1-2017

Child Temperamental Regulation and Classroom Quality in Head Start: Considering the Role of Cumulative Economic Risk

Kathleen Moritz Rudasill
University of Nebraska-Lincoln, kmrudasill@vcu.edu

Leslie R. Hawley
University of Nebraska-Lincoln, lhawley2@unl.edu

Jennifer LoCasale-Crouch
University of Virginia, jl3d@virginia.edu

Eric S. Buhs
University of Nebraska-Lincoln, ebuhs2@unl.edu

Published in *Journal of Educational Psychology* 109:1 (2017), pp 118–130. doi:10.1037/edu0000123
Submitted November 3, 2014; revised February 11, 2016; accepted February 20, 2016; published January 2017.

Child Temperamental Regulation and Classroom Quality in Head Start: Considering the Role of Cumulative Economic Risk

Kathleen Moritz Rudasill
Department of Educational Psychology, University of Nebraska-Lincoln

Leslie R. Hawley
Nebraska Center for Research on Children, Youth, Families, and Schools, University of Nebraska-Lincoln

Jennifer LoCasale-Crouch
Center for Advanced Study of Teaching and Learning, University of Virginia

Eric S. Buhs
Department of Educational Psychology, University of Nebraska-Lincoln

Corresponding author — Kathleen Moritz Rudasill, Department of Educational Psychology, University of Nebraska-Lincoln, 221 Teachers College Hall, P.O. Box 880345, Lincoln, NE 68588-0345. Email: krudasill2@unl.edu

Abstract
There is growing recognition that cumulative economic risk places children at higher risk for depressed academic competencies (Crosnoe & Cooper, 2010; NCCP, 2008; Sameroff, 2000). Yet, children's temperamental regulation and the quality of the early childhood classroom environment have been associated with better academic skills. This study is an examination of prekindergarten classroom quality (instructional support, emotional support, organization) as a moderator between temperamental regulation and early math and literacy skills for children at varying levels of cumulative economic risk. The sample includes children enrolled in Head Start programs drawn from the FACES 2009 study. Three main findings emerged. First, for lower and highest risk children, more instructional support was associated with better math performance when children had high levels of temperamental regulation but poorer performance when children had low temperamental regulation. Second, among highest risk children, low instructional support was protective for math performance for children with low temperamental regulation and detrimental for those with high temperamental regulation. Third, for highest risk children, high classroom organization predicted better literacy scores for those with high temperamental regulation. Children with low temperamental regulation were expected to perform about the same, regardless of the level of classroom organization. Implications are discussed.

Keywords: classroom quality, temperamental regulation, behavioral regulation, cumulative economic risk, Head Start

Optimizing children's learning capacity is among the foremost social and political goals of the current national agenda. Early childhood has been identified as a sensitive period encompassing experiences that affect the development of salient socioemotional and cognitive competencies that are linked to long-term adjustment outcomes (e.g., Entwisle, 1995). A well-established body of research points to the critical role of poverty in the development of academic competencies in early childhood and beyond (Duncan et al., 2007; Raver, Gershoff, & Aber, 2007); there is also growing recognition that *cumulative* economic risk (i.e., having multiple indicators of economic risk) places children at greater risk for poor academic competencies (Appleyard, Egeland, van Dulmen, & Sroufe, 2005; Crosnoe & Cooper, 2010; Sameroff, 2000). Children's temperamental regulation (i.e., the temperament-based ability to attend to appropriate stimuli and inhibit inappropriate behaviors) has also been linked to academic success, with low levels of temperamental regulation associated with poorer academic competence (Martin, Drew, Gaddis, & Moseley, 1988; Rudasill, Gallagher, & White, 2010). However, evidence suggests that high-quality early childhood classrooms can be protective for outcomes of children with economic (Crosnoe & Cooper, 2010; Hamre & Pianta, 2005) or temperamental risk (e.g., poor attention, Rudasill et al., 2010; over- or undercontrolled, Vitiello, Moas, Henderson, Greenfield, & Munis, 2012). Yet, there have been virtually no investigations of how children's temperamental regulation may compound (at low levels) or mitigate (at high levels) cumulative economic risk, and the potential protective role of high-quality classrooms across varying levels of cumulative economic risk.

This study is an examination of temperamental regulation, classroom quality, and cumulative economic risk as predictors of children's academic skills at the end of prekindergarten in a sample of children who are enrolled in Head Start programs. Children served by Head Start programs are typically economically disadvantaged (at least 90% of families enrolled in Head Start must be below the poverty line), and many have multiple economic risk factors, including low levels of parental education and single-parent households. Although low income is often used as an indicator of economic risk, evidence suggests that a combination of risk factors (e.g., low income and low parental education) is most detrimental for children's academic outcomes (Crosnoe & Cooper, 2010). A novel contribution of this study is the examination of varying levels of cumulative economic risk among an already at-risk population of children, thus, enabling investigation of temperamental regulation and classroom quality for children at the highest levels of risk.

Theoretical Framework

This study is grounded in bioecological theory (Bronfenbrenner & Morris, 1998), positing that development occurs through bidirectional interactions (i.e., proximal processes) between an individual and people, objects, and symbols in the environment. Proximal processes are influenced by the characteristics of the individual and his or her environment. From this perspective, we use a Child × Environment model (Ladd, 2003) to conceptualize complex interactions between characteristics of children (temperamental regulation and cumulative economic risk) and preschool classroom quality (instructional support, emotional support, and organization) as predictors of children's academic skill development in prekindergarten. Higher classroom quality is conceptualized as a potential buffer for children with poor temperamental regulation, high cumulative economic risk, or both. This complex set of interactions serves as a potentially powerful example for examining the assertions of Child × Environment models.

Although evidence suggests classroom quality may mitigate negative effects of demographic or temperamental risk (Curby et al., 2011; Hamre & Pianta, 2005), there is no research that examines classroom quality as a moderator of the association between temperamental regulation and academic performance for children at *varying levels* of higher cumulative economic risk. It may be that low temperamental regulation and high cumulative economic risk are, in combination, too maladaptive to be mitigated by high-quality classrooms. For example, Crosnoe et al. (2010) found that, for children from low-income families with less stimulating home environments, being in stimulating preschool and first-grade environments was not sufficient to compensate for deficits in reading and math skill development relative to skills of higher income peers or lower income peers from more stimulating home environments. Thus, although we expect better math and literacy performance for children with higher levels of temperamental regulation who are in higher quality preschool classrooms, hypotheses regarding the extent to which this may vary in interaction with different levels of cumulative economic risk in family contexts have not been empirically examined.

Cumulative Economic Risk

Children from low-income families typically begin kindergarten with far fewer skills than their higher income classmates (U.S. Department of Health & Human Services, 2010) and, without intervention, they fall even further behind as they move through elementary grades (McClelland et al., 2006). Though higher economic risk has consistently predicted poorer outcomes for children, growing evidence suggests that assessing the *accumulation* of risk factors, rather than the presence or absence of a *specific* risk factor, provides a more accurate estimate of the potential impact on development (Appleyard et al., 2005; Crosnoe & Cooper, 2010; Sameroff, 2000). Sameroff, Bartko, Baldwin, Baldwin, and Seifer (1998) found that the level of cumulative risk experienced in preschool, for example, was associated both with concurrent behavior problems in preschool as well as downstream mental health, problem behavior, and academic problems in adolescence. According to Appleyard et al. (2005), the timing of risk is particularly important, with the number of risks in early childhood independently predicting an increase in behavior problems over and above indices of adolescent risk. On the basis of this evidence, we used a similar strategy in the current study.

Crosnoe and Cooper (2010) showed that classroom-level factors moderated the association between child cumulative risk status and achievement in elementary school. This suggests that even for groups of children assumed to experience similar levels of economic risk (e.g., Head Start participants) variation in children's cumulative economic risk may interact with potential protective factors to impact children's outcomes. Thus, failure to examine variations in outcomes as a consequence of cumulative risks may mask some profound differences.

Child Temperament

Temperament is an individual's general style of responding to the environment and refers to innate individual differences in reactivity and regulation (Rothbart et al., 2000). Temperament is "the relatively enduring biological makeup of the [individual], influenced over time by heredity, maturation, and experience" (Rothbart & Derryberry, 1981, p. 40). It is widely acknowledged to be relatively stable through early elementary school and beyond (e.g., Caspi & Silva, 1995; Rothbart & Posner, 2005) and to result from and influence interactions between genes and the environment (Shiner et al., 2012). Although temperamental reactivity refers to one's initial emotional and behavioral reaction to stimuli in the environment, temperamental regulation operates on reactivity (Rothbart & Posner, 2005); it is the temperamentally based ability to regulate an initial emotional and behavioral reaction. For example, a child's initial response (reactivity) in a dispute with a peer may be to cry or lash out; temperamental regulation may, however, help the child curb the initial reaction and enact a more moderate, socially acceptable response. Because of the extensive literature linking temperamental regulation and related constructs (e.g., executive function) to children's academic readiness and success (e.g., Blair, 2002), temperamental regulation is the focus of this study.

