environment Scale Manual.* Palo Alto, California: Consulting Psychologists Press

Morrison, G. R., Ross, S. M., & Kemp, J. E. (2001). *Designing effective instruction*, 3rd ed. New York: John Wiley.

Murray , H. A. (1938). Explorations in personality. New York: Oxford University Press, 1938.

Myers, R. E., & Fouts, J. T. (1992). A cluster analysis of high school science classroom environments and attitude toward science. *Journal of Research in Science Education, 29,* 929–937. doi: 10.1002/tea.3660290904

National Center for Education Statistics (NCES). (2011). State profiles. Retrieved from http://nces.ed.gov/nationsreportcard/states/

Nelson, P.M., Demers, J., & Christ, T.J. (2014). The Responsive Environmental Assessment for Classroom Teaching (REACT): The dimensionality of student perceptions of the instructional environment. *School Psychology Quarterly.*

Netemeyer, R.G.,Bearden, W.O., & Sharma, S. (2003). *Scaling procedures: Issues and applications*. Sage Publications, Incorporated.

Nolen, S.B. (2003). Learning environment, motivation, and achievement in high school science. *Journal of Research in Science Teaching*, *40(4),* 347-368. doi: 10.1002/tea.10080

Nye, B. Konstantopoulos, S., Hedges, L. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26(3),* 237-257. doi: 10.3102/01623737026003237

Palardy, G.J. & Rumberger, R.W. (2008). Teacher effectiveness in first grade: The importance of background qualifications, attitudes, and instructional practices for student learning. *Educational Evaluation and Policy Analysis*, *30,* 111-140. doi: 10.3102/0162373708317680

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2009). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest, 9(3),* 105-119. doi: 10.1111/j.1539-6053.2009.01038.x

Patrick, H., Ryan, A. M., & Kaplan, A. (2007). Early adolescents' perceptions of the classroom social environment, motivational beliefs, and engagement. *Journal of Educational Psychology, 99,* 83–98.

Payne, A. (1994). The role of home literacy environment in the development of language ability in preschool children from low-income families. *Early Childhood Research Quarterly, 9(3),* 427-440. doi: 10.10160/0885-2006(94)90018-3

Pianta, R. C., Steinberg, M. S., & Rollins, K. B. (1995). The first two years of school: Teacher-child relationships and deflections in children's classroom adjustment. *Development and Psychopathology*, *7,* 295-295.

Pianta, R. C., & Walsh, D. J. (1998). Applying the construct of resilience in schools: Cautions from a developmental systems perspective. *School Psychology Review, 27*(3), 407-417

Pianta,R.C. & Hamre, B.K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38(2),* 109-119. doi: 10.3102/0013189X09332374

Pomeroy, E. (1999). The teacher-student relationship in secondary school: insights from excluded students. *British Journal of Sociology of Education, 20*(4), 465-482.

Porter, A.C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher, 31(7),* 3-14. doi: 10.3102/0013189X031007003

Posamentier, A.S. & Jaye, D. (2006). What successful math teachers do, grades 6-12. Thousand Oaks, CA: Corwin Press

Quay, H.D. (1973). Special education: Assumptions, techniques, and evaluative criteria. *Exceptional Children, 40*, 165-170.

R Development Core Team (2011). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/

Raudenbush, S. W., & Bryk, A. S. (2006). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA [u.a.: Sage Publications].

Raudenbush, S., Bryk, A., Cheong, Y. F., Congdon, R., & du Toit, M. (2000). Hierarchical Linear and Nonlinear Modeling-HLM5. *Lincolnwood, IL: SSI Scientific Software*.

Raudenbush, S. W., et al. (2011). Optimal Design Software for Multi-level and Longitudinal Research (Version 3.01) [Software]. Available from www.wtgrantfoundation.org.

Raviv, A., Raviv, A., & Reisel, E. (1990). Teachers and students: Two different perspectives: Measuring social climate in the classroom. *American Educational Research Journal, 27,* 141–157.

Ready, D., & Wright, D.L. (2010). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: The role of child background and classroom context. *American Educational Research Journal, 48(2),* 335-360.

