CHILD DEVELOPMENT

Child Development, July/August 2017, Volume 88, Number 4, Pages 1321–1337

Warmth and Demand: The Relation Between Students' Perceptions of the Classroom Environment and Achievement Growth

Lia E. Sandilos, Sara E. Rimm-Kaufman, and Julia J. Cohen
University of Virginia

Theory suggests that African American students benefit from warm and demanding teachers. This study examines the relation between students' perceptions of 634 teachers' warm demander characteristics and achievement growth in fourth and fifth grades ($M_{\text{student age}}$ = 9–11.5 years). Analyses explored whether relations were moderated by the proportion of African American students in the classroom or the ethnic match or mismatch between African American students and their teachers. Results indicated that students' perceptions of teachers' demand (challenge and control) related to student achievement growth. Findings showed a stronger relation between challenge and academic growth in classrooms with more African American students, but no significant findings were identified for ethnic match or mismatch.

In U.S. schools, teachers are charged with the task of establishing learning environments that foster measurable academic growth in diverse learners (Zaslow, Martinez-Beck, Tout, & Halle, 2011). As a result, education researchers strive to develop a deeper understanding of the instructional practices and interactions that are most beneficial to students. In particular, the persistent disparities in achievement that exist between minority and nonminority students has prompted research examining culturally responsive pedagogical practices that support ethnic minority children (Villegas & Lucas, 2002). Classroom processes and teacher–child interactions hold promise for reducing these disparities (Pianta & Walsh, 1996). Yet, prior research has indicated that the quality of instruction and teacher–child interactions is lower in low-income elementary schools, which have disproportionately higher numbers of ethnic minority students (National Center for Education and Evaluation, 2011; Pianta, Belsky, Houts, & Morrison, 2007). Given these patterns, it is essential that research continues to explore the social interactions and instructional patterns that can support positive developmental outcomes for ethnic minority students (Cabrera, Beeghly, & Eisenberg, 2012).

Warm demander pedagogy is one example of a culturally responsive teaching framework that describes effective interactions and instructional practices with African American students. The warm demander theory was developed from small-scale qualitative work with samples of both African American students and teachers, and posits that teachers who are high in both warmth and demand toward their students produce the best outcomes for African American students (Ford & Sassi, 2014; Ware, 2006). In theory, teachers who embody a high-warmth and high-demand teaching style promote note only a culture of academic achievement by managing their classrooms very efficiently but also conveying to students that they care for their well-being and hold high expectations for academic achievement (Ware, 2006). However, since the emergence of this theory from qualitative examinations of effective teachers, little empirical work has been done to test the relation between warm demander teachers and student achievement outcomes.

The present study addresses a problem at the intersection of two lines of research. One line of research measures and examines effective classroom instruction, by considering interactions between teachers and students (Pianta & Hamre, 2009). The second line of research examines warm demander

The data used in this study were collected as part of the Measures of Effective Teaching (MET) Project funded by the Bill & Melinda Gates Foundation (2009–2011). The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B130013 to the University of Virginia. The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education.

Correspondence concerning this article should be addressed to Lia E. Sandilos, Curry School of Education, University of Virginia, 268 Ruffner Hall 405 Emmet Street, Charlottesville, VA 22904. Electronic mail may be sent to lia.sandilos@gmail.com.

 0009-3920/2017/8804-0024
DOI: 10.1111/cdev.12685

teaching practices that benefit ethnic minority students, specifically African American students (Ware, 2006). Bridging these two lines of research has potential to create a deeper understanding of the classroom processes that are most beneficial and culturally responsive to students. To that end, this study explored the extent to which the warm demander teaching style relates to fourth- and fifth-grade student achievement growth in reading and math and considered the extent to which these practices are particularly important for African American students. Furthermore, this study explored the influence of teacher ethnicity on the relation between warm demander practices and African American students' achievement growth.

Teachers as Warm Demanders

The warm demander theory emerges from a broader literature on culturally responsive pedagogy, which explores effective teaching through a deeper understanding of learning styles and interaction patterns of students from diverse backgrounds (racial, ethnic, economic, linguistic; Gay, 2002). The term *warm demander*, coined by Vasquez (1989), refers to a teaching style in which teachers are nurturing or caring toward their students but do not lower academic standards or expectations and are effective disciplinarians. Ethnographic work on warm demanders focuses on effective instruction with African American students and emphasizes teachers' critical role in promoting educational equity by holding all students to a very high standard academically while maintaining strong caring relationships and effective classroom management (Ford & Sassi, 2014). Thus, warmth is defined as the teacher's ability to exhibit unconditional positive regard, convey a sense of caring for students' well-being, show authentic interest in children's lives, and demonstrate mutual respect (Bondy & Ross, 2008). Demand refers to the teacher's ability to challenge students through academically rigorous instruction and to insist that students exert effort and perform to a high standard. Demand also encompasses the teacher's ability to manage or control classroom behavior to minimize distractions and increase time engaged in learning (Bondy & Ross, 2008; Ware, 2006).

Warm demander practices have been found to be especially salient for African American students. African American high school students reported stronger trust in their teachers when they perceived teachers as both caring and holding high expectations (Gregory & Weinstein, 2008). Additionally, academically successful African American high school graduates from single-parent, low-income homes reported having supportive and warm mentors who challenged them to succeed (Williams & Bryan, 2013). African American students have been historically marginalized within the education system, and teachers may be able to use warm demander practices to counteract negative societal messages and to convey high expectations for success in school.

Moreover, several qualitative studies have suggested that African American teachers might better exhibit warm demander teaching style than Caucasian teachers because they draw upon a shared cultural history with African American students (e.g., Cholewa, Goodman, West-Olatunji, & Amatea, 2014; Ford & Sassi, 2014; Irvine, 2002; Ware, 2006). A shared ethnic or cultural background can contribute to positive student–teacher relationships because students have the opportunity to view their teacher as a role model who is racially similar to themselves (Dee, 2005). Furthermore, some theorize that African American students are able to perceive cultural commonalities between themselves and their teacher, contributing to a lower risk for negative expectancy effects and stereotype threat when an ethnic match exists (Dee, 2005). Teachers' perceptions of students may also be influenced by ethnic match or mismatch. For example, prior studies of elementary and middle school samples indicated that Caucasian teachers were more likely to report a higher prevalence of problems among ethnic minority students (e.g., inattention, lack of homework completion, difficulty following directions, disorganized home environment) than teachers of similar ethnic minority backgrounds (Dee, 2005; Rimm-Kaufman, Pianta, & Cox, 2000). Given that warm demander theory originated from observations of African American teachers and students, ethnic match may be an important consideration in understanding the practices of warm demanding teachers.

Important qualitative work generated the warm demander theory based on observations of African American students and teachers. An essential next step for the warm demander theory is to better understand the influence of these practices on student achievement outcomes in a large and diverse sample of students. Validation of this pedagogical style is important to improve teacher preparation and professional development. Current views of cultural competence emphasize that teachers hold high expectations for students regardless of students' background characteristics (Ladson-Billings,

2009). Enhanced empirical understanding of the warm demander style may provide important nuance for future recommendations. Specifically, establishing links between warm demander teaching style and student academic growth may lead to interventions to reduce racial academic disparities.

Unfortunately, there is no quantitative measure of warm demandingness. However, the widely used Tripod Survey (Ferguson, 2008) is a valid tool that can be used reliably to operationalize key facets of the theory. The Tripod uses student-report data to measure teachers' level of academic challenge, classroom control, and care—all aspects of the warm demander teaching style.

Considering the Student Perspective

Students' perception of their experiences in the classroom provides a unique lens for measuring classroom quality. Two students sitting in the same classroom may have very different experiences. Students' experiences vary because each student constantly appraises his or her immediate context or environment (e.g., classrooms, teachers, peers). These subjective appraisals lead students to either engage in or avoid the available learning opportunities (Skinner, Kindermann, Connell, & Wellborn, 2009). When measures are designed well and implemented properly, surveying students about their perceptions of the classroom environment can provide critical information about how students perceive their teachers' ability to meet their needs (Brock, Nishida, Chiong, Grimm, & Rimm-Kaufman, 2008; Kane & Staiger, 2012). Surveying students is more common in secondary and higher education; however, the use of student perception measures has gained credibility as a tool for understanding upper elementary classrooms as well (Downer, Stuhlman, Schweig, Martínez, & Ruzek, 2014; Polikoff, 2014).