Temperamental Regulation

Children's temperamental regulation begins to develop in the second or third year of life, with the greatest development occurring during the preschool and early elementary years (Rothbart & Bates, 2006). Temperamental regulation includes the abilities to attend to stimuli, inhibit inappropriate responses, and align behavior with context-specific expectations (Rothbart & Bates, 2006). It is similar to executive function (EF), which also indicates regulated behavior and incorporates attention and inhibitory control. However, EF emphasizes the cognitive basis of regulation, and includes components beyond those in temperamental regulation such as working memory and planning (Liew, 2011; Wolfe & Bell, 2003).

High levels of temperamental regulation appear to pay particularly big dividends in the school setting. Not only do more regulated children have an easier time with basic behavioral expectations in the classroom (e.g., sitting still, taking turns, remaining quiet, and following directions; Blair, 2002; Rimm-Kaufman, Curby, Grimm, Nathanson, & Brock, 2009) but also they are advantaged when it comes to benefiting from instructional activities. They may be more likely to persist with difficult tasks, work efficiently in loud or distracting environments, and pay attention to teacher instruction—skills associated with better academic outcomes (Bierman, Nix, Greenberg, Blair, & Domitrovich, 2008; McClelland et al., 2007). In an examination of links between children's regulatory and academic skills prior to kindergarten, McClelland and colleagues (2007) found that children's growth in regulation was positively associated with their growth in emergent literacy, vocabulary, and math skills across the prekindergarten year. Relatedly, Bierman et al. (2008) found robust associations between Head Start children's regulation (operationalized as EF) in the fall of prekindergarten and their language, emergent literacy, and behavioral outcomes at the end of the year. Together, these findings highlight the importance of further inquiry into the role of regulation in young children's academic lives, particularly into the factors that may ameliorate or exacerbate poor regulation skills.

Classroom Quality

Observations of classroom instructional support, emotional support, and organization have been used to assess the nature of interactions (proximal processes) between teachers and children in classrooms (i.e., classroom quality). Mounting evidence points to the importance of high-quality preschool classrooms for promoting children's early academic skills and readiness for kindergarten (e.g., Burchinal, Howes et al., 2008; Vitiello et al., 2012), as well as the capacity for high quality classrooms to buffer children at-risk for academic difficulties due to low socioeconomic status (SES; e.g., Burchinal, Howes et al., 2008; Maier, Vitiello, & Greenfield, 2012). The Classroom Assessment Scoring System (CLASS; Pianta et al., 2008) is an observational tool designed to capture these classroom processes, and indicators of instructional support, emotional support, and classroom organization from this assessment have been consistently predictive of children's academic and behavioral outcomes in preschool and early elementary grades (e.g., Mashburn et al., 2008; Pianta, Belsky, Vandergrift, Houts, & Morrison, 2008).

Instructional support. Instructional support refers to teachers' interactions with children that promote concept and skill development through scaffolding, questioning, and feedback between teachers and children. Growing evidence suggests that better instructional support promotes greater achievement. In a large study of academic and social skills among children in public prekindergarten programs, Mashburn et al. (2008) found that classroom instructional support positively predicted children's literacy, language, and math scores above and beyond a host of child and teacher characteristics.

Emotional support. Emotional support in the preschool classroom is also important for children's early school success, particularly for children at risk for academic difficulty (O'Connor & McCartney, 2007). Teachers who provide child-centered classroom environments, marked by positive climate, warmth, and teacher sensitivity (La Paro et al., 2004) are likely to have pupils who thrive academically (e.g., O'Connor, Cappella, McCormick, & McClowry, 2014).

Classroom organization. Given the growing body of work indicating the importance of children's early self-regulatory and executive functioning skills, recent attention has focused on the organizational and managerial aspects of the classroom that appear to be most important for the development of these skills (Ursache, Blair, & Raver, 2012). Young children demonstrate better behavioral and cognitive self-control in the classroom and less time off-task when teachers proactively manage their behavior and attention (Rimm-Kaufman et al., 2009). In addition, evidence suggests that well-organized and managed classrooms benefit children's cognitive and academic development (Downer, Sabol, & Hamre, 2010).

Classroom Quality as a Moderator Between Child Characteristics and Outcomes

Accumulating evidence on the value of high-quality classrooms suggests that such classrooms are likely to be protective for children at risk for academic and/or behavioral problems. In a seminal study, Hamre and Pianta (2005) tested instructional support and emotional support in first-grade classrooms as moderators between children's risk in kindergarten (both demographic and functional risk) and their academic achievement and relationship quality with teachers in first grade. They found that demographic risk (indicated by lower maternal education level) was moderated by instructional support for predicting children's achievement (aggregated across scores on all Woodcock-Johnson subtests of cognition and achievement), such that children in high instructional support first-grade classrooms with less educated mothers performed just as well as their peers with more educated mothers; but in low instructional support classrooms, children with less educated mothers were substantially outperformed by their peers with more educated mothers. Classroom emotional support emerged as a similar mechanism for children's functional risk (the presence or absence of two or more risk factors drawn from sustained attention, externalizing behavior, and social and academic competence) and achievement and teacher–child conflict. Rudasill et al. (2010) extended this work and examined lower levels of temperamental attention in preschool as a potential risk factor for later achievement. Emotional support in third-grade classrooms was construed as a moderator of the link between earlier attention and later reading and math achievement. They found effects consistent with findings from Hamre and Pianta (2005); high emotional support was protective. That is, in high emotional support classrooms, children with low attention performed as well as their peers with high attention, but in low emotional support classrooms, better attention was associated with better performance on math and reading assessments.

Recent research has provided some additional evidence of the protective capacity of classroom quality among low-SES preschool children. Maier et al. (2012) explored Head Start children's teacher-rated psychosocial strengths (i.e., initiative, self-control, and attachment) as predictors of language and literacy skills and the extent to which psychosocial strengths were moderated by preschool classroom quality. Results indicated that psychosocial strengths and classroom organization exerted main effects on children's language and literacy skills, but there was no evidence of interactions between classroom quality and psychosocial strengths. In another study of Head Start children, Dominguez, Vitiello, Fuccillo, Greenfield, and Bulotsky-Shearer (2011) examined classroom quality as a moderator between children's classroom-based problem behaviors and their approaches to learning (ATL). High emotional support classrooms were protective; in those classrooms, children with problems interacting with teachers and problems in instructional activities demonstrated the same proficiency with ATL as their counterparts with few problems. Unexpectedly, in high instructional support classrooms, children with more problems interacting with teachers demonstrated lower levels of ATL than their peers with few problems; but in low instructional support classrooms, children with more problems demonstrated the same ATL as children with few problems. Thus, low instructional support appeared to be protective for children with more problematic interactions with teachers.

Vitiello et al. (2012) conducted a study of Head Start children that evaluated classroom quality as a potential moderator between children's temperament and language, literacy, and math gains across the prekindergarten year. Head Start teachers rated children's temperament as overcontrolled (e.g., shy, fearful, having intense reactions), resilient (e.g., well regulated, positive, friendly), or undercontrolled (e.g., low in regulation, impulsive, high in activity) at the beginning of the school year. Teachers also rated children's language, literacy, and math skills at multiple points during the school year. Classroom quality was assessed through observations during the spring of the prekindergarten year. Results indicated that high emotional support was protective for resilient children's increases in language and literacy skills; that is, resilient children showed more improvement as emotional support increased. However, overcontrolled children's increases across the year did not vary as a function of classroom emotional support. In addition, high instructional support was protective for overcontrolled children's increases in language and literacy skills, but low instructional support was protective for resilient children's improvement. That is, as instructional support increased, overcontrolled children had more growth in language and literacy across the school year; however, resilient children had more growth in language and literacy as instructional support decreased. Collectively, findings from these studies suggest that classroom quality may not consistently ameliorate children's risk; different aspects of classroom quality may be sources of support or stress when examined in interaction with child characteristics such as temperamental regulation or economic risk.

The Present Study

This study is informed by evidence from three bodies of literature. The first is a group of high-quality studies showing academic decrements and disadvantages for children from homes with greater cumulative economic risk, particularly in early childhood. The second is evidence supporting the notion that children's academic achievement is predicated in part on children's temperamental regulation. The third shows that classroom quality is related to children's outcomes. Emerging work combining these literatures suggests that classroom quality may be protective for children at risk for academic difficulty. At the same time, there is some evidence that classroom quality does not work in the same way for all children (e.g., Dominguez et al., 2011). Some of these studies have used Head Start samples (Dominguez et al., 2011; Maier et al., 2012; Vitiello et al., 2012) or used tests of demographic or temperamental risk (Curby et al., 2011; Hamre & Pianta, 2005; Rudasill et al., 2010). However, none has included examinations of cumulative economic risk as a moderator of associations between temperament, classroom quality, and academic skills in the prekindergarten year. The current study is critical because the interplay of temperament and classroom quality factors on children's academic achievement remains understudied, particularly for children at varying levels of cumulative economic risk.