Ready, D.D., & Wright, D.L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: The role of child background and classroom context. *American Educational Research Journal, 48(2),* 335-360. doi: 10.3102/0002831210374874

Reschly, D.J. (2008). School psychology paradigm shift and beyond. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology V* (pp. 3-15). Bethesda, MD: National Association of School Psychologists.

Resnick, L.B., Wang, M.C., & Kaplan, J. (1973). Task analysis in curriculum design: A hierarchically sequenced introductory mathematics curriculum. *Journal of Applied Behavior Analysis, 6(4),* 679-709. doi:10.1901/jaba.1973.6-679.

Reddy, R., Rhodes, J. E., & Mulhall, P. (2003). The influence of teacher support on student adjustment in the middle school years: A latent growth curve study. *Development and psychopathology*, *15*(1), 119-138.

Rockoff, J. E. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *American Economic Review, 94*, 247–252.

Rosenshine, B. (1970). Evaluation of classroom instruction. *Review of Educational Research, 40(2),* 279-300.

Ross, J. A. (1992). Teacher efficacy and the effect of coaching on student achievement. *Canadian Journal of Education, 17(1),* 51-65.

Rowe, E.W., Kim, S., Baker, J.A., Kamphaus, R.W., & Horne, A.M. (2010). Student personal perception of classroom climate: Exploratory and confirmatory factor analyses. *Educational and Psychological Measurement, 70(5),* 858-879. doi: 10.1177/0013164410378085

Ruhl, K.L., Hughes, C.A., & Schloss, P.J. (1987). Using the pause procedure to enhance lecture recall. *Teacher Education and Special Education, 10(1),* 14-18.

Ryan, A.M., & Patrick H. (2001). The classroom social environment and changes in adolescents'' motivation and engagement during middle school. *American Educational Research Journal, 38(2),* 437-460. doi: 10.3102/00028312038002437

S. 844--112th Congress: Race to the Top Act of 2011. (2011). In *GovTrack.us (database of federal legislation).* Retrieved March 18, 2012, from http://www.govtrack.us/congress/bill.xpd?bill=s112-844

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

Samuels, S. J. (1979). The method of repeated readings. *The Reading Teacher, 41,* 756–760.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, *99*(6), 323-338.

Schunk, D.H. (1996). Goal and self-evaluative influences during children's cognitive skill learning. *American Educational Research Journal, 33(2),* 359-382. doi: 10.3102/00028312033002359

Shapiro, E. S. (2010). *Academic skills problems: Direct assessment and intervention* (3rd ed.). New York: Guilford.

Sinclair, B. B., & Fraser, B. J. (2002). Changing classroom environments in urban middle schools. *Learning Environments Research*, *5*(3), 301-328.

Sirin, S.R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75(3),* 417-453. doi:10.3102/00346543075003417

Skiba, R. J., Peterson, R. L., & Williams, T. (1997). Office referrals and suspension: Disciplinary intervention in middle schools. *Education & Treatment of Children*.

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of educational psychology*, *85*(4), 571.

Slavin, R.E. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of Educational Research*, *60(3),* 471-499.

Sroufe, A.L. (1997). Psychopathology as an outcome of development. Development and Psychopathology, 9, 251-268. doi: 10.1017/S0954579497002046

Steffe, L. P., Gale, J. E., & Alternative Epistemologies in Education Conference. (1995). *Constructivism in education*. Hillsdale, N.J: Lawrence Erlbaum.

Sternberg, R.J., Grigorenko, E.L., & Zhang, L. (2008). Styles of learning and thinking matter in instruction and assessment. *Perspectives on Psychological Science, 3(6),* 486-506.

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, *83*(10), 758–765.

Sugai, et al. (2000). Applying positive behavior support and functional behavioral assessment in schools. *Journal of Positive Behavior Interventions, 2(3),* 131-143. doi: 10.1177/109830070000200302

Sullivan, A. L., Klingbeil, D., & Van Norman, E. (2013). Beyond behavior: Multilevel analysis of the influence of sociodemographics and school characteristics on students' risk of suspension. *School Psychology Review*, *47*(1),

Supovitz, J. A., & Turner, H. M. (2000). The effects of professional development on science teaching practices and classroom culture. *Journal of research in science teaching*, *37*(9), 963-980.