Students' views of classrooms reflect the actual characteristics of the classroom experience, as well as students' expectations and past experiences. The student perspective yields valuable information about pedagogical practices and facilitates understanding of the diversity of views students have about their schooling (Howard, 2001). In the present study, the Tripod was used to measure three elements of warm demander teaching: challenge, control, and care (Ferguson, 2010).

Challenge

Challenge, one aspect of demand, refers to the rigor of instructional content as well as the amount of effort students are expected to expend on learning (Ferguson, 2010). Related terms include *teacher expectations*, *academic press*, or *demandingness* (Goddard, Sweetland, & Hoy, 2000; Lee, 2012; Lee & Smith, 1999). One common element among these various constructs is teachers' facilitation of academically rigorous learning environments in which students are expected to exert the effort needed for academic success and to persist in the face of difficulty (Ferguson, 2010). When teachers challenge their students successfully, the teachers' expectation for thoughtful work and high effort will convey to students they are *all* competent and capable of performing to a high standard (Fulmer & Turner, 2014; Stipek, 2002). Prior research indicates that teachers' expression of high expectations and facilitation of a challenging environment influence academic performance directly (Brophy & Good, 1970; Goddard et al., 2000; Rosenthal & Jacobsen, 1968). For example, elementary and middle school-age students whose teachers held high expectations for the entire class made twice as much reading growth in a year compared to students of similar ability in classrooms with low-expectation teachers (Rubie-Davies, 2007). In a large sample of urban elementary schools with more than 50% of the student sample being low income and African American, teachers' self-reported academic press accounted for approximately half of the variance in reading and math achievement between schools (Goddard et al., 2000).

Students' perceptions of teachers' performance expectations may be integral to the relation between academic challenge and outcomes. Good (1981) outlined a model of teacher expectations and student perceptions: (a) the teacher expects certain student behaviors or academic performance, (b) the teacher behaves in specific ways toward students based on those expectations, (c) teacher behavior alters students' academic self-concept, (d) over time students conform to the behaviors that are expected of them. Thus, when teachers demonstrate class-wide expectations for academic excellence, students will be more likely to exhibit behavior consistent with those expectations.

Existing research also suggests that the association between teacher expectations and student achievement may be stronger for African American students than for their Caucasian peers (Jussim, Eccles, & Madon, 1996). By fourth grade, students have the ability to recognize that teachers can hold differential beliefs about their academic ability and that those expectations may relate to students' group membership (McKown & Weinstein, 2003).

For example, one study examined elementary students' academic trajectories to determine if their present teacher had lower expectations for performance than was reasonable given their previous academic track record. Results indicated that teachers who held lower expectations, or underestimated student ability, had a more deleterious effect on the academic performance of African American students than their Caucasian peers (McKown & Weinstein, 2002). As the Good (1981) model posited, expectations can have a powerful influence on student performance. The findings from McKown and Weinstein (2002) underscore the ways in which high expectations may be especially supportive for African American students. In addition to challenge, it is important to understand how classroom control, the second aspect of *demand* in warm demander teaching, influences outcomes.

Control

Teachers who demonstrate effective management or control of the classroom establish clear routines and have higher levels of student productivity (Pianta & Hamre, 2009). Classroom control is characterized by teachers' ability to redirect off-task or disruptive behavior in ways that do not upset students' focus and the flow of a lesson (Ferguson, 2010). Generally, teachers who take a proactive approach to management (i.e., use clear guidance about behavioral expectations, take time to establish routines and classroom rules) tend to have more well-managed classrooms across grade levels and school contexts (Woolfolk Hoy & Weinstein, 2011). Positive management strategies can help to improve students' ability to regulate their own behavior (Rimm-Kaufman, Curby, Grimm, Nathanson, & Brock, 2009), which, in turn, supports academic achievement (e.g., Ponitz, Rimm-Kaufman, Grimm, & Curby, 2009). Students' perception of teachers' classroom management is critical to examine with ethnic minority student populations given that research has revealed inequitable disciplinary outcomes for African American students, and for male students in particular (Gregory & Weinstein, 2008; Silva, Langhout, Kohfeldt, & Gurrola, 2015). In light of these disparities, it is valuable to gain deeper insight into African American students' perceptions of classroom management and how those perceptions relate to outcomes. However, limited research has been done in this area.

Most research examines teachers' perceptions of their own classroom management (e.g., Cluniess-Ross, Little, & Kienhuis, 2008; Mitchell, Bradshaw, & Leaf, 2010) or observed indicators of high-quality classroom management strategies (e.g., Pianta & Hamre, 2009; Rimm-Kaufman et al., 2009). A few studies have explored students' perceptions of classroom management and their relation to academic outcomes. In one qualitative study of middle and high school students, aspects of classroom management that were most salient to students included their teachers' ability to clearly convey rules, expectations, and consequences for misbehavior. The degree to which the teacher was caring also emerged, warranting further discussion below (Cothran, Kulinna, & Garrahy, 2003).

Care

Caring teachers, by definition, establish a classroom climate in which students feel emotionally safe and sense that their teacher is concerned for their well-being and future success. Caring teachers not only demonstrate positive interpersonal interactions with students, but they also foster an environment in which students feel they can take academic risks (Ferguson, 2010; Reyes, Brackett, Rivers, White, & Salovey, 2012; Rimm-Kaufman & Chiu, 2007).

Caring teachers may be particularly beneficial to children who experience social inequities in the education system (e.g., low income background, membership in a minority group). Howard (2001) found that African American elementary students preferred learning environments in which they perceived their teacher as caring. Roorda, Koomen, Spilt, and Oort (2011) found that the magnitude of the relation between teacher–child relationships and achievement in a classroom increased as the number of students from low-socioeconomic status and ethnic minority backgrounds increased. Students' perceptions of teachers' care has also been linked to engagement and motivation in elementary and secondary grades (e.g., Klem & Connell, 2004; Wentzel, 1997). Less is known about the direct relation between students' perceptions of teachers' caring behaviors, or *warmth*, and students' academic growth in math and reading in upper elementary grades, particularly when examined simultaneously with teachers' level of demand (i.e., challenge and control).

Developmental Significance

The fourth and fifth grades, which span middle childhood to early adolescence, are important years in which to examine perceptions of the learning

environment. During this phase of development, students transition physically, emotionally, and psychologically into adolescence and also prepare for an environmental transition from elementary to middle school. Children's sense of identity and need for autonomy advance, and social relationships and interactions with peers and adults change, as well (Eccles, 1999). As children forge their individuality, race and culture often come to the forefront of their experiences and they begin to understand their own cultural background within a broader social context (McKown & Weinstein, 2003).

Children's academic self-concept is also forming rapidly during this stage and their interactions with teachers and experiences of success or failure within the school system have greater leverage in shaping their long-term identity as a student (Eccles, 1999). For some children, ethnic identity may be signaling one message of what is expected of them in school while educators may be conveying a different message, particularly through the expression of academic expectations. As such, teachers' warm demander style has the potential to be vital in helping students at this developmental stage establish a positive academic identity and, in turn, support greater achievement. Taken together, this sensitive phase of development includes a complex interplay of personal development, ethnic background, and classroom experiences.

Goals of the Current Study

The present study tests two propositions of the warm demander theory: (a) that warm and demanding interactions relate to positive developmental outcomes for fourth- and fifth-grade African American students, and (b) that African American teachers may be particularly effective warm demanders (Irvine, 2002). We test the warm demander theory by examining the combination of challenge, control, and care and its relation to achievement growth on both high- and low-stakes math and reading assessments. In doing so, we rely on the students' perception of challenge, control, and care because of the unique perspective they bring to understanding day-to-day classroom life. We pursued three research questions: (a) Do students' perceptions of teachers as warm demanders, as measured by challenge, control, and care relate to student achievement growth? (b) Is the relation between students' perceptions of teachers' warmth and demand and student achievement growth moderated by the proportion of African American students in the classroom? (c) Is the relation between students' perceptions of teachers' warmth and demand and student achievement growth in a subsample of classrooms with predominantly African American students (> 80%) moderated by the ethnic match or mismatch between an African American teacher and students?