Research Questions and Hypotheses

We used data drawn from the Head Start Family and Child Experiences Survey (FACES) 2009, a large, longitudinal study following children from their entrance into Head Start in the fall of 2009 (ages 3 and 4 years) to the end of kindergarten. With these data, we examined whether prekindergarten classroom quality moderated associations between temperament and literacy/math skills differently for children at varying levels of cumulative economic risk. We expected that classroom quality would have a buffering effect for children with low temperamental regulation or high cumulative economic risk, or both, such that children in higher quality classrooms would display better math and literacy skills at the end of prekindergarten than their peers in lower quality classrooms. We also hypothesized the following: (a) children with higher temperamental regulation would display better math and literacy skills at the end of prekindergarten; (b) children with lower cumulative economic risk would display better math and literacy skills at the end of prekindergarten; (c) children in preschool classrooms where teachers provide higher levels of instructional support, emotional support, and organization would have better math and literacy skills at the end of prekindergarten; (d) children in high-quality preschool classrooms would perform well in assessments of math and literacy skills, regardless of individual differences in temperamental regulation or cumulative economic risk. Conversely, children in low-quality preschool classrooms with poor temperamental regulation or high cumulative economic risk would be outperformed by their more regulated or lower risk peers.

Method

Sample and Participants

Sampling design. The FACES 2009 study used a complex multistage sampling design to create a nationally representative sample of 3- and 4-year-old children participating in Head Start. Data were collected from 3- and 4-year-old cohorts starting in fall 2009 through either spring 2011 (4-year-old cohort) or 2012 (3-year-old cohort). Data were obtained using a four-stage sampling approach that included: (a) Head Start programs (60); (b) centers within Head Start programs (two per program); (c) classrooms within centers (up to three per center); and (d) children within classrooms (10 per classroom; U.S. Department of Health and Human Services, 2013). This sequential sampling method was based on the Chromy procedure, which uses a combination of probability proportional to size (PPS), implicit and explicit stratified sampling to select the programs, centers, and classrooms (U.S. Department of Health and Human Services, 2013). Explicit strata included census region, urbanicity, and racial/ethnic minority enrollment. Implicit strata included percentage in program whose primary home language is English, percentage of children with disabilities, and percentage of dual-language learners. PPS was used in the first three stages with selection of programs, centers, and classrooms. The fourth and final stage of this approach sampled equal numbers of children within classrooms. The complete dataset for the 4-year-old cohort in fall 2009 included 12 first-stage strata, 60 programs, 365 classrooms, and an average of 4 students per classroom. For more information regarding the sampling procedures, see the FACES user's guide (U.S. Department of Health and Human Services, 2013).

Because FACES data were drawn from a multistage sample rather than a traditional simple random sampling design, specific analytical procedures were used to account for the complex sampling to avoid biased parameter estimates and incorrect standard errors (Hahs-Vaughn, McWayne, Bulotsky-Shearer, Wen, & Faria, 2011; Kaplan & Ferguson, 1999). Model- and design-based approaches are the two general methods for handling data obtained from multistage complex sampling (Muthén & Satorra, 1995). To determine the best approach for our analyses, we evaluated the consistency of parameter estimates from both approaches. Design- and model-based analyses were conducted using the software program SAS 9.3 (SAS Institute, 2011). The model-based approach (HLM) was estimated using the MIXED procedure, and the design-based approach was estimated using the SURVEYREG procedure. The model-based approach applies a three level (L1: child; L2: classroom, L3: center) model with sampling weights. The design-based approach estimates a single level model and uses the sampling weights as well as the cluster and stratum design variables. Results from both approaches were consistent in terms of the fixed effect estimates. There were slight differences in the variances across the two approaches, but there was no impact on our conclusions. Further, although a model-based approach in MIXED using maximum likelihood (ML) estimation methods is sometimes preferred when there are missing data, this was not the case with our analyses because the effective sample size using ML methods through the MIXED procedure and listwise deletion through the SURVEYREG procedure was the same (Allison, 2012). On the basis of the consistency in our conclusions across the two approaches and the recommendations in the FACES user's guide (U.S. Department of Health and Human Services, 2013), we determined the most appropriate method for accounting for the complex sampling in this study was to use a design-based approach via the SURVEYREG procedure in SAS.

In our analysis, we specified the stratum (Strata = STRAT) and primary sampling unit (cluster = PSU; Head Start programs) variables to adjust the standard errors to account for the multistage sampling (i.e., clustering). Strata act as independent populations from which Head Start programs or PSUs were sampled (Hahs-Vaughn et al., 2011). Using the SURVEYREG procedure, we applied the child longitudinal weight variable (weight = P12WT) to adjust parameter estimates for differential probabilities of selection and response in our child-level outcomes. Multiple weights are available within the FACES data set, with the choice of weight depending on the specific variables included in the analysis. According to the FACES user guide, sampling weights were created using a three-step process. First, a probability of selection was calculated at each stage of sampling (program, center, classroom, child) and within each explicit sampling stratum. Next, the inverse of the probability of selection was calculated at each stage and stratum. The inverse of the probability of selection is called the sampling weight. The sampling weight is used in the final step, where at each stage the sampling weight is multiplied by the inverse of the weighted response rate. It is assumed that eligibility status of each sampled unit is known at each stage, so this process adjusts weights to account for respondents and nonrespondents (U.S. Department of Health and Human Services, 2013).

Empirical sample. The current study included only the 4-year-old cohort because data were collected from this cohort in both the fall (2009) and spring (2010) of the prekindergarten year. Additionally, we purposely selected only children in the 4-year-old cohort that remained in the same classroom from fall to spring. Table 1 contains the weighted and unweighted sample statistics.

Table 1. Descriptive Statistics

	Weighted			Unweighted		
	n	M	SE	n	M	SD
T1 Age	785	52.37	0.23	937	52.29	3.83
T1 Applied Problems score	693	383.71	1.29	832	382.72	23.81
T2 Applied Problems score	741	396.6	1.20	878	395.5	23.89
T1 Letter-Word score	718	310.75	1.94	860	310.58	25.46
T2 Letter-Word score	752	331.31	1.65	890	330.81	27.31
T2 Temperament	759	–0.13	0.23	899	–0.03	2.40
T2 Classroom Organization	709	4.68	0.05	851	4.72	0.61
T2 Instructional Support	709	2.26	0.05	851	2.27	0.60
T2 Emotional Support	709	5.26	0.06	851	5.29	0.51

	Frequency	Percentage	Frequency	Percentage
T1 Race				
Non-Hispanic, White	141	20.52	181	19.36
Minority	643	79.48	754	80.64
T1 Gender				
Male	388	50.16	469	50.05
Female	397	49.84	468	49.95
T1 Risk				
Low	625	86.97	718	87.35
High	92	13.03	104	12.65
T1 AP Language				
English	667	88.47	803	85.7
Spanish	118	11.53	134	14.3
T2 AP Language				
English	730	93.93	876	93.49
Spanish	55	6.07	61	6.51
T1 LW Language				
English	666	88.35	802	85.59
Spanish	119	11.65	135	14.41
T2 LW Language				
English	728	93.72	874	93.28
Spanish	57	6.28	63	6.72

Estimates were weighted using the P12WT variable and standard errors are calculated for the weighted means. T = time; AP = Applied Problems; LW = Letter Word

Missing data. Owing to concerns regarding potential bias and/or representativeness of our sample, we examined different sources of missingness using the MI function in SAS 9.3. Although more than 1,300 children were included in the 4-year-old cohort, in our empirical sample only 715 children had complete data on all of the outcome variables related to mathematics, and 758 had complete data on the literacy outcomes. Sample sizes varied due to the presence/absence of variables in the model. The sources of missingness included missing outcome variables (math, N = 229; literacy, N = 209), missing sampling weight variable (math, N = 106; literacy, N = 110), and the combination of missing covariates, variables of interest, and/or design variables (math, N = 278; literacy, N = 274). These numbers were slightly reduced again when we limited our sample to only children who remained in the same classroom from fall to spring (math, N = 166; literacy, N = 176). Thus, our final empirical sample was N = 549 and N = 582 for models with math and literacy outcomes, respectively.

