Journal of Applied Developmental Psychology 77 (2021) 101346

Contents lists available at ScienceDirect

Journal of Applied Developmental Psychology

journal homepage: www.elsevier.com/locate/jappdp

Unpacking pre-K classroom organization: Types, variation, and links to school readiness gains☆,☆☆

Lillie Moffett [a,*], Amanda Weissman [a], Christina Weiland [a], Meghan McCormick [b], JoAnn Hsueh [b], Catherine Snow [c], Jason Sachs [d]

[a] University of Michigan, USA
[b] MDRC, USA
[c] Harvard Graduate School of Education, USA
[d] Boston Public Schools, USA

ARTICLE INFO

Keywords:
Teacher organization
Classroom observation
Prekindergarten

ABSTRACT

Classroom organization is an important facet of prekindergarten quality but is typically measured at a *global level* and as a *single construct*. Little is known about how experiences of *different facets* of classroom organization—namely, exposure to teacher organizational strategies—vary across *individual children* in the same classroom and predict gains in children's academic and executive functioning (EF) skills. We coded the total number of minutes 263 prekindergartners were exposed to teacher strategies for classroom organization (e.g., verbal directions, behavior management, rituals and routine cues, and modeling of materials and transitions). Time spent exposed to a global construct of teacher organizational strategies—as well as the CLASS classroom organization score—did not predict children's gains, but time exposed to specific types of organization did. Findings illustrate the utility in separating a global construct into specific types, and offer a first look at child-level measurement of teacher organization.

Introduction

Classroom organization is theorized to be one of the primary drivers of high-quality learning experiences in preschool classrooms (Pianta & Hamre, 2009). Facets of classroom organization such as better organization of materials, smooth transitions, and clear routines are thought to make children feel more secure, increase instructional time, and ultimately support greater school readiness gains in the preschool year. Policymakers have taken note of these linkages. Observed classroom organization is included in 38 state quality rating improvement systems (QRIS) for preschool (Tout et al., 2017), a key policy lever that aims to enhance quality in early childhood education.

Empirically, however, the leading measure of classroom organization – a subcomponent of the Classroom Assessment Scoring System (CLASS; Hamre & LaParo, 2008) – generally has shown small or null relations with children's school readiness gains in preschool (Authors, 2020; Hong et al., 2015; Weiland & Yoshikawa, 2013; Williford, Maier, Downer, Pianta, & Howes, 2013). Two potential reasons may be that the CLASS measures the *global quality* of classroom organization across all children and examines classroom organization as a *single construct*. Classroom organization includes classroom-level components such as the physical set-up of the classroom, which is naturally measured as a global construct (Pianta, LaParo, & Hamre, 2008). Yet, another component of classroom organization includes teacher strategies for

☆ The research reported here was conducted as a part of a study funded by Grant R305N160018 – 17 from the Institute of Education Sciences to MDRC with subcontracts to the University of Michigan, the Boston Public Schools, and the Harvard Graduate School of Education. Lillie Moffett's and Amanda Weissman's work was supported by the Institute of Education Sciences Predoctoral Fellowship at the University of Michigan (Grant R305B150012). And Lillie Moffett is further supported by the NSF Graduate Research Fellowship (Grant 1650114).☆☆ Thanks to the Boston Public Schools, Brian Gold, Blaire Horner, Annie Taylor, the BPS Department of Early Childhood coaches and staff, the BPS Department of Research, the MDRC team (Jennifer Yeaton, Mirjana Pralica, Rama Hagos, Desiree Alderson, Marissa Strassberger and Sharon Huang), the Harvard Graduate School of Education research team (Sibyl Holland, Maia Gokhale, and field-based data collection staff), the University of Michigan research team (Paola Guerrero Rosada, Kehui Zhang, Margaret Michalowski, Christina Daniel, and Sarah Niemann), and Frederick Morrison and Carol Connor for advising our team on the ISI Measure.