We hypothesized that students' perceptions of their teachers as warm demanders would relate positively to student achievement growth in fourth- and fifth-grade elementary classrooms. Consistent with the warm demander theory, we hypothesized that the proportion of African American students in the classroom would moderate this relation, such that the warm demander characteristics would be more important for African American students' academic growth than non-African American students. We also hypothesized that the presence of an African American teacher (ethnic match) would strengthen the relation between warm demander characteristics and academic growth in classrooms with predominantly African American students. We expected ethnic mismatch would attenuate this relation.

Method

Participants

This study used data from the 1st year of the 2-year Measures of Effective Teaching (MET) Project, a large-scale observational study of classroom teaching conducted by the Bill & Melinda Gates Foundation (2009–2011). The primary purpose of MET study was to evaluate existing measures of teaching quality and instructional effectiveness (Kane & Staiger, 2012).

Full Sample

Participants in the present study included a total of 634 fourth- (n = 320) and fifth-grade (n = 314) teachers from five large districts located in New York, North Carolina, Florida, Tennessee, and Colorado. Teachers included in the sample were considered *generalists* because they taught all major subjects to their classroom students. The majority of teachers were female, and a little more than half had a master's degree or higher, with a wide range of experience teaching in their district (M = 6 years, range = < 1–34). The ethnic composition of teachers was largely African American and Caucasian (Table 2).

African American Subsample

A subsample of 223 fourth- (n = 107) and fifth-grade (n = 116) classrooms identified as having *predominantly* (80%–100%) African American students were used to explore the interaction of teacher–student ethnic match or mismatch with student perception ratings. For the subsample, the majority of teachers were female, nearly three-quarters had a master's degree or higher with an average of 5 years of experience teaching in their district (range = 1–19 years). The ethnic composition of teachers was three-quarters African American and nearly one-quarter Caucasian (Table 2).

Demographic characteristics of the classroom compositions of the full sample and African American subsample are also included in Table 2. Average age of students in a classroom ranged from 9 to 11.5 years (M = 9.97) in the full sample, and 9.20–11.5 years (M = 10.04) in the African American subsample.

Measures

Tripod 7Cs Student Perceptions Survey

The Tripod 7Cs (Ferguson, 2008) is a 36-item survey designed to capture students' perspective of the classroom environment through seven composites: challenge, control, care, clarify, captivate, confer, and consolidate. All items are coded as 1 (*no, never*), 2 (*mostly not*), 3 (*maybe/sometimes*), 4 (*mostly yes*), and 5 (*yes, always*). Composite scores are calculated by averaging item-level scores. We focus on challenge, control, and care composites because of their alignment with characteristics of warm demander teachers. Correlations among the three composites ranged from .44 to .61 (Table 3).

The challenge composite consists of four items tapping into the rigor of instructional content and the level of hard work expected of students in the classroom. Item examples include, "In this class, my teacher accepts nothing less than our full effort" and "My teacher pushes everybody to work hard." The challenge items demonstrated high internal consistency (α = .79) and moderate item-level correlations (r = .37–.69). The control composite consists of four items inquiring about teachers' ability to maintain a well-behaved and productive classroom. Item examples include, "Students in this class behave the way my teacher wants them to" and "Our class stays busy and doesn't waste time." One item, "Students behave so badly in this class it slows down our learning," was reverse scored. The control composite exhibited high internal consistency (α = .80) and moderate item-level correlations (r = .35–.64). The care composite consists of seven items that ask students about their teachers' level of care and emotional supportiveness in the classroom. Examples of items include, "I like the way my teacher treats me when I need help" and "My teacher in this class makes me feel that he or she really cares about me." The care composite had high internal consistency reliability (α = .92) and moderate to high correlations among items (r = .44–.80). A full list of items is included in Table 1. Individual student scores were aggregated to create a classroom level score for each of the challenge, care, and control composites to prepare for analyses.

Achievement Growth Measured by Value-Added Estimates

Student achievement growth on low- and high-stakes reading and math assessments was measured through value-added models (Kane & Staiger, 2012). MET researchers used student assessment data and background information to create value-

Table 1
Tripod Items for Challenge, Control, and Care

Items
Challenge
My teacher pushes everybody to work hard.
My teacher pushes us to think hard about things we read.
In this class, my teacher accepts nothing less than our full effort.
In this class, we have to think hard about the writing we do.
Control
Students in this class behave the way my teacher wants them to.
Our class stays busy and does not waste time.
Students behave so badly in this class that it slows down our learning (reversed).
Everybody knows what they should be doing and learning in this class.
Care
I like the way my teacher treats me when I need help.
My teacher in this class makes me feel that he or she really cares about me.
The teacher in this class encourages me to do my best.
My teacher gives us time to explain our ideas.
My teacher seems to know if something is bothering me.
If I am sad or angry my teacher helps me feel better.
My teacher is nice to me when I ask questions.

Note. Items are scored on a range of 1 (*no, never*) to 5 (*yes, always*).

added scores for teachers. A teacher's value-added estimate was computed for each of the four student assessment measures and standardized across the entire sample of teachers. Value-added scores for each assessment were used as key outcomes. The data included within a value-added estimate consisted of (a) students' current assessment scores in a particular measure, (b) their scores in a measure of that content domain from a year prior, (c) the aggregate classroom score on that content domain assessment in the students' classroom a year prior, (d) individual student background variables, and (e) student background variables averaged across the classroom (White & Rowan, 2013). Student background variables included ethnicity, English language learners (ELL) status, age, gender, special education status, gifted status, and receipt of free or reduced lunch. Students' prior year state test scores were used as an estimate of prior achievement for value-added scores in the MET-selected assessments (Stanford 9 Open-Ended Reading Assessment [SAT-9] and Balanced Assessment in Mathematics [BAM]) because those measures were administered exclusively during the MET study. Teachers received an aggregate value-added score for each of the four assessment measures. More information regarding the calculation of MET value-added estimates can be found in the MET User's Guide (White & Rowan, 2013).

Value-added scores are interpreted as the teachers' ability to facilitate growth in an academic area, conditional on the composition of students in a classroom. A standardized value-added score of zero would indicate that students are performing as expected given prior achievement and background data, a negative score would indicate that students are performing lower than expected and a positive score would indicate that students are performing higher than expected. As with other measures of educational accountability and effectiveness, there are cautions surrounding value-added models, such as the limitations of using a single number to quantify teacher effectiveness and the potential instability of scores (Braun, 2005). Despite these cautions, these models are commonly used in teacher evaluations and have been linked empirically to long-term student outcomes (e.g., college attendance, job salary; Chetty, Friedman, & Rockoff, 2011).

The following section describes the four individual assessments used to create the value-added scores. MET researchers designated the BAM and SAT-9 as low-stakes assessments and the state math and English tests as high-stakes assessments. The latter are high-stakes assessments because data from these tests are used to make important decisions about students and teachers (von der Embse, Kilgus, Solomon, Bowler, & Curtiss, 2015).

Balanced Assessment in Mathematics. The BAM is an open-ended mathematics assessment that requires written mathematical responses from students. The BAM is designed to test students' conceptual knowledge, higher level reasoning, and problem-solving skills. The BAM measures students' skill at modeling or formulating problems, transforming and manipulating mathematical formulas, inferring or drawing conclusions from tasks, and communicating about mathematics findings. The assessment contains four to five tasks that take approximately 1 hr to complete.

Stanford Open-Ended Assessment in Reading (SAT-9). The SAT-9 is a measure of students' reading comprehension of various types of text (fictional, nonfictional, and practical material such as newspaper stories or advertisements). Students participating in the MET study were presented with one narrative reading passage and asked to respond to nine open-ended comprehension questions. The SAT-9 was administered within a single class period.