In general, our largest source of missing data was missing sampling weight variables and the combination of missing values for other variables of interest and the sampling weight. The FACES user guide notes that sampling weights provided with the data account for selection into the sample, attrition over time, and participant nonresponse (i.e., missing data at the instrument level; U.S. Department of Health and Human Services, 2013). That is, sampling weight variables have been adjusted for missingness because of the individual's chance of being selected and whether their parents, teachers, or center directors responded on an instrument in the FACES survey. Despite the fact that missingness was adjusted with the sampling weight variable, we tested whether demographic and child outcome variables from the first time point (fall 2009) significantly predicted who had missing data at the second (spring 2010) time point. The results from analyses using SURVEYLOGISTIC in SAS 9.3 showed that none of the demographic and child outcome variables in fall 2009 were significantly related to whether or not children remained in the study in spring 2010. We used this same process to test for differences between students who moved from one classroom to another during the school year and found that none of the demographic and child outcome variables from fall 2009 significantly predicted differences between students who moved and those who did not.

Measures

Academic skills. Children's early math skills were measured with the Applied Problems subtest from the Woodcock-Johnson Tests of Achievement—III (WJ-III; Woodcock, McGrew, & Mather, 2007) or the Problemas Aplicados subtest from the Batería III: Woodcock-Muñoz (WM-III; Woodcock, Muñoz-Sandoval, McGrew, & Mather, 2007). Applied Problems will be used to refer to this subscale from both the WJ-III and the WM-III. This subtest measures children's ability to analyze information and solve problems (typically using counting, addition, or subtraction). Children's early literacy skills were measured with the Letter-Word subtest from the WJ-III or the Identification de Letras y Palabras from the WM-III. Letter-Word will be used to refer to this subscale from both the WJ-III and the WM-III. This subtest measures children's ability to correctly label letters and words shown individually on a book page. We used the growth or W-scores from both the WJ-III and WM-III because the mean of the W-scores will change as children progress, demonstrating growth and facilitating comparisons over time as scores are linked to the same developmental scale (U.S. Department of Health and Human Services, 2013). The FACES 2009 user's manual (U.S. Department of Health and Human Services, 2013) reports reliability for the W-scores with 4-year-old children as α = .94 and α = .93 for the Applied Problems subtest on the WJ-III and the WM-III, respectively, and as α = .98 and α = .84, for the Letter-Word subtest on the WJ-III and the WM-III, respectively.

To evaluate the potential for differential prediction of spring WJ-III and WM-III scores, we tested the interaction between academic outcome scores (Applied Problems/Letter-Word) at Time 1 (fall) and assessment language. None of the interactions were statistically significant, indicating that scores from the WJ and WM did not function differently here (highest B = −0.16, SE_b = 0.10, p = .13; Letter-Word).

Temperamental regulation. Temperamental regulation was measured using observations of children's behavior during direct assessments via the Leiter-R Examiner Rating Scales (Roid & Miller, 1997) that have been used previously to measure children's regulation (e.g., Chazan-Cohen et al., 2009). We used three **subscales** : Attention (10 items; "Pays attention to details within tasks"), Activity (4 items; "Focuses without fidgeting, restlessness, or gazing elsewhere"), and Impulse Control (8 items; "Inhibits verbalizations appropriately; does not 'blurt out'"). Assessors rated children from 0 (*rarely/never*) to 3 (*usually/always*) on each item. Results from a large, randomized trial conducted with 2- to 4-year-olds (Olds et al., 2004) suggest that scores from these scales demonstrate convergent evidence with scores for language development and executive functioning; specifically, at-risk children exposed to a home visitation intervention experienced significant improvements in their behavior during testing (as measured using the Leiter-R Examiner Rating Scales; ES = .38), as well as language (ES = .31), and executive function (ES = .47). The FACES user's manual reports Cronbach's alphas for these scales ranging from 0.93–0.97 (U.S. Department of Health and Human Services, 2013). The three subscales were highly correlated (all *r*s > 0.80), so we created a composite factor score from the subscales using M*plus* version 7.0 (Muthén & Muthén, 1998-2012). Factor scores are used as proxies for latent variables because they are the scores that would have been observed if it were possible to measure a latent factor directly (Brown, 2012). This composite factor score was used in all subsequent analyses (composite reliability ω = 0.96; Geldhof, Preacher, & Zyphur, 2014), and the variable is referred to as Temperamental Regulation (weighted M = −0.25, SE = 0.23).

Classroom quality. The Classroom Assessment Scoring System (CLASS; Pianta, LaParo, & Hamre, 2008) was used to evaluate classroom quality. Observations were conducted in the spring for at least 4 hr on one day (conducted in four 30-min cycles; 20 min of observation followed by 10 min of recording and coding). Trained researchers conducted classroom

1. Item-level information as well as interrater agreement, training, and qualification of raters for the Leiter-R was not provided in the dataset or manual.

2. M*plus* uses maximum a posteriori scoring to generate factor scores (MAP; Embretson & Reise, 2000).

observations using the CLASS and were required to achieve agreement with master coders on three videotapes (within 1 point on 80% of the CLASS dimensions; Pianta et al., 2008) prior to conducting observations in the field.

For the current study, we included domain scores for Instructional Support, Emotional Support, and Classroom Organization. Instructional Support (weighted M = 2.26, SE = 0.05) is calculated as the average of three dimensions measuring language modeling, quality of feedback, and concept development. Emotional Support (weighted M = 5.26, SE = 0.06) is a composite measure of four dimensions measuring classroom positive climate, negative climate (reversed), teacher sensitivity, and regard for student perspectives. Classroom Organization (weighted M = 4.68, SE = 0.05) comprises the average of three dimensions measuring behavior management, classroom productivity, and instructional learning formats. All dimensions were scored using a scale from 1 (*low*) to 7 (*high*). The three CLASS domain variables were moderately correlated with one another. The strongest relationship was found between Emotional Support and Classroom Organization (r = .70), whereas weaker relationships were found between Instructional Support and Classroom Organization (r = .49), and Instructional Support and Emotional Support (and r = .51). According to the FACES user guide, coefficient alphas ranged from .79 for Instructional Support to .91 for Emotional Support. Additionally, joint observations with master coders at the beginning, middle, and end of the field data collection period indicated that interrater agreement (within 1 point of master coder) was high (95%–96%) across all time points (U.S. Department of Health and Human Services, 2013). Instructional Support and Emotional Support scores have been associated with preschoolers' academic skills and behavior. For example, Instructional Support was positively associated with language, literacy, and math skills (βs ranged from .33 to .69) and Emotional Support was associated with social competence (β = .06), and behavior problems (β = −.05) in preschoolers (Mashburn et al., 2008). Organization in kindergarten classrooms was associated with more self-control (behavioral, β = .32; cognitive, β = .24) and positive work habits (β = .22; Rimm-Kaufman et al., 2009).

Cumulative economic risk. Consistent with other studies of cumulative risk (e.g., Appleyard et al., 2005; Crosnoe & Cooper, 2010; Evans, 2003), the Cumulative Economic Risk variable was calculated as the simple sum of dichotomous (0 = no, 1 = yes) scores for three indices of economic risk: single-parent household, mother's education less than high school diploma, and household income below federal poverty threshold. Correlations among the three risk indices ranged from r = −.10 to .19. Cumulative Economic Risk scores were calculated in the fall of 2009 and ranged from 0–3. The majority of children in the empirical sample had either one or two risk factors (0 risk factors, 13%; one risk factor, 34%; two risk factors, 40%; three risk factors, 13%). We were particularly interested in understanding how children with the highest level of Cumulative Economic Risk differed from other children in the sample, so we divided children into two Cumulative Economic Risk groups: (a) highest risk (i.e., children with three risk factors; $N_{AppliedProblems}$ = 63 and $N_{LetterWord}$ = 70) and (b) lower risk (children with zero, one, or two risk factors; $N_{AppliedProblems}$ = 455 and $N_{LetterWord}$ = 479).

Control variables. Child demographic variables (gender, age, and race/ethnicity recoded as non-Hispanic White, and minority), fall assessment scores for Applied Problems or Letter-Word, and assessment language (English or Spanish) were used as control variables in analyses.

Data Analysis

Outcomes were children's performance on math and literacy assessments at the end of prekindergarten (spring 2010). Predictors included the control variables, performance on the math or literacy assessment at the beginning of prekindergarten (fall 2009), Temperamental Regulation, classroom quality (Instructional Support, Emotional Support, Organization), Cumulative Economic Risk, and all two- and three-way interactions between Temperamental Regulation, classroom quality, and Cumulative Economic Risk. See the online supplemental material appendix for a general expression of the statistical model used in this study. Before modeling, we tested the data for the presence of heteroscedasticity and multicollinearity. Given the complex data structure, traditional methods for assessing model assumptions were not available, so we used evidence from both survey and traditional approaches to evaluate model assumptions (e.g., Lohr, 2012). Our results showed no evidence of heteroscedasticity or multicollinearity.