Sutherland, K.S., Alder, N., & Gunter, P.L. (2003). The effect of varying rates of opportunities to respond academic requests on the classroom behavior of students with EBD. *Journal of Emotional Behavioral Disorders, 11(4),* 239-248. doi: 10.1177/10634266030110040501

Sutherland, K.S., Alder, N., & Gunter, P.L. (2003). The effect of varying rates of opportunities to respond to academic requests on the classroom behavior of students with EBD. *Journal of Emotional and Behavioral Disroders, 11(4),* 239-248. doi: 10.1177/10634266030110040501

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. *Journal of Special Education, 41(4),* 223-233. doi: 10.1177/0022466907310372

Taylor, B.M., Pearson, D.M., Peterson, D.S., & Rodriguez, M.C. (2004). The CIERA school change framework: An evidence-based approach to professional development and school reading improvement. *Reading Research Quarterly, 40(1)*, 40-69.

Taylor, B.M., Peterson, D.S., Pearson, D.P, & Rodriguez, M.C. (2002). Looking inside classrooms: Reflecting on the "how" as well as the "what" in effective reading instruction. *The Reading Teacher, 56(3),* 270-279.

Taylor, B. M., & Ysseldyke, J. E. (2007). *Effective instruction for struggling readers, K-6*. New York: Teachers College Press.

Tessmer, M. (1990). Environmental analysis: A neglected stage of instructional design. *Educational Technology Research and Development, 38(1),* 55-64.

Thorp, H., Burden, R.L., & Fraser, B.J. (1994). Assessing and improving classroom environment. *School Science Review, 75*, 107–113.

Tilly, W.D., III (2008). The evolution of school psychology to science-based practice: Problem solving and the three-tiered model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 17-36). Bethesda, MD: National Association of School Psychologists.

Titsworth, B.S. (2001). Immediate and delayed effects of interest cues and engagement cues on students' affective learning. *Communication Studies, 52(3),* 169-179. doi: 10.1080/10510970109388552

Treptow, M.A., Burns, M.K., & McComas, J.J. (2007). Reading at the frustration, instructional, and independent levels: Effects of student time on task and comprehension. *School Psychology Review, 36,* 159–166.

Tucker, J. A. (1989). Basic flashcard technique when vocabulary is the goal. Unpublished teaching materials, University of Tennessee at Chattanooga. Chattanooga, TN: Author.

Tyler R.W. (1949). *Basic principles of curriculum and instruction.* Chicago, IL: University of Chicago Press.

U.S. Congress (2001). *No Child Left Behind Act of 2001*. Washington, DC: Author.

VanDerHeyden, A. M. & Witt, J. C. (2008). Best practices in can't do/won't do assessment. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 131-139). Bethesda, MD: National Association of School Psychologists

VanTassel-Baska, J., Quek, C., & Feng, A.X. (2007). The development and use of a structured teacher observation scale to assess differentiated best practice. *Roeper Review, 29(2)*, 84-92. doi: 10.1080/02783190709554391

Walberg , H. J. , & Anderson , G. J. (1968). Classroom climate and individual learning. *Journal of Educational Psychology, 59,* 414-419. doi: 10.1037/h0026490

Waldrip, B., Reene, P., Fisher, D., Dorman, J. (2008). Changing students' perceptions of teacher interpersonal behaviors in science. *Research in Science Education, 38(2),* 213-235. doi: 10.1007/s11165-007-9045-x

Wang, M. C.; Haertel, G. D. & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research, 63(3),* 249–294.

Wang, M., & Holcombe, R. (2010). Adolescents' perceptions of school environment, engagement, and academic achievement in middle school. *American Educational Research Journal, 47(3),* 633-662.

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

Way, N., Reddy, R., & Rhodes, J. Students' perceptions of school climate during the middle school years: Associations with trajectories of psychological and behavioral adjustment. *American Journal of Community Psychology, 40(3-4),* 194-213. doi: 10.1007/s10464-007-9143-y

Way, N., Reddy, R., Rhodes, J. (2007). Students' perceptions of school climate during the middle school years: Associations with trajectories of psychological behavioral adjustment. *American Journal of Community Psychology, 40(3),* 194-213. 10.1007/s10464-007-9143-y

Wehby, J., & Lane, K. L. (2009). Classroom management. In A. Akin-Little, S. G. Little, M. A. Bray, & T. J. Kehle (Eds.), Behavioral interventions in schools: Evidence-based positive strategies (pp. 141–156). Washington, DC: American Psychological Association.