State standardized tests in mathematics (MATH) and English language arts. State standardized tests in MATH and English language arts (ELA) varied by district. The standardized test formats were typically multiple choice. Only rank-based *z* scores for the state standardized test measures are provided in the MET data set.

Demographic Data

A demographic survey was used to collect relevant background data on participating teachers. Local district administrative data were used to derive aggregate classroom demographic information for this study (Kane & Staiger, 2012). Analyses included the proportion of students identified as African American students (i.e., selected to test the warm demander theory), male, receiving special education, and ELL.

Procedure

Participating districts were recruited through an opportunity sampling procedure, which resulted in six large districts across the country agreeing to participate in the MET study. One district was excluded from the present study because it did not have participating fourth- and fifth-grade classrooms. Once districts were identified, school

principals were recruited and the principals identified eligible teachers. To be an eligible participant, the teacher had to be part of a grade level or subject area that had at least two other teachers to form an "exchange group." The requirement of an exchange group was established so that students could be randomly assigned to one of three teachers in the 2nd year of the study. Alternative schools (special education, vocational schools, etc.) were excluded. Teachers who engaged in team-teaching situations in which it would be difficult to link students to one teacher also were excluded.

Incentives were used for recruitment and retention. Schools received a total of $1,500, the use of which could be determined by the schools' administration, as well as $500 a year to pay for a project coordinator. Video recording equipment used for MET classroom observations was donated to the schools at the conclusion of the study. Each participating teacher received a $1,500 incentive for participating in the study ($1,000 at the start of the study and $500 at the conclusion).

Data used in the present study were collected during spring 2010 of the 1st year of the MET study. No randomization occurred during the 1st year of the MET study. The BAM and SAT-9 were administered between April and June 2010. The state standardized testing occurred between March and June 2010. Students were administered the Tripod 7Cs survey between February and June 2010. The measures selected by MET researchers (i.e., BAM, SAT-9, and Tripod) were administered to all consenting students within participating teachers' classrooms. The MET data set is restricted because it contains sensitive and potentially identifying information about school districts. The findings reported in this article follow the secure data reporting policies outlined by MET. Additional information about the MET data set can be found in MET reports (Kane & Staiger, 2012).

Data Analysis

Regression models were conducted in MPlus version 7 (Muthén & Muthén, 1998–2012) with all data aggregated to the classroom level. The average number of schools included in each analysis ranged from 104 to 108, with an average cluster size of approximately six classrooms per school. Intraclass correlations (ICCs) were examined for the outcomes and variables of interest to determine the need for nesting classrooms within schools. ICCs ranged from .08 to .22 revealing a shared variance among teachers within the same school. Thus, the TYPE = COMPLEX function in Mplus was utilized in each analysis to account for the nonindependence of observations through the use of robust standard errors.

Of the 634 generalist classrooms included in the analysis, 60 teachers had some amount of missing data. Of the 223 majority African American classrooms, 37 teachers had some amount of missing data. Based on the missing at random (MAR) assumption (Enders, 2010), all models were estimated using full information maximum likelihood. Although the MAR assumption could not be directly tested, it was believed to be reasonable for these data because no significant differences were found between participants with missing data and participants with complete data. One exception to this was the proportion of students within a classroom receiving free or reduced price lunch, a proxy for socioeconomic status, which was not provided for one entire district. As such, this variable was included as an auxiliary variable in the analyses to adjust standard errors for systematically missing data. Grade level and district fixed effects were included in the analysis as dummy variables. Following the MET secure data requirements, district fixed effects estimates were not reported in the Results section or Tables. Other covariates consisted of the proportion of students identified as male, ELL, and special education status. These variables were included because they demonstrated significant correlations with either the Tripod focal predictors or outcome variables.

To address the warm demander hypothesis, the percent of African American students per classroom was included in analyses as a continuous proportion variable. To explore the influence of ethnic match or mismatch between teachers and students, a subsample of classrooms with 80% or more African American students was examined. If the teacher in this subsample was also African American, then that classroom was coded as a dichotomous teacher–student ethnic match variable (*match* 1 = yes, 0 = no). If the teacher was not African American, then the classroom was coded as dichotomous ethnic mismatch variable (*mismatch* 1 = yes, 0 = no).

For each regression model, variables were entered in a stepwise fashion to first determine if main effects were present and then, to determine the presence of significant interaction effects. To address the first research question, the three focal Tripod composites were used to predict value-added scores. Classroom proportion of African American students was then included as a

moderator in each model to explore the second research question. To examine the third research question, Tripod composites were used to predict value-added scores in a smaller subsample of classrooms with predominantly African American students, and match or mismatch variables were entered as moderators to determine if there was a significant interaction between Tripod predictors and teacher–student ethnic match or mismatch.

Results

Raw score means, standard deviations, and ranges are reported for the Tripod composites and BAM and SAT-9 value-added outcomes (Table 2). State-standardized testing data were only available in standardized form. Correlations between Tripod composites and value-added outcomes ranged from .08 to .22 (Table 3). Within each regression model, all data were *z* scored to aid with interpretation of the findings.

To explore the influence of challenge, control, and care on value-added scores in the full sample, the main effects of those variables were tested in an initial set of regression models (Model 1, Table 4). In the model predicting BAM, the classroom aggregate rating of control related to higher value-added scores ($\beta = .14$, $p < .01$). For SAT-9, the classroom aggregate ratings of challenge ($\beta = .11$, $p < .05$) and control ($\beta = .16$, $p < .05$) related to higher value-added scores. Similarly for the MATH value-added model, average student perceptions of challenge ($\beta = .18$, $p < .001$) and control ($\beta = .15$, $p < .01$) related to higher value-added scores. In the model predicting ELA, aggregate classroom ratings of challenge ($\beta = .18$, $p < .01$) and control ($\beta = .18$, $p < .01$) related to higher value-added scores.

Notably, when included in models with challenge and control variables, the main effect of care was consistently nonsignificant. However, when care was entered into the model alone, it was significantly and positively related to the value-added outcomes. This finding indicates that the influence of the care composite was diminished in the presence of challenge and control composites. Because the warm demander theory posits that these teaching qualities should be present simultaneously in the classroom, models with all three Tripod composites entered together were selected as the final models.

To explore the moderation effect of the proportion African American students in a classroom on the relation between Tripod ratings and value-added scores, interaction effects were tested in a second set of regression models (Model 2, Table 4). For BAM, there was a significant positive interaction between challenge and the proportion of African American students in a classroom ($\beta = .13$, $p < .05$). Simple slopes were examined to determine the nature of the interaction. At high levels (+1 *SD* above the mean) of proportion of African American students, the slope was statistically significant ($b = .05$, $SE = .02$, $t = 2.47$, $p = .014$) such that the association between challenge and BAM value-added scores was stronger in classrooms with higher numbers of African American students (Figure 1).

Likewise for SAT-9 value-added scores, a significant positive interaction was found between challenge and the proportion of African American students in a classroom ($\beta = .12$, $p < .05$). An examination of simple slopes revealed significant slopes at average ($b = .03$, $SE = .02$, $t = 2.06$, $p = .040$) and high proportions ($b = .07$, $SE = .03$, $t = 2.50$, $p = .013$) of African American students (Figure 1). The findings indicated that the strength of the association between challenge and SAT-9 value-added scores increased as the proportion of African American students increased.

No significant interactions were identified in the model prediction MATH value-added scores. As such, only the main effects of this model were interpreted.

Table 2
Percentages of Sociodemographic Characteristics for Full Sample (N = 634) and African American Subsample (n = 223)

Characteristic	Full sample (%)	African American subsample (%)
Teacher	—	—
Female	91	91
African American	40	75
Caucasian	54	22
Latino(a)	5	< 1
Master's degree or higher	54	75
Classroom	—	—
Female	50	50
African American	52	93
Caucasian	17	1
Latino(a)	23	5
Asian	5	< 1
English language learners	15	4
Special education	9	5
Free and reduced price lunch	44	27

Note. Asian and Native American teachers comprised < 1% of the full teacher sample. Native American students comprised < 5% of the full classroom sample.