The interactions of child Temperamental Regulation, classroom quality, and Cumulative Economic Risk were examined in separate regression models for each of the two outcomes (math and literacy performance) using a top-down approach, which is an iterative, exploratory model-building approach (Ryoo, 2011) in which we started with the most complex model (i.e., all possible interactions with the central variables of interest) and removed interactions that were nonsignificant (p > .05). Each model always included the control variables and variables of substantive interest (i.e., Temperamental Regulation, CLASS variables, Cumulative Economic Risk), but only significant higher order interactions (three-way and two-way interactions) were retained. We also tested the interactions between each of the covariates with the central variables to test the assumption that they do not interact with our variables of interest. There were some significant interactions between our control variables and central variables, so these significant interactions were included in the final models. The final models were the end result of this model building process.

Variables were centered to aid in parameter interpretation. Continuous variables were centered at the sample average. Discrete variables were coded so the reference group was comprised of non-Hispanic, White males who completed assessments in English and indicated lower cumulative economic risk. Given this centering method, the estimated regression effects can be interpreted as the change in the predicted outcome (math or literacy scores) for every one-unit change in a particular variable, holding all other variables constant at the centering point.

Our main research question focused on whether classroom quality moderates associations between temperament

3. We evaluated two coding schemes for the race/ethnicity variable with both of our models: dichotomous coding of non-Hispanic White, and minority; and four categorical codes of the following: (a) non-Hispanic, White; (b) Hispanic/Latino; (c) African American, non-Hispanic; and (d) other. In both models, the more complicated categorical scheme did not indicate significant differences among the four groups. Subsequently, we used the dichotomous coding scheme.

and children's early math and literacy skills at the two levels of Cumulative Economic Risk (highest, three risk factors; lower, zero, one, or two risk factors). We also examined associations between the lower order effects (i.e., two-way interactions and main effects) and achievement; however, these were interpreted only in the absence of higher order effects. To facilitate interpretation of interactions of Temperamental Regulation, classroom quality, and Cumulative Economic Risk, we refer to high (1 *SD* > mean), medium (mean), and low (1 *SD* < mean) levels of each continuous variable. Cohen's f^2 was used as a measure of local effect size following the procedures described in Selya et al. (2012), with "small" effects defined as $f^2 \geq 0.02$ and < .15, "medium" effects defined as $f^2 \geq 0.15$ and < .35, and "large" effects defined as $f^2 \geq 0.35$ (Cohen, 1992).

We conducted post hoc simple slopes tests to probe the significant three-way interactions and controlled for familywise Type I error by risk category using the Bonferroni-Holm procedure (Holm, 1979). Because the main analysis showed predicted values for children's math and literacy skills at the mean value of each CLASS variable (e.g., Instructional Support), simple slopes tests were conducted at high (1 *SD* > mean) and low (1 *SD* < mean) levels of the CLASS variables for the highest and lower risk groups. The *p* values for the slopes within each risk category were ordered from smallest to largest and alpha (.05) was then divided by 3, 2, and 1, respectively to obtain adjusted *p* values. Starting with the smallest *p* value, the adjusted *p* values (p_{BH}) were sequentially compared to .05 until the first nonsignificant test (i.e., $p_{BH} > .05$) was identified, and the remaining ones were declared nonsignificant.

Results

Classroom Quality, Cumulative Economic Risk, Temperamental Regulation, and Children's Skills

Applied Problems. The left side of Table 2 provides the final model and the unstandardized parameter estimates for Applied Problems. The results showed a significant three-way interaction between Temperamental Regulation, Instructional Support, and Cumulative Economic Risk ($B = 3.85$, $SEb = 1.11$, $p < .01$, $f^2 = .01$). Simple slopes analyses were conducted at high and low values of Instructional Support (one unweighted standard deviation above and below the mean) within each of the Cumulative Economic Risk categories (Aiken & West, 1991). All simple slopes are reported as standardized values to coincide with the figures. With the highest Cumulative Economic Risk group, the simple slopes of Temperamental Regulation on Applied Problems scores were −0.18 (*p*BH < .01) and 0.26 (*p*BH < .01) at low and high values of Instructional Support, respectively. With the lower Cumulative Economic Risk group, the simple slopes of Temperamental Regulation on Applied

Table 2. Estimated Regression Coefficients for Spring (T2) Applied Problems and Letter-Word Scores

	Applied Problems				Letter-Word			
Parameter	Estimate	SE	p	f^2	Estimate	SE	p	f^2
Intercept	402.21	1.75	<.01		333.06	2.46	<.01	
T1 Race (Minority)	−4.58	1.52	<.01	.01	−3.35	2.48	.18	.00
T1 Risk (Highest)	2.19	1.83	.24	.00	3.61	3.55	.31	.00
T1 Gender (Female)	2.38	1.39	.09	.01	−0.24	1.75	.89	.00
T1 Age	0.56	0.19	.01	.02	0.66	0.20	<.01	.02
T1 Assessment Score	0.49	0.04	<.01	.36	0.63	0.04	<.01	.44
T1 Assessment Language (Spanish)	−12.52	3.82	<.01	.04	4.92	3.55	.17	.00
T2 Assessment Language (Spanish)	2.23	4.76	.64	.00	−3.32	3.78	.38	.00
T2 Child Temperament	1.59	0.45	<.01	.03	1.88	0.73	.01	.01
T2 Classroom Organization	1.81	1.92	.35	.00	−1.25	2.53	.62	.00
T2 Instructional Support	0.22	1.91	.91	.00	−0.47	2.84	.87	.00
T2 Emotional Support	−2.77	2.21	.22	.00	0.26	2.10	.90	.00
T1 Assessment Language × Emotional Support[a]	—	—	—	—	−17.70	5.44	<.01	.02
T2 Assessment Language × Classroom Organization[a]	−9.21	4.22	<.01	.01	11.81	5.21	.03	.00
Race × Instructional Support[a]	—	—	—	—	8.43	3.24	.01	.01
Temperament × Classroom Organization	—	—	—	—	0.10	1.09	.93	.00
Temperament × Instructional Support	1.37	0.61	.03	.01	—	—	—	—
Temperament × Emotional Support	—	—	—	—	−0.50	0.84	.55	.00
Risk × Classroom Organization	—	—	—	—	8.47	5.21	.11	.00
Risk × Instructional Support	−3.24	3.67	.38	.00	—	—	—	—
Risk × Emotional Support	—	—	—	—	−12.88	5.69	.03	.01
Temperament × Risk	−0.63	0.74	.40	.00	1.65	1.41	.25	.00
Temperament × Classroom Organization × Risk	—	—	—	—	4.88	1.92	.01	.01
Temperament × Instructional Support × Risk	3.85	1.11	<.01	.01	—	—	—	—
Temperament × Emotional Support × Risk	—	—	—	—	−5.45	2.16	.02	.00

Estimates were weighted using the P12WT variable; Continuous variables were mean centered. The reference group for categorical variables was non-Hispanic, White males who completed assessments in English and had lower cumulative economic risk. "Small" effects are defined as $f^2 \geq .02$ to < .15, "medium" effects defined as $f^2 \geq .15$ to .35, and "large" effects defined as $f^2 \geq .35$ (Cohen, 1992). Boldface type indicates coefficients statistically significant at p < .05. T = time.

a. Interactions with control variables were included in the models if preliminary tests were statistically significant, but these interactions are not interpreted because they are not of substantive interest.

Problems scores were 0.01 (pBH > .05) and 0.12 (pBH < .01) at low and high values of Instructional Support, respectively.

Figure 1 shows the predicted Applied Problems scores for lower and highest-risk students. Holding all other variables constant at their centering point, highest-risk students in classrooms with low Instructional Support were predicted to score lower on the Applied Problems assessment as their Temperamental Regulation increased. In contrast, highest-risk students in classrooms with high Instructional Support were predicted to score higher on the Applied Problems assessment as Temperamental Regulation increased. Among the lower risk students, those in classrooms with high Instructional Support were predicted to score higher on the Applied Problems assessment as Temperamental Regulation increased.

Letter-Word. The right side of Table 2 provides the final model and the unstandardized parameter estimates corresponding to the individual variables and interactions for Letter-Word. Our results showed a significant interaction with Temperamental Regulation, Classroom Organization, and Cumulative Economic Risk (B = 4.88, SEb = 1.92, p < .05, f^2 = .01). When risk was highest, the simple slopes of Temperamental Regulation were −0.04 (p_{BH} > .05) and 0.21 (pBH < .05) at low and high values of Classroom Organization, respectively. When risk was lower, the simple slopes of Temperamental Regulation were 0.04 (p_{BH} > .05) and 0.05 (p_{BH} > .05), at low and high values of Classroom Organization, respectively. As Figure 2 shows, holding all other variables constant at their centering point, among the highest-risk students, those in classrooms with high Organization were predicted to perform better on the Letter-Word assessment as their Temperamental Regulation increased.

There was also a significant three-way interaction between Temperamental Regulation, Emotional Support, and Cumulative Economic Risk (B = −5.45, SEb = 2.16, p < .05, f^2 = .00). However, because of the negligible f^2 value, we did not interpret this interaction.