Wehby, J.H., & Lane, K.L. (2009). Behavioral interventions in schools: Evidence-based positive strategies. In S. Little, M. Bray, & T. Kehle (Eds.), School Psychology (pp. 141-156). Washington, DC, US: American Psychological Association. doi: 10.1037/11886-009

West, J., Denton, K., & Reaney, L. (2000). The kindergarten year: Findings from the early childhood longitudinal study, kindergarten class of 1998-99. Washington DC: U.S. Department of Education.

Westling, D.L. (2010). Teachers and challenging behavior. *Remedial and Special Education, 31(1),* 48-63. doi: 10.1177/0741932508327466

Wheldall, K., Beaman, R., & Mok, M. (1999). Does the Individualized Classroom Environment Questionnaire (ICEQ) measure classroom climate? *Educational Psychological Measurement, 59(5),* 847-854. doi: 10.1177/00131649921970170

White, K.R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin, 91(3),* 461-481. doi: 10.1037/0033-2909.91.3.461

Wigfield, A., Eccles, J.S., Mac Iver, D., Reuman, D.A., & Midgley, C. (1991). Transitions during early adolescence: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Development Psychology, 37(4),* 552-565. doi: 10.1037/0012-1649.27.4.552.

Yarbrough, J.L., Skinner, C.H., Lee, Y.U., & Lemmons, C. (2008). Decreasing transition times in a second grade classroom. *Journal of Applied School Psychology, 20(2),* 85-107. 10.1300/J370v20n02_06

Yarrow, A., Millwater, J., & Fraser, BJ (1997). Improving university and primary school classroom environments through preservice teachers' action research. *International Journal of Practical Experience in Professional Education, 1 (1),* 68-93.

Ysseldyke, J. E., & Christenson, S. (2002). *Functional assessment of academic behavior.* Longmont, CO: Sopris West.

Ysseldyke, J.E. & Burns, M. (2008). Functional assessment of instructional environments for the purpose of making data-driven instructional decisions. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology V* (pp. 410-427). Bethesda, MD: National Association of School Psychologists.

Ysseldyke, J.E., & Christenson, S.L. (1993). *TIES-II: The instructional environment system.* Longmont: Sopris West

Yu, C. Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes* (Doctoral dissertation, University of California Los Angeles)

# Appendix A

| **The Classroom Environment**<br>Those academic and behavioral supports in the classroom that are under the direct control of the teacher and supported by empirical research. | |
|---|---|
| **Component** | **Definition** |
| *Social and Behavioral Support* | The classroom management techniques used are effective for the student; classroom rules are explicitly communicated and consistently reinforced. |
| *Instructional Expectations* | Reasonable and high expectations are established and communicated to the student along with a shared belief that he or she will learn and achieve. Assignments and activities are provided with clear directions of reasonable length and students understand expectations to succeed. |
| *Cognitive Emphasis* | Concepts and strategies are taught to promote retention and understanding of key skills and concepts. |
| *Academic Engaged Time* | The student is actively engaged in responding to academic content; the teacher monitors the extent to which the student is actively engaged and redirects the student when the student is unengaged; time is used productively. |
| *Motivational Supports* | Effective strategies for heightening student interest and effort are used with the student. There is a positive tone in the classroom and frequent appropriate use of sincere praise and reinforcement. |
| *Instructional Presentation* | Instruction is presented in a clear and effective manner; the instructional lesson contains sufficient information for the student to understand the kinds of behaviors or skills that are to be demonstrated; there is substantive teacher/student interaction during the lesson. |
| *Strategic Feedback:* | The student receives relatively immediate and specific information on his/her performance or behavior; when the student makes mistakes, correction is provided. |
| *Instructional Differentiation* | The procedures and pace of instruction accommodate the individual student as necessary (to ensure understanding and success on skills when applied independently); the level of instruction is relevant in light of the student's experience. |
| *Curriculum Differentiation* | The content and skills that are taught and practiced are adjusted to accommodate the individual student as necessary (to maintain a zone of proximal development); instructional goals and materials are matched to the student's skills. |
| *Relevant Practice* | The student is given adequate opportunity to practice with appropriate materials and achieve a high success rate. Classroom tasks are clearly important in achieving instructional goals. |
| *Assessment of Progress* | There are explicit long term (annual or semester) and short-term goals (weekly). Student progress is assessed frequently (weekly quizzes, unit tests) and these data are used to instructional decisions. |