Table 3
Correlations Among Tripod Predictors and Value-Added Outcomes

	1	2	3	4	5	6	7
1. Challenge	—	.44***	.44***	.10*	.15***	.21***	.22**
2. Control		—	.61***	.15**	.19***	.20***	.22***
3. Care			—	.08	.12**	.14**	.13**
4. Balanced Assessment in Mathematics				—	.44***	.36***	.34***
5. SAT-9					—	.21***	.35***
6. MATH						—	.56***
7. English language arts							—
Raw *M* (*SD*)	16.68 (1.22)	13.94 (1.60)	29.06 (2.42)	17.97 (5.88)	106.23 (86.97)	—	—
Range	12.50–16.68	8.87–19.67	19.51–34.63	3.90–33.54	5.95–521.97	—	—

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

For ELA, a significant positive interaction was found between challenge and the proportion of African American students ($\beta = .14$, $p < .01$) in a classroom. Simple slopes showed significance at average ($b = .04$, $SE = .01$, $t = 2.96$, $p = .003$) and high ($b = .07$, $SE = .02$, $t = 3.11$, $p = .002$) proportions of African American students. The association between challenge and ELA value-added scores increased as the number of African American students increased (Figure 1).

In a subsample with 80% or more African American students per classroom (Table 5), analyses revealed a main effect of challenge for BAM ($\beta = .19$, $p < .05$), SAT ($\beta = .21$, $p < .01$), MATH ($\beta = .23$, $p < .01$), and ELA ($\beta = .32$, $p < .001$). No significant interactions were found between Tripod scores and teacher–student ethnic match or mismatch; only the main effects of the care, control, and challenge on value-added scores were interpreted.

An alternative angle for predicting achievement growth is to consider variability in students' perceptions of the classroom environment, in addition to central tendency. For example, if 20 children reporting on their teacher all have disparate views, that classroom experience is potentially different from a classroom in which 20 children convey similar views of their teacher. As such, we explored this angle through the use of the coefficient of variation. The coefficient of variation is a statistical technique used to measure variation in an observed variable (Bedeian & Mossholder, 2000) and was used as an indicator of variability in students' ratings of challenge, control, and care. Main effect findings indicated that higher variability in student perceptions of demand (challenge and control) was associated with lower achievement. Regarding the interactions, as the variability in students' perceptions of their teacher's level of challenge increased, value-added achievement scores decreased. The negative relation between variability in challenge and achievement was stronger in classrooms with a higher proportion of African American students.

Discussion

The results of this study partially support the warm demander theory, with *demand* emerging as a particularly important construct as it significantly and positively related to students' academic growth. Main effect analyses indicated that students who perceived their teachers as more challenging and higher in control showed greater growth on both high- and low-stakes math and reading assessments. Moderation analyses indicated that the positive relation between challenge and academic growth was stronger in classrooms with greater numbers of African American students. Findings showed that teacher–student ethnic match or mismatch was not associated with student academic growth in this sample.

Consistent with our hypotheses, student perceptions of an academically rigorous and well-managed classroom environment contributed to academic growth in fourth- and fifth-grade classrooms on both high- and low-stakes achievement measures in math and reading. Prior research corroborates the importance of challenge and control (e.g., Pianta & Hamre, 2009; Rubie-Davies, 2007). Appropriately challenging learning contexts, in which teachers convey high expectations and offer academically rigorous content, may contribute to students' self-perceptions and internalizations of their own ability to exert effort and achieve

Table 4
Main Effects of Tripod and Moderation Effects of Proportion African American Students to Predict Value-Added Scores

	Low-stakes MET assessments				High-stakes state assessments			
	Model 1 BAM (n = 575) R^2 = .03*	Model 2 BAM (n = 574) R^2 = .04**	Model 1 SAT-9 (n = 578) R^2 = .05*	Model 2 SAT-9 (n = 577) R^2 = .06**	Model 1 MATH (n = 594) R^2 = .07***	Model 2 MATH (n = 593) R^2 = .07***	Model 1 ELA (n = 594) R^2 = .07**	Model 2 ELA (n = 593) R^2 = .09**
Fourth grade	.00 (.05)	.01 (.05)	.02 (.04)	−.02 (.04)	.00 (.05)	.01 (.05)	.02 (.05)	.03 (.05)
Proportion special education	.02 (.11)	.06 (.04)	.03 (.05)	.03 (.05)	−.02 (.04)	−.02 (.04)	.02 (.04)	.02 (.05)
Proportion ELL	.03 (.07)	.02 (.06)	.03 (.05)	.03 (.05)	.04 (.04)	.04 (.04)	.03 (.05)	.02 (.04)
Proportion Male	.00 (.05)	.02 (.05)	−.02 (.05)	.00 (.04)	.02 (.04)	.03 (.04)	.01 (.04)	.03 (.06)
Challenge	.06 (.06)	.07 (.06)	.11 (.05)*	.12 (.06)*	.18 (.05)***	.19 (.01)***	.18 (.06)**	.19 (.01)***
Control	.14 (.05)**	.14 (.05)**	.16 (.07)*	.16 (.07)*	.15 (.05)**	.15 (.01)**	.18 (.06)**	.17 (.01)**
Care	−.03 (.06)	−.04 (.06)	−.03 (.05)	−.04 (.05)	−.04 (.06)	−.05 (.02)	−.06 (.07)	−.07 (.02)
African American	—	.02 (.08)	—	−.02 (.09)	—	−.01 (.02)	—	−.01 (.01)
Challenge × African American	—	.13 (.06)*	—	.12 (.06)*	—	.07 (.02)	—	.14 (.01)**
Control × African American	—	−.01 (.05)	—	−.07 (.06)	—	.01 (.01)	—	.03 (.01)
Care × African American	—	−.04 (.06)	—	−.02 (.05)	—	−.02 (.02)	—	−.04 (.01)

Note. Standardized estimates are reported with the standard error in parenthesis. District was also controlled for as a fixed factor in the analysis. Model 1 = main effects model; Model 2 = moderation model; African American = classroom proportion variable; BAM = Balanced Assessment in Mathematics; ELA = English language arts; ELL = English language learners; MET = Measures of Effective Teaching. *p < .05. **p < .01. ***p < .001.