Discussion

Our examination of Head Start classroom quality as a moderator between children's temperamental regulation and their early math and literacy skills across varying levels of cumulative economic risk produced three main findings. First, for children in both risk groups (highest and lower risk), more instructional support was associated with better math performance for those with high levels of temperamental regulation, but poorer performance for those with low temperamental regulation. Second, among highest risk children, low instructional support was protective for those with low temperamental regulation and detrimental for those with high temperamental regulation. Third, for highest risk children, high classroom organization predicted higher literacy scores for children with high temperamental regulation. Children with low temperamental regulation were expected to perform about the same, regardless of the level of classroom organization. Each of these findings will be discussed in turn.

Instructional Support as Protective

Classroom instructional support emerged as a moderator between children's temperamental regulation and math performance (see Figure 1). For lower and highest-risk children, results indicate that high instructional support may function as a protective factor for math performance of children with higher temperamental regulation, but may be detrimental for children with lower temperamental regulation. Also, for highest risk children, low instructional support appears to be detrimental for children with high temperamental regulation but protective for those with low temperamental regulation.

Although not fully congruent with our hypotheses, these findings are similar to those from a study of Head Start children by Dominguez et al. (2011) where children with more problems in teacher interactions (and perhaps similar to children with low temperamental regulation in the current study) had lower levels of approaches to learning (ATL) skills in high instructional support classrooms, whereas in low instructional support classrooms, problems in teacher interactions were unrelated to ATL skills. Our findings are also related to results from Vitiello et al. (2012) where resilient children (similar to children characterized by higher temperamental regulation in the current study) benefitted more from high instructional support than their over- or undercontrolled peers (similar to our low temperamental regulation children). Like both Dominguez et al. (2011) and Vitiello et al. (2012), our results suggest that higher quality classrooms may be variably beneficial, depending on children's characteristics.

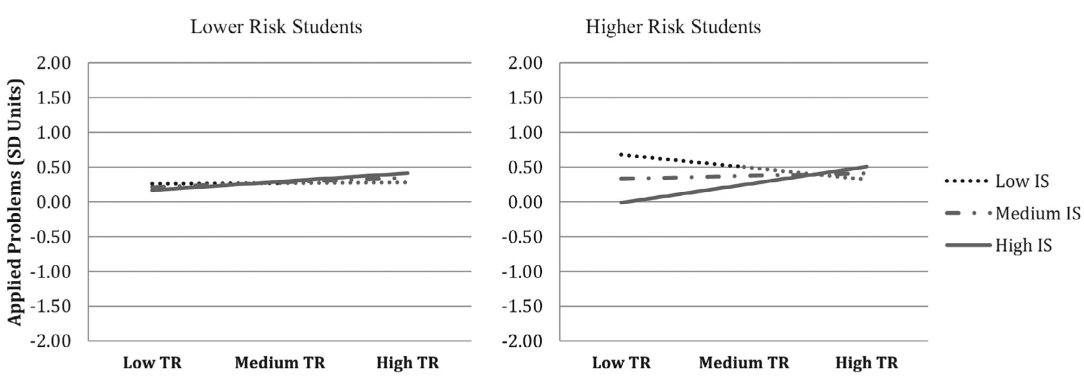

Figure 1. Predicted values in standard deviation units for lower risk students (left image) and higher risk students (right image) for Applied Problems assessment scores based on temperamental regulation (TR) and preschool classroom instructional support (IS), holding all other variables constant at their centering point. High, medium, and low values were based on the unweighted standard deviation (1 SD above, at the mean, and 1 SD below the mean).

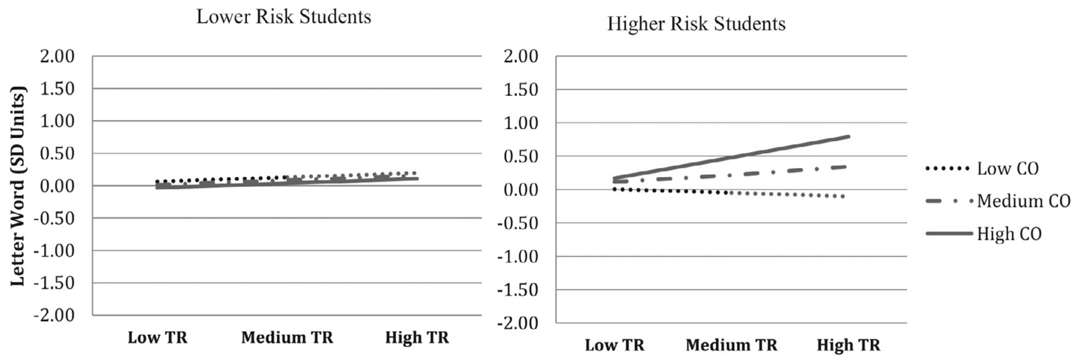

Figure 2. Predicted values in standard deviation units for lower risk students (left image) and higher risk students (right image) for Letter-Word assessment scores based on temperamental regulation (TR) and preschool classroom organization (CO) holding all other variables constant at their centering point. High, medium, and low values were based on the unweighted standard deviation (1 SD above, at the mean, and 1 SD below the mean).

Results reported here suggest that higher quality teacher–child interactions, particularly instructional support, are likely to promote greater math skill development but perhaps only when children have the regulatory skills necessary to benefit from a more intensive and demanding level of instruction (see Burchinal, Howes, et al., 2008; Mashburn et al., 2008; Pianta et al., 2008). In the current study, this was especially true for the highest risk children—if they had low temperamental regulation, their math performance was expected to decrease as instructional support increased. Some studies indicating that instructional support promotes positive outcomes for all children have included economically disadvantaged samples (e.g., Mashburn et al., 2008). However, studies showing that instructional support can be protective for acquisition of academic and social skills among children at-risk due to demographic or temperament characteristics have included samples with relatively low levels of economic risk (e.g., Curby et al., 2011; Hamre & Pianta, 2005). It appears that the early math skill development of children with the highest level of cumulative economic risk is more dependent on their abilities to regulate than for children from lower risk homes. Together with our finding that high instructional support was protective only for math skills of children with high temperamental regulation (and detrimental for children with low temperamental regulation), our results suggest that prioritizing interventions targeting children's regulatory development may be particularly beneficial for facilitating learning in children with more cumulative economic risk.

The fact that high instructional support appeared detrimental for children with less temperamental regulation among the highest risk group may also be understood in the context of the instructional support scores obtained across classrooms in this study. As is typical in classroom quality research (e.g., Mashburn et al., 2008), the average level of instructional support in this study was very low (unweighted *M* = 2.27). Even classrooms with "high" instructional support (2.87 in this sample; 1 *SD* above the unweighted *M*) had scores falling short of a classification as "moderately" instructionally supportive as defined in the CLASS manual (Pianta et al., 2008). Thus, the generally low levels of instructional support in this sample may have masked some associations that would perhaps have emerged in a sample with greater variability. Likewise, it could be that observed levels of instructional support in this study were not high enough to impact the math performance of children with low temperamental regulation. Indeed, Burchinal, Vandergrift, Pianta, and Mashburn (2010) identified 3.25 as a threshold score for instructional support on the scale used here, such that effects of instructional support on children's academic skills were larger in classrooms of moderate or high instructional support (i.e., 3.25 or above) compared with classrooms with low instructional support (i.e., under 3.25)—a category that would have included all of the classrooms in our data set.

Classroom Organization, Temperamental Regulation, and Cumulative Economic Risk

Classroom organization moderated the association between temperamental regulation and literacy performance but only for children with highest cumulative economic risk. That is, similar to findings for instructional support, we found that high classroom organization appeared protective for children with highest economic risk but only if they also had high temperamental regulation (see Figure 2). On the other hand, children with low temperamental regulation were predicted to perform about the same, regardless of the level of organization in their classroom. These results indicate that the benefits of classroom organization may matter more for children with already high levels of regulation, especially in situations where children face elevated cumulative economic risk. Although classroom organization is intended to facilitate children's self-regulation development and, in turn, promote academic success (Downer et al., 2010), findings from the current study suggest that training children's regulation skills may be necessary for optimizing high-risk children's early academic outcomes (O'Connor et al., 2014).