# Appendix B

| The Classroom Teaching Environment | |
|---|---|
| **Instructional Strategies** | |
| My teacher helps me learn ways to answer different kinds of questions. | Cognitive Emphasis |
| We learn tricks, strategies or shortcuts to learn and remember things. | Cognitive Emphasis |
| We learn and practice problem solving in class. | Cognitive Emphasis |
| My teacher explains things clearly. | Instructional Presentation |
| My teacher explains things in more than one way. | Instructional Presentation |
| My teacher tells me what he/she's going to teach before the lesson begins. | Instructional Presentation |
| **Progress Monitoring Strategies** | |
| My teacher and I set goals for my learning. | Instructional Expectations |
| My teacher explains how I am doing in class. | Assessment of Progress |
| We track how much we learn in class | Assessment of Progress |
| My teacher helps me make plans for how I'll do my work. | Cognitive Emphasis |
| **Differentiated Instruction Strategies** | |
| My teacher knows what subjects or skills are easier for me. | Instructional Match |
| My teacher gives extra review when I need it. | Instructional Match |
| My teacher helps me pick books or materials that are on my level. | Curricular Match |
| I have enough time to work on new things I learn. | Relevant Practice |
| **Formative Feedback Strategies** | |
| My teacher shows me how to correct my mistakes on my work. | Specific Feedback |
| My teacher comes to check my work when I am working alone. | Specific Feedback |
| When I make mistakes, my teacher helps me understand why | Mastery Learning (New) |
| My teacher explains why learning is important. | Motivational Strategies |
| My teacher cares more about how well I understand things we learn than how well I do on tests and quizzes | Mastery Learning (New) |
| **Positive Reinforcement Strategies** | |
| My teacher tells me when I do well in class. | Spec. Feedback |
| I am rewarded for doing good work in my class. | Motivational Strategies |
| My teacher uses praise or rewards for good behavior. | Motivational Strategies |
| My teacher says nice things about my work. | Motivational Strategies |
| My teacher tells me when I do a good job. | Social and Behavioral Support |
| **Academic Engagement Strategies** | |
| There are other learning activities to do when I finish my work early. | Academic Engagement |
| My teacher keeps me thinking during the lesson. | Academic Engagement |
| **General Perception Items** | |
| My teacher makes learning interesting. | Motivational Strategies |
| My teacher makes class fun. | Motivational Strategies |
| I like this class. | Motivational Strategies |
| If I'm upset, my teacher listens. | Social and Behavioral Support |

# Appendix C

## TEACHER SURVEY

**Demographic Information: (CIRCLE ONE)**

**Gender:** A. Male B. Female

**Ethnicity:** A. Hispanic/Latino B. American Indian or Alaskan Native C. Asian D. Black or African-American E. Native Hawaiian or Other Pacific Islander F. White G. Other

**Teaching Experience**: 1-3 years 4-10 years 10-15 years 15+ years

**REACT Questions**: Please place an X in the chosen response column for each statement.

| Questions | Agree | Mostly Agree | Mostly Disagree | No |
|---|---|---|---|---|
| The procedures for administering the REACT were practical. | | | | |
| The procedures for the feedback meeting were practical. | | | | |
| I found my students' ratings easy to interpret. | | | | |
| I believe my students' perceptions are meaningful. | | | | |
| The questions on the REACT are appropriate for middle school students. | | | | |
| The questions on the REACT cover a wide range of content. | | | | |
| The questions on the REACT provide information **that I can act on** to address students' **behavior problems.** | | | | |
| The questions on the REACT provide information **that I can act on** to address students' **academic problems.** | | | | |
| The strategies in the feedback form are appropriate for middle school students. | | | | |
| The strategies in the feedback form are clearly presented. | | | | |
| The strategies in the feedback form are useful. | | | | |

I would use the REACT on my own if it were available

**a.** Yes, with **no need** for additional support with data collection/data analysis
**b.** Yes, with additional support for data collection/data analysis
**c.** No