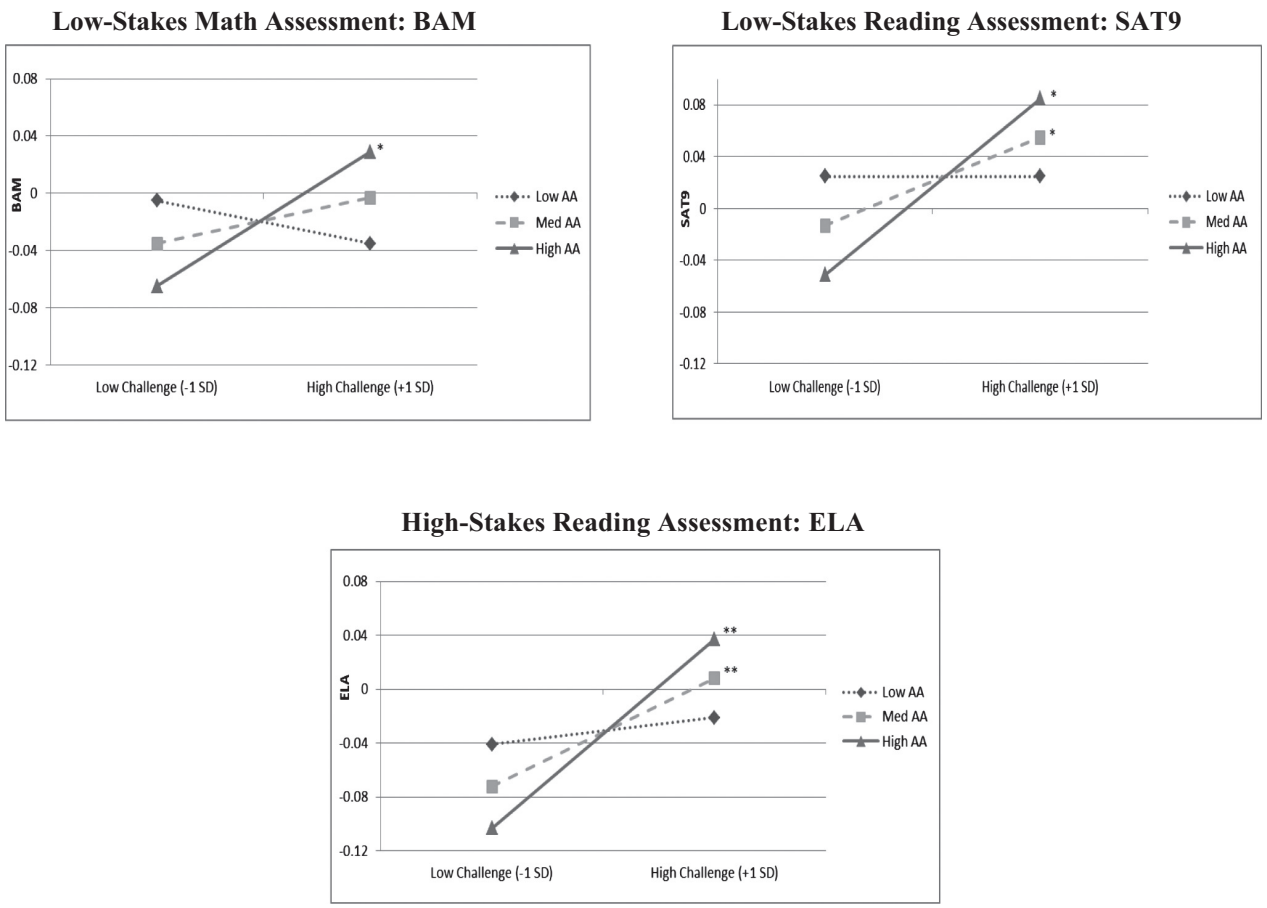

Figure 1. Interaction between challenge and proportion of African American (AA) students to predict the low- and high-stakes assessments.
*$p < .05$. **$p < .01$.

Table 5
Main Effect of Tripod in the African American Student Subsample

	Low-stakes MET assessments		High-stakes state assessments	
	BAM (n = 191) R^2 = .04	SAT-9 (n = 192) R^2 = .09	MATH (n = 199) R^2 = .10**	ELA (n = 199) R^2 = .14**
Fourth grade	.09 (.09)	.02 (.07)	.05 (.09)	.05 (.09)
Proportion special education	.07 (.06)	.00 (.08)	.01 (.06)	.01 (.09)
Proportion ELL	−.03 (.25)	−.22 (.18)	.11 (.17)	.08 (.19)
Proportion male	.03 (.10)	−.02 (.06)	.02 (.07)	.07 (.06)
Challenge	.19 (.09)*	.21 (.10)*	.23 (.08)**	.32 (.08)***
Control	.01 (.06)	−.02 (.11)	.12 (.08)	.12 (.08)
Care	−.02 (.09)	.01 (.08)	.07 (.12)	−.05 (.10)

Note. Standardized estimates are reported in the table with standard errors in parenthesis. District was controlled for as a fixed factor in the analysis. BAM = Balanced Assessment in Mathematics; ELA = English language arts; ELL = English language learners; MET = Measures of Effective Teaching. *$p < .05$. **$p < .01$. ***$p < .001$.

academically (Lee, 2012; Rubie-Davies, 2007). Moreover, classroom environments that are high on control, such that time is used productively and there are minimal behavioral disruptions, are more conducive to learning (Woolfolk Hoy & Weinstein, 2011).

The present findings extend existing work by using student reports to indicate that control and challenge are important and complementary instructional practices. Given these encouraging and consistent findings, teacher preparation and professional development providers should think critically about how to best support novice and experienced teachers alike in developing these practices in tandem. Additionally, the instructional methods employed by teachers likely serve as the foundation for students' perceptions of their environment. Thus, the positive influence that teachers' *demand* has on the classroom should be considered in the context of instructional practices that are regarded as high quality, such as scaffolding, concept development, and proactive redirection of behavior. Research on teachers' demand could be expanded by examining this construct in the presence and absence of high-quality teaching strategies.

Results regarding students' perception of teacher caring were more complex to interpret. Analyses examining the contribution of challenge, control, and care simultaneously showed no statistically significant association between care and value-added scores. However, as noted in the Results section, analyses that considered the contribution of care alone showed a statistically significant contribution of care to achievement growth. These nuanced findings require careful interpretation involving both methodological and substantive considerations. Challenge, control, and care correlate with one another and thus, the teaching practices situated at the nexus between care and challenge, or between care and control, may be associated with improved achievement. Within a warm demander framework, care and challenge are particularly difficult to tease apart because one way for teachers to implicitly demonstrate care for their students is to hold high expectations for those students (Gay, 2002; Ladson-Billings, 2009). Yet another explanation stems from the selection of achievement on standardized tests as the study outcome. High performance on such assessments requires academic challenge in an orderly environment. Simply demonstrating caring behaviors without providing time and opportunity to learn, as well as academic work well-matched to the student about to take the test, may create a positive classroom climate, but may not facilitate the academic learning essential for achievement. Care may play a larger role when examining outcomes proximal to student learning behaviors, such as students' engagement in learning (e.g., Hughes & Kwok, 2007; Rimm-Kaufman, Baroody, Larsen, Curby, & Abry, 2015; Roorda et al., 2011). The findings raise important questions for future research. For instance, in classrooms that present high challenge and control, does care relate to achievement indirectly by boosting student emotional functioning and learning behaviors?

As mentioned previously, the *demand* (i.e., challenge and control) component of the warm demander theory was associated with growth on high- and low-stakes achievement tests. Challenge emerged as a particularly vital aspect of teaching when exploring classrooms with greater numbers of African American students. Findings related to teachers' level of demand resemble recommendations for culturally responsive teaching practices with African American students, indicating the value of high expectations for students of color (Gay, 2002; Ladson-Billings, 2009).

The importance of challenge in classrooms with greater proportions of African American students also has implications for contemporary issues in educational research and policy. Recent work reveals disparities in the quality of education in low-income and urban areas with larger minority student populations compared to other settings (Pianta et al., 2007). These disparities have sparked debates in the education system regarding the equality of access to high-quality education for African American and Latino students (National Center for Education Evaluation, 2011). Results from the present study indicate that students' perception of challenging instruction is a particularly powerful aspect of teaching for African American students. Although the design of this study does not warrant causal inferences, the work does suggest the importance of future work on interventions that boost expectations and prepare teachers to challenge students, particularly African American students.

Warm demander theory developed out of observations of highly effective African American teachers who could leverage a shared cultural background to develop strong relationships with African American students (Irvine & Fraser, 1998; Ware, 2006). Given that this theory emerged from classrooms with ethnic match between teachers and students, it was necessary to explore the influence of teacher race in the current study. Results indicated that with the current sample, ethnic match or mismatch between teachers and students did not interact significantly with challenge, control, or care in classrooms with predominantly African American students. This finding is not altogether surprising given that other empirical research regarding racial or ethnic congruence between teachers and

students has been fairly mixed and inconclusive (Dee, 2005; Kline, Le, & Hamilton, 2001; Rimm-Kaufman et al., 2000; Saft & Pianta, 2001). The lack of significance in this study can be viewed positively in that challenging environments appear to benefit the African American elementary students regardless of their teachers' ethnicity. However, the finding also requires further inquiry. The MET data are rich, but not exhaustive. For instance, the data collected do not tap elements of African American culture embedded in warm demander practices, such as using culturally specific communication styles (Ford & Sassi, 2014) or referencing shared cultural history and racial identity (Ford & Sassi, 2014; Ware, 2006). In the presence of data with more cultural nuance or more varied outcomes, such as engagement in learning rather than test performance, it is quite possible that the ethnic match between teachers and students could play a larger role in the results.