Limitations

This study has several limitations that warrant mention. First, in any longitudinal study, particularly using data from multiple sources, data are likely to be missing due to attrition and participant nonresponse. This is even more likely when a study includes a large proportion of children at-risk due to economic disadvantage, as was the case here. Although a number of children in the original sample had missing data, the use of sampling weights mitigates concerns regarding generalizability and some nonresponse bias. Second, the correlational nature of this study prevents any strong conclusions about causal relationships between children's temperamental regulation, cumulative economic risk, classroom quality, and academic performance. There is a need for experimental work that may more precisely tease apart the mechanisms by which certain classroom processes and regulation-focused interventions may be effective for improving children's learning. Third, effect sizes from this study are small, as is typical of classroom-based research (e.g., Bulotsky-Shearer, Dominguez, & Bell, 2012; Hamre & Pianta, 2005; Reyes, Brackett, Rivers, White, & Salovey, 2012) as well as with studies of specific subsets of children from large representative samples (e.g., Adelson, McCoach, & Gavin, 2012). Fourth, although our measure of cumulative economic risk included the most robust indicators of economic disadvantage (i.e., income below poverty and mothers' education below high school; Crosnoe & Cooper, 2010), research suggests that there are other indicators related to economic risk, such as food insufficiency or household density, that may add meaningfully to our understanding of the impact of cumulative risk on children's outcomes (e.g., Burchinal, Vernon-Feagans, Cox, & Key Family Life Project Investigators, 2008). Fifth, we were unable to measure the density of high cumulative economic risk as a classroom-level variable because only 10 children per Head Start classroom were included in the study. Sixth, our measure of temperament was restricted to regulation; future research in this area should include indices of temperamental reactivity, especially considering the emerging evidence of the importance of these reactive components (e.g., shyness, anger) to children's academic and social success in early childhood classrooms (e.g., Justice, Cottone, Mashburn, & Rimm-Kaufman, 2008; Vitiello et al., 2012). Seventh, classroom observations took place on one day in the spring; it is possible that classroom quality varied more than this sampling strategy revealed. However, multiple studies using the CLASS to assess classroom quality in a single day have found consistent results (Araujo et al., 2014; Burchinal, Howes, et al., 2008; Leyva et al., 2015; Mashburn et al., 2008). Although the single-day observation methodology has limits, evidence from the CLASS manual (Pianta et al., 2008) suggests that there is consistency with observed teacher behavior that can be accounted for by conducting multiple observation cycles on one day and averaging the score, as was done here. Finally, there was a large difference in sample size between the highest and lower risk groups. These sample size differences can lead to some instability in the model due to the small group size of children in the highest-risk category (<100).

Implications

Findings reported here have implications for research and practice. Our results most clearly point to the importance of considering cumulative economic risk when examining the potential effects of classroom quality, teacher behavior, and interventions with at-risk children. Building on work showing that accumulated risk factors predict adjustment outcomes (e.g., Burchinal et al., 2008; Crosnoe & Cooper, 2010; Sameroff et al., 1998), our results suggest that the accumulation of several key economic risk factors (i.e., income below the poverty cutoff, low parent education levels, single-parent family status) creates a level of stress that may prevent children from accessing or benefitting from features of high-quality classrooms that have appeared protective for other children. Although a relatively narrow intervention focus on classroom quality may be attractive because that environment may be more amenable to intervention efforts, our results suggest that the effectiveness of a more supportive environment may depend on children's regulatory abilities when considered in the context of cumulative economic risk. It may also be more effective to incorporate home and parenting environments into intervention protocols. Work by Crosnoe et al. (2010) shows that high-quality interactions in both the home and childcare or school settings were necessary to mitigate reading deficits for low-income children.

In terms of practice, our results complement other work showing the value of children's regulation for their lifelong success (Bierman et al., 2008; McClelland et al., 2007; Moffitt, Poulton, & Caspi, 2013). Here, children with high temperamental regulation were predicted to perform better on math and literacy assessments than their less regulated peers when in classrooms with higher levels of instructional support and organization. This indicates that, in addition to efforts to improve the quality of interactions between children and teachers for promoting positive childhood outcomes, a focus on building children's regulatory skills, particularly for children at the highest levels of cumulative risk, may be a promising avenue for interventions designed to narrow the income achievement gap.

Acknowledgment — This research was supported by a grant from the American Educational Research Association, which receives funds for its "AERA Grants Program" from the National Science Foundation (NSF) under NSF Grant DRL-0941014. Opinions reflect those of the authors and do not necessarily reflect those of the granting agencies.

References

Adelson, J. L., McCoach, D. B., & Gavin, M. K. (2012). Examining the effects of gifted programming in mathematics and reading using the ECLS-K. *Gifted Child Quarterly*, *56*, 25–39. 10.1177/0016986211431487

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. London: Sage Publications.

Allison, P. D. (2012). *Handling missing data by maximum likelihood*. SAS Global Forum: Statistics and Data Analysis (Paper 312–2012). Retrieved from http://support.sas.com/resources/papers/proceedings12/312-2012.pdf

Appleyard, K., Egeland, B., van Dulmen, M. H. M., & Sroufe, L. A. (2005). When more is not better: The role of cumulative risk in child behavior outcomes. *Journal of Child Psychology and Psychiatry*, *46*, 235–245. 10.1111/j.1469-7610.2004.00351.x

Araujo, M. C., Carneiro, P., Cruz-Aguayo, Y., & Schady, N. (2014). *A helping hand? Teacher quality and learning outcomes in kindergarten*. Unpublished manuscript.

Bierman, K. L., Nix, R. L., Greenberg, M. T., Blair, C., & Domitrovich, C. E. (2008). Executive functions and school readiness intervention: Impact, moderation, and mediation in the Head Start REDI program. *Development and Psychopathology*, *20*, 821–843. 10.1017/S0954579408000394

Blair, C. (2002). School readiness. Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist*, *57*, 111–127. 10.1037/0003-066X.57.2.111

Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental processes. In R. M. Lerner (Ed.), *Theoretical models of human development* (5th ed., Vol. *1*, pp. 993–1028). New York, NY: Wiley.

Brown, T. A. (2012). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

Bulotsky-Shearer, R. J., Dominguez, X., & Bell, E. R. (2012). Preschool classroom behavioral context and school readiness outcomes for low-income children: A multilevel examination of child-and classroom-level influences. *Journal of Educational Psychology*, *104*, 421–438. 10.1037/a0026301

Burchinal, M., Howes, C., Pianta, R. C., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Predicting child outcomes at the end of kindergarten from the quality of pre-kindergarten teacher-child interactions and instruction. *Applied Developmental Science*, *12*, 140–153. 10.1080/10888690802199418

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

Burchinal, M., Vernon-Feagans, L., Cox, M., & the Key Family Life Project Investigators. (2008). Cumulative social risk, parenting, and infant development in rural low-income communities. *Parenting: Science and Practice*, *8*, 41–69. 10.1080/15295190701830672

Caspi, A., & Silva, P. A. (1995). Temperamental qualities at age three predict personality traits in young adulthood: Longitudinal evidence from a birth cohort. *Child Development*, *66*, 486–498. 10.2307/1131592

Chazan-Cohen, R., Raikes, H., Brooks-Gunn, J., Ayoub, C., Pan, B. A., Kisker, E. E., ... Fuligni, A. S. (2009). Low-income children's school readiness: Parent contributions over the first five years. *Early Education and Development*, *20*, 958–977. 10.1080/10409280903362402

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159. 10.1037/0033-2909.112.1.155

Crosnoe, R., & Cooper, C. E. (2010). Economically disadvantaged children's transitions into elementary school: Linking family processes, school context, and educational policy. *American Educational Research Journal*, *47*, 258–291. 10.3102/0002831209351564

Crosnoe, R., Leventhal, T., Wirth, R. J., Pierce, K. M., Pianta, R. C., & the NICHD Early Child Care Research Network. (2010). Family socioeconomic status and consistent environmental stimulation in early childhood. *Child Development*, *81*, 972–987. 10.1111/j.1467-8624.2010.01446.x

Curby, T. W., Rudasill, K. M., Edwards, T., & Perez-Edgar, K. (2011). The role of classroom quality in ameliorating the academic and social risks associated with difficult temperament. *School Psychology Quarterly*, *26*, 175–188. 10.1037/a0023042

Domínguez, X., Vitiello, V. E., Fuccillo, J. M., Greenfield, D. B., & Bulotsky-Shearer, R. J. (2011). The role of context in preschool learning: A multilevel examination of the contribution of context-specific problem behaviors and classroom process quality to low-income children's approaches to learning. *Journal of School Psychology*, *49*, 175–195. 10.1016/j.jsp.2010.11.002

Downer, J., Sabol, T. J., & Hamre, B. (2010). Teacher–child interactions in the classroom: Toward a theory of within- and cross-domain links to children's developmental outcomes. *Early Education and Development*, *21*, 699–723. 10.1080/10409289.2010.497453

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

Embretson, S. E., & Reise, S. T. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.

Entwisle, D. (1995). The role of schools in sustaining early childhood program benefits. *The Future of Children*, *5*, 133–144. 10.2307/1602371

Evans, G. W. (2003). A multimethodological analysis of cumulative risk and allostatic load among rural children. *Developmental Psychology*, *39*, 924–933. 10.1037/0012-1649.39.5.924

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*, 72–91. 10.1037/a0032138

Hahs-Vaughn, D. L., McWayne, C. M., Bulotsky-Shearer, R. J., Wen, X., & Faria, A. M. (2011). Methodological considerations in using complex survey data: An applied example with the Head Start Family and Child Experiences Survey. *Evaluation Review*, *35*, 269–303. 10.1177/0193841X11412071

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*, 949–967. 10.1111/j.1467-8624.2005.00889.x

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.