Limitations and Future Directions

Several limitations should be noted. First, the MET data set is not a nationally representative sample of teachers and students (Kane & Staiger, 2012), which limits the generalizability of the findings. Forty percent of the teachers in the full sample were African American and three quarters of the teachers were African American in the subsample compared to roughly 7% in public school national averages (National Center for Education Statistics [NCES], 2013). Compared to most research on U.S. educators, these data afforded a large and diverse set of teachers and a larger than typical representation of African American teachers. However, one necessary next step is to use a nationally representative sample with a range of grade levels to examine questions of ethnic match or mismatch. Schools are a critical developmental context for youth as they establish their ethnic and racial identity, and ethnic match between teachers and students may hold different meaning and importance at varying developmental levels or grades.

In addition to representativeness, this study examines challenge, control, and care in a descriptive manner, such that these constructs cannot be linked causally to value-added outcomes. A future randomized control trial of an intervention designed to enhance challenge, control, and care would facilitate causal inferences. Yet another need is to understand the contribution of challenge, control, and care on other important student learning behaviors including engagement, motivation, and efficacy.

To our knowledge, there is no existing quantitative measure of warm demander teacher characteristics and thus, the Tripod measure used in the MET study was the closest proxy for the constructs of warmth and demand. Important aspects of culturally specific interactions (e.g., communication style, discussions of shared cultural history, and racial or ethnic identity) are missing from the Tripod, however. A new quantitative measure of warm demander characteristics is needed to measure culturally informed practices for African American students.

Finally, limitations to the value-added models used to determine achievement growth must be acknowledged. For instance, there is limited evidence for the reliability and stability of value-added scores. Despite these limitations, value-added scores are used with increasing frequency in teacher evaluation nationwide. Many districts use value-added scores to make high-stakes decisions about employment and salary increases (Chetty et al., 2011). A beneficial aspect of value-added models is that the data included within the estimate (e.g., students' scores in a measure of that content domain from a year prior, individual student and aggregate classroom demographic information) helps to control for the potential limitations of using nonrandomized classroom data, such as the purposeful sorting of students into certain classrooms.

As mentioned in the Results section, variability in student ratings within a classroom also was examined and findings revealed that higher variability in ratings was related to lower levels of achievement growth. Another avenue for future research is to further explore this question to better understand the adverse effect that variability in student perceptions has on achievement. For example, it is possible that variability is not a product of differentiation, but rather it is an artifact of a teacher interacting more positively with, and providing more challenging experiences for, some students and not others. Our finding that the inverse relation between variability in perceptions and achievement outcomes was strengthened as the proportion of African American students in a classroom increased is consistent with prior research indicating that African American students may be more attuned to differential expectations conveyed by their teachers than their Caucasian peers (Jussim et al., 1996; McKown & Weinstein, 2003).

Conclusion

This study examined the warm demander theory and showed the salience of demand (challenge and control) on academic growth in fourth- and fifth-grade classrooms. The work suggests that challenge may be even more important for African American students, regardless of the ethnicity of the teacher. The work contributes knowledge in education at a time when the United States is deeply concerned about improving the academic outcomes for ethnic minority students. The findings are based on diverse classrooms and lend insight into how students' perceptions of classroom processes relate to selected high-stakes achievement outcomes used as policy-relevant benchmarks. In closing, we highlight three implications of the work. First, in light of the current findings, teachers should be encouraged to hold high expectations for *all* students, particularly African American students in their classrooms. Second, professional development programs need to emphasize the value of classroom management and academic rigor without negating the importance of creating a caring environment. Third, the field of education has many theories similar to the warm demander pedagogy that have been described well and explored using small qualitative data sets. The availability of new national data sets such as the MET study provides researchers with new opportunities to test those theories with larger and more varied samples.

References

Bedeian, A. G., & Mossholder, K. W. (2000). On the use of coefficient of variation as a measure of diversity. *Organizational Research Methods*, *3*, 285–297. doi:10.1177/109442810033005

Bondy, E., & Ross, D. D. (2008). The teacher as warm demander. *Educational Leadership*, *66*, 54–58.

Braun, H. I. (2005). *Using student progress to evaluate teachers: A primer on value-added models*. Princeton, NJ: Educational Testing Service.

Brock, L. L., Nishida, T. K., Chiong, C., Grimm, K. J., & Rimm-Kaufman, S. E. (2008). Children's perceptions of the classroom environment and social and academic performance: A longitudinal analysis of the contribution of the Responsive Classroom approach. *Journal of School Psychology*, *46*, 129–149. doi:10.1016/j.jsp.2007.02.004

Brophy, J. E., & Good, T. L. (1970). Teachers' communication of differential expectations for children's classroom performance: Some behavioral data. *Journal of Educational Psychology*, *61*, 365–374.

Cabrera, N. J., Beeghly, M., & Eisenberg, N. (2012). Positive development of minority children: Introduction to the special issue. *Child Development Perspectives*, *6*, 207–209. doi:10.1111/j.1750-8606.2012.00253.x

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2011). *The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood*. National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w17699

Cholewa, B., Goodman, R. D., West-Olatunji, C., & Amatea, E. (2014). A qualitative examination of the impact of culturally responsive education practices on the psychological well-being of students of color. *The Urban Review*, *46*, 574–596. doi:10.1007/s11256-014-0272-y

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 behaviour. *Educational Psychology: An International Journal of Experimental Educational Psychology*, *28*, 693–710. doi:10.1080/01443410802206700

Cothran, D. J., Kulinna, P. H., & Garrahy, D. A. (2003). "This is kind of giving a secret away. . .": Students' perspectives on effective class management. *Teaching and Teacher Education*, *19*, 435–444. doi:10.1016/S0742-051X(03)00027-1

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *The American Economic Review*, *95*, 158–165. doi:10.1257/000282805774670446

Downer, J. T., Stuhlman, M., Schweig, J., Martínez, J. F., & Ruzek, E. (2014). Measuring effective teacher-student interactions from a student perspective: A multi-level analysis. *Journal of Early Adolescence*, *35*, 722–758. doi:10.1177/0272431614564059

Eccles, J. S. (1999). The development of children ages 6 to 14. *The Future of Children*, *9*, 30–44. Retrieved from http://www.jstor.org/stable/1602703

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford.

Ferguson, R. F. (2008). *The tripod project framework*. Cambridge, MA: Harvard University.

Ferguson, R. F. (2010, October). *Student perceptions of teaching effectiveness* (Discussion Brief). National Center for Teacher Effectiveness and the Achievement Gap Initiative, Harvard University.

Ford, A. C., & Sassi, K. (2014). Authority in cross-racial teaching and learning (re) considering the transferability of warm demander approaches. *Urban Education*, *49*, 39–74. doi:10.1177/0042085912464790

Fulmer, S. M., & Turner, J. C. (2014). The perception and implementation of challenging instruction by middle school teachers. *The Elementary School Journal*, *114*, 303–326. doi:10.1086/674053

Gay, G. (2002). Preparing for culturally responsive teaching. *Journal of Teacher Education*, *53*, 106–116. doi:10.1177/0022487102053002003

Goddard, R. D., Sweetland, S. R., & Hoy, W. (2000). Academic emphasis of urban elementary schools and student achievement in reading and mathematics: A multilevel analysis. *Educational Administration Quarterly*, *36*, 682–702. doi:10.1177/00131610021969164

Good, T. L. (1981). Teacher expectations and student perceptions: A decade of research. *Educational Leadership*, *38*, 415–422.

Gregory, A., & Weinstein, R. S. (2008). The discipline gap and African Americans: Defiance and cooperation in the high school classroom. *The Journal of School Psychology*, *46*, 455–475. doi:10.1016/j.jsp.2007.09.001

Howard, T. C. (2001). Telling their side of the story: African-American students' perceptions of culturally relevant teaching. *The Urban Review*, *33*, 131–149.

Hughes, J., & Kwok, O. (2007). Influence of student-teacher and parent-teacher relationships on lower achieving readers' engagement and achievement in the primary grades. *Journal of Educational Psychology*, *99*(1), 39–51. doi:10.1037/0022-0663.99.1.39

Irvine, J. J. (Ed.). (2002). *In search of wholeness: African American teachers and their culturally specific classroom practices*. New York, NY: Palgrave.

Irvine, J., & Fraser, J. (1998, May 13). "Warm demanders": Do national certification standards leave room for the culturally responsive pedagogy of African American teachers? *Education Week*, p. 56.