Justice, L. M., Cottone, E. A., Mashburn, A., & Rimm-Kaufman, S. E. (2008). Relationships between teachers and preschoolers who are at risk: Contribution of children's language skills, temperamentally based attributes, and gender. *Early Education and Development*, *19*, 600–621. 10.1080/10409280802231021

Kaplan, D., & Ferguson, A. J. (1999). On the utilization of sample weights in latent variable models. *Structural Equation Modeling*, *6*, 305–321. 10.1080/10705519909540138

Ladd, G. W. (2003). Probing the adaptive significance of children's behavior and relationships in the school context: A child by environment perspective. In R. V. Kail (Ed.), *Advances in Child Development and Behavior* (Vol. *3*, pp. 43–103). San Diego, CA: Academic Press.

La Paro, K. M., Pianta, R. C., & Stuhlman, M. (2004). The Classroom Assessment Scoring System: Findings from the Prekindergarten Year. *The Elementary School Journal*, *104*, 409–426.

Leyva, D., Weiland, C., Barata, M., Yoshikawa, H., Snow, C., Treviño, E., & Rolla, A. (2015). Teacher-child interactions in Chile and their associations with prekindergarten outcomes. *Child Development*, *86*, 781–799. 10.1111/cdev.12342

Liew, J. (2011). Effortful control, executive functions, and education: Bringing self-regulatory and social-emotional competencies to the table. *Child Development Perspectives*, *0*, 1–7.

Lohr, S. L. (2012). *Using SAS for the design, analysis, and visualization of complex surveys*. SAS Global Forum 2012 (Paper 343–2012). Retrieved from http://support.sas.com/resources/papers/proceedings12/343-2012.pdf

Maier, M. F., Vitiello, V. E., & Greenfield, D. B. (2012). A multilevel model of child- and classroom-level psychosocial factors that support language and literacy resilience of children in Head Start. *Early Childhood Research Quarterly*, *27*, 104–114. 10.1016/j.ecresq.2011.06.002

Martin, R. P., Drew, K. D., Gaddis, L. R., & Moseley, M. (1988). Prediction of elementary school achievement from preschool temperament: Three studies. *School Psychology Review*, *17*, 125–137.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., . . .Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, *79*, 732–749. 10.1111/j.1467-8624.2008.01154.x

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

McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology*, *43*, 947–959. 10.1037/0012-1649.43.4.947

Moffitt, T. E., Poulton, R., & Caspi, A. (2013). Lifelong impact of early self-control. *American Scientist*, *101*, 352–359. 10.1511/2013.104.352

Muthén, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. In P. Marsden (Ed.), *Sociological methodology* (pp. 267–316). Washington, DC: American Sociological Association. 10.2307/271070

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

O'Connor, E. E., Cappella, E., McCormick, M. P., & McClowry, S. (2014). An examination of the efficacy of INSIGHTS in enhancing the academic and behavioral development of children in early grades. *Journal of Educational Psychology*, *106*, 1156–1169. 10.1037/a0036615

O'Connor, E. E., & McCartney, K. (2007). Examining teacher-child relationships and achievement as part of an ecological model of development. *American Educational Research Journal*, *44*, 340–369. 10.3102/0002831207302172

Olds, D. L., Robinson, J., Pettitt, L., Luckey, D. W., Holmberg, J., Ng, R. K., . . .Henderson, C. R., Jr. (2004). Effects of home visits by paraprofessionals and by nurses: Age 4 follow-up results of a randomized trial. *Pediatrics*, *114*, 1560–1568. 10.1542/peds.2004-0961

Pianta, R. C., Belsky, J., Vandergrift, N., Houts, R., & Morrison, F. J. (2008). Classroom effects on children's achievement trajectories in elementary school. *American Educational Research Journal*, *45*, 365–397. 10.3102/0002831207308230

Pianta, R., LaParo, K., & Hamre, B. (2008). *The Classroom Assessment Scoring System Pre-K manual*. Charlottesville, VA: University of Virginia.

Raver, C. C., Gershoff, E. T., & Aber, J. L. (2007). Testing equivalence of mediating models of income, parenting, and school readiness for White, Black, and Hispanic children in a national sample. *Child Development*, *78*, 96–115. 10.1111/j.1467-8624.2007.00987.x

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*, 700–712. 10.1037/a0027268

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. 10.1037/a0015861

Roid, G. H., & Miller, L. J. (1997). *Social emotional rating scale—Examiner version. Leiter International Performance Scale—Revised (Leiter-R)*. Wood Dale, IL: Stoelting.

Rothbart, M. K., Ahadi, S. A., & Evans, D. E. (2000). Temperament and personality: Origins and outcomes. *Journal of Personality and Social Psychology*, *78*, 122–135. 10.1037/0022-3514.78.1.122

Rothbart, M. K., & Bates, J. E. (2006). Temperament. In W. Damon & N. Eisenberg (Eds.), Handbook of child psychology: Vol. *3*. *Social, emotional, and personality development* (6th ed., pp. 99–166). New York, NY: Wiley.

Rothbart, M. K., & Derryberry, D. (1981). Development of individual differences in temperament. In M. E. Lamb & A. L. Brown (Eds.), *Advances in developmental psychology* (Vol. *1*, pp. 37–86). Hillsdale, NJ: Erlbaum.

Rothbart, M. K., & Posner, M. I. (2005). Genes and experience in the development of executive attention and effortful control. *New Directions for Child and Adolescent Development*, *2005*, 101–108. 10.1002/cd.142

Rudasill, K. M., Gallagher, K., & White, J. M. (2010). Temperamental attention and activity, classroom emotional support, and academic achievement in third grade. *Journal of School Psychology*, *48*, 113–134. 10.1016/j.jsp.2009.11.002

Ryoo, J. H. (2011). Model selection with the linear mixed model for longitudinal data. *Multivariate Behavioral Research*, *46*, 598–624. 10.1080/00273171.2011.589264

Sameroff, A. J. (2000). Dialectical processes in developmental psychopathology. In A. Sameroff, M. Lewis, & S. Miller (Eds.), *Handbook of developmental psychopathology* (2nd ed., pp. 23–40). New York, NY: Kluwer Academic/Plenum Press Publishers. 10.1007/978-1-4615-4163-9_2

Sameroff, A. J., Bartko, W. T., Baldwin, A., Baldwin, C., & Seifer, R. (1998). Family and social influences on the development of child competence. In M. Lewis & C. Feiring (Eds.), *Families, risk, and competence* (pp. 161–183). Mahwah, NJ: Erlbaum.

SAS Institute Inc. (2011). *SAS/STAT® 9.3 User's Guide*. Cary, NC: SAS Institute Inc.

Selya, A. S., Rose, J. S., Dierker, L. C., Hedeker, D., & Mermelstein, R. J. (2012). A practical guide to calculating Cohen's f^2, a measure of local effect size, from PROC MIXED. *Frontiers in Psychology*, *3*, 111. 10.3389/fpsyg.2012.00111

Shiner, R. L., Buss, K. A., McClowry, S. G., Putnam, S. P., Saudino, K. J., & Zentner, M. (2012). What is temperament now? Assessing progress in temperament research on the twenty-fifth anniversary of Goldsmith et al. (1997). *Child Development Perspectives*, *6*, 436–444.

Ursache, A., Blair, C., & Raver, C. C. (2012). The promotion of self-regulation as a means of enhancing school readiness and early achievement in children at risk for school failure. *Child Development Perspectives*, *6*, 122–128. 10.1111/j.1750-8606.2011.00209.x

U.S. Department of Health and Human Services, Administration for Children and Families. (2010). *Head Start Impact Study: Final report*. Washington, DC: Author.

U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation. (2013). *Head Start Family and Child Experiences Survey (FACES): 2009 Cohort [United States]. ICPSR34558-v1*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. Retrieved from http://doi.org.vufind-dev.unl.edu:2048/10.3886/ICPSR34558.v1

Vitiello, V. E., Moas, O., Henderson, H. A., Greenfield, D. B., & Munis, P. E. (2012). Goodness of fit between children and classrooms: Effects of child temperament and preschool classroom quality on achievement trajectories. *Early Education and Development*, *23*, 302–322. 10.1080/10409289.2011.526415

Wolfe, C. D., & Bell, M. A. (2004). Working memory and inhibitory control in early childhood: Contributions from physiology, temperament, and language. *Developmental Psychobiology*, *44*, 68–83.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2007). *Woodcock-Johnson III Normative Update Tests of Achievement*. Rolling Meadows, IL: Riverside Publishing.

Woodcock, R. W., Muñoz-Sandoval, A. F., McGrew, K. S., & Mather, N. (2007). *Batería III: Woodcock-Muñoz*. Rolling Meadows, IL: Riverside Publishing.