Jussim, L., Eccles, J. S., & Madon, S. (1996). Social perception, social stereotypes, and teacher expectations: Accuracy and the quest for the powerful self-fulfilling prophecy. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 28, pp. 281–388). New York, NY: Academic Press.

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teachers: Combining high-quality observations with student surveys and achievement gains*. Policy and Practice Brief Prepared for the Bill and Melinda Gates Foundation. Retrieved from http://metproject.org/downloads/MET_Gathering_Feedback_Practioner_Brief.pdf

Klem, A. M., & Connell, J. P. (2004). Relationships matter: Linking teaching support to student engagement and achievement. *Journal of School Health*, *74*, 262–273. doi:10.1111/j.1746-1561.2004.tb08283.x

Kline, S., Le, V. N., & Hamilton, L. (2001). *Does matching student and teacher racial/ethnic group improve math? (DRU-2529-EDU)*. RAND Education. Retrieved from www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA390470

Ladson-Billings, G. (2009). *The Dreamkeepers: Successful teachers of African American children* (2nd ed.). San Francisco, CA: Jossey-Bass.

Lee, J. (2012). The effects of the teacher-student relationship and academic press on student engagement and academic performance. *International Journal of Educational Research*, *53*, 330–340. doi:10.1016/j.ijer.2012.04.006

Lee, V., & Smith, J. B. (1999). Social support and achievement for young adolescents in Chicago: The role of school academic press. *American Education Research Journal*, *36*, 907–945. doi:10.3102/00028312036004907

McKown, C., & Weinstein, R. (2002). Modeling the role of child ethnicity and gender in children's differential response to teacher expectations. *Journal of Applied Social Psychology*, *32*, 159–184.

McKown, C., & Weinstein, R. (2003). The development and consequences of stereotype consciousness in middle childhood. *Child Development*, *74*, 498–515. doi:10.1111/1467-8624.7402012

Mitchell, M. M., Bradshaw, C. P., & Leaf, P. J. (2010). Student and teacher perceptions of school climate: A multilevel exploration of patterns of discrepancy. *Journal of School Health*, *80*, 271–279. doi:10.1111/j.1746-1561.2010.00501.x

Muthén, L.K., & Muthén, B.O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.

National Center for Education Evaluation. (2011, April). *Do low-income students have equal access to the highest-performing teachers?* (NCEE 2011-4016). Retrieved from http://ies.ed.gov/ncee/pubs/20114016/pdf/20114016.pdf

National Center for Education Statistics (NCES). (2013). *Digest of education statistics*. Retrieved from https://nces.ed.gov/programs/digest/d13/tables/dt13_209.10.asp

Pianta, R. C., Belsky, J., Houts, R., & Morrison, F. (2007). Opportunities to learn in America's elementary classrooms. *Science*, *315*, 1795–1796. doi:10.1126/science.1139719

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

Pianta, R. C., & Walsh, D. J. (1996). *High risk children in schools*. New York, NY: Routledge.

Polikoff, M. S. (2014). The Stability of Observational and Student Survey Measures of Teaching Effectiveness. *The American Journal of Education*, *121*, 183–212. doi:10.1086/679390

Ponitz, C. C., Rimm-Kaufman, S. E., Grimm, K. J., & Curby, T. W. (2009). Kindergarten classroom quality, behavioral engagement, and reading achievement. *School Psychology, Review*, *38*, 102–120.

Reyes, M. R., Brackett, M. A., Rivers, S. E., White, M., & Salovey, P. (2012). Classroom emotional climate, student engagement, and academic achievement. *Journal of Educational Psychology*, *104*(3), 700–712. doi:10.1037/a0027268

Rimm-Kaufman, S. E., Baroody, A. E., Larsen, R. A. A., Curby, T. W., & Abry, T. (2015). To what extent do teacher–student interaction quality and student gender contribute to fifth graders' engagement in mathematics learning? *Journal of Educational Psychology*, *107*, 170–185. doi:10.1037/a0037252

Rimm-Kaufman, S. E., & Chiu, Y. J. (2007). Promoting social and academic competence in the classroom. *Psychology in the Schools*, *44*, 397–413. doi:10.1002/pits.20231

Rimm-Kaufman, S. E., Curby, T. W., Grimm, K. J., Nathanson, L., & Brock, L. L. (2009). The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Developmental Psychology*, *45*, 958–972. doi:10.1037/a0015861

Rimm-Kaufman, S. E., Pianta, R. C., & Cox, M. J. (2000). Teachers' judgments of problems in the transition to kindergarten. *Early Childhood Research Quarterly*, *12*, 363–385. doi:10.1016/S0885-2006(00)00049-1

Roorda, D. L., Koomen, H. M. Y., Spilt, J. L., & Oort, F. J. (2011). The influence of affective teacher-student relationships on students' school engagement and achievement: A meta-analytic approach. *Review of Educational Research*, *81*, 493–529. doi:10.3102/0034654311421793

Rosenthal, R., & Jacobsen, L. (1968). *Pygmalion in the classroom: Self-fulfilling prophecies and teacher expectations*. New York, NY: Holt, Rhinehart, and Winston.

Rubie-Davies, C. M. (2007). Classroom interactions: Exploring the practices or high- and low-expectation teachers. *British Journal of Educational Psychology*, *77*, 289–306. doi:10.1348/000709906X101601

Saft, E. W., & Pianta, R. C. (2001). Teachers' perceptions of their relationships with students: Effects of child, age, gender and ethnicity of teachers and children. *School Psychology Quarterly*, *16*, 125–141. doi:10.1521/scpq.16.2.125.18698

Silva, J. M., Langhout, R. D., Kohfeldt, D., & Gurrola, E. (2015). "Good" and "bad" kids? A race and gender analysis of effective behavioral support in elementary school. *Urban Education*, *50*, 787–811. doi:10.1177/0042085914534859

Skinner, E. A., Kindermann, T. A., Connell, J. P., & Wellborn, J. G. (2009). Engagement and disaffection as organizational constructs in the dynamics of motivational development. In K. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 232–241). New York, NY: Routledge.

Stipek, D. (2002). Good instruction is motivating. In A. Wigfield & J. Eccles (Eds.), *Development of achievement motivation* (pp. 309–332). San Diego, CA: Academic Press.

Vasquez, J. (1989). Contexts of learning for minority students. *The Education Forum*, *52*, 243–253. doi:10.1080/00131728809335490

Villegas, A. M., & Lucas, T. (2002). Preparing culturally responsive teachers: Rethinking the curriculum. *Journal of Teacher Education*, *53*, 20–32. doi:10.1177/0022487102053001003

von der Embse, N. P., Kilgus, S., Solomon, H. J., Bowler, M., & Curtiss, C. (2015). Initial development and factor structure of the educator test stress inventory. *Journal of Psychoeducational Measurement*, *33*, 223–237. doi:10.1177/0734282914548329

Ware, F. (2006). Warm demander pedagogy: Culturally responsive teaching that supports a culture of achievement for African American students. *Urban Education*, *41*, 427–456. doi:10.1177/0042085906289710

Wentzel, K. (1997). Student motivation in middle school: The role of perceived pedagogical caring. *Journal of Educational Psychology*, *89*, 411–419. doi:10.1037/0022-0663.89.3.411

White, M., & Rowan, B. (2013). *User guide to the measures of effective teaching longitudinal database (MET LBD)*. Ann Arbor, MI: Inter-University Consortium for Political and Social Research.

Williams, J. M., & Bryan, J. (2013). Overcoming adversity: High achieving African American youth's perspectives on educational resilience. *Journal of Counseling & Development*, *91*, 291–300. doi:10.1002/j.1556-6676.2013.00097.x

Woolfolk Hoy, A., & Weinstein, C. S. (2011). Student and teacher perspectives on classroom management. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, contemporary issues* (pp. 181–223). Mahwah, NJ: Erlbaum.

Zaslow, M., Martinez-Beck, I., Tout, K., & Halle, T. (2011). *Quality measurement in early childhood settings*. Baltimore, MD: Paul H. Brookes.