* Corresponding author.
E-mail address: lillmoff@umich.edu (L. Moffett).

https://doi.org/10.1016/j.appdev.2021.101346
Received 12 April 2021; Received in revised form 20 August 2021; Accepted 8 October 2021
Available online 23 October 2021

organizing and orienting students, which teachers might implement with some children more often than others. For instance, a teachers' behavior management portion of her CLASS classroom organization score could be based on instances in which she effectively managed the behavior of a few students, but other children may not have received the same organizational attention. Teachers could also practice other organizational strategies, like providing extra directions and orientation before an activity, to a different subset of children from those eliciting more behavior management. Empirically, prior studies have shown that some aspects of children's classroom experiences such as positive and negative engagement with teachers and peers and exposure to literacy instruction can vary across children enrolled in the same preschool classroom, and that this variation predicts gains in literacy (Connor, Morrison, & Slominski, 2006; Vitiello, Booren, Downer, & Williford, 2012). However, we do not know whether there is similarly child-level variation in exposure to different types of teacher organizational efforts.

To address this gap in the literature, using data from a large-scale public prekindergarten program, we examined whether there is *child-level* variation in exposure to *specific types* of teacher organization—like teachers' uses of behavior management, verbal directions, modeling of materials and transitions, and ritual and routine cues. And we further explore whether this within-classroom variation predicts child learning gains above and beyond a global measure of classroom organization.

Exploring measurement of classroom organization at the child level

Classroom organization encompasses classroom physical setup, teachers' behavior management techniques, teachers' overall efficiency with routines, and organization of materials (Pianta & Hamre, 2009). Theoretically, classroom organization maximizes the time students spend engaged in academic tasks, which in turn maximizes opportunities for learning and increased gains in academic skills such as children's math and language skills (Brophy, 1979; Fisher et al., 1980). Classroom organization is also theorized to provide external supports—like teacher behavior management—for children to then learn how to develop internal supports of their own behavior and executive functioning skills (Rimm-Kaufman, Curby, Grimm, Nathanson, & Brock, 2009).

Despite theory linking children's learning gains in preschool to classroom organization, the empirical evidence on the associations between classroom organization as a global construct and children's gains in preschool has been mixed. Some studies have found small, positive relations between classroom organization and preschoolers' gains in EF (Choi et al., 2016; Hamre, Hatfield, Pianta, & Jamil, 2014), language (Xu et al., 2013), and math (Downer et al., 2012). Others, however, have found largely null linear associations between classroom organization and gains in language, literacy, math, and EF skills in preschool (Authors, 2020; Hong, Howes, Marcella, Zucker, & Huang, 2015; Williford et al., 2013). A recent metanalysis that included 19 peer-reviewed studies found a small positive relation between classroom organization and preschoolers' gains in inhibitory control (an EF skill) but no consistent relation between classroom organization and children's language or math gains in preschool (Perlman et al., 2016).

It is possible that measuring classroom organization at the *global level* is one of the drivers of these null findings. Conceptually, classroom organization does comprise some global components that all children experience together, such as the teacher organizing materials before a whole-group activity or the teacher providing behavior reminders to the entire class. However, individual experiences of classroom organization—like a teacher providing more behavior reminders or directions to some children more than others—may differ across children in the same classroom because preschool settings tend to emphasize a high level of child choice and fewer instances of whole-class activities (NAEYC, 2009). By design, this emphasis leads children within the same classroom to engage in different learning opportunities and different amounts of time exposed to scaffolded interactions from the teacher throughout the school day (Bulotsky-Shearer, Fantuzzo, & McDermott, 2008; Pianta, 2001). For example, one study in 701 classrooms in 11 states found that some preschoolers spent as much as 41% of the school day in individual or peer free-play, while other children spent as little as 11% of time in free play and instead engaged in more academic activities and teacher-scaffolded interactions (Chien et al., 2010). It is difficult to tease apart whether these differences are due to child choice and seeking out these experiences or teacher encouragement to engage. But the net effect is that some children will be exposed to more teacher time—and thus potentially more exposure to organizational support—than others. Differential classroom organization experiences could also be driven by individual variation in children's need for teacher directions and behavior reminders. These child-level differences are why teacher professional development programs for behavior management highlight how to tailor behavior plans for individual children (e.g., Webster-Stratton, Reid, & Stoolmiller, 2008). Taken together, the literature demonstrates that some children receive more attention from the teacher than others, but we do not yet know the extent to which individual children receive different organizational supports that may in turn influence their academic outcomes.

Exploring the separation of classroom organization into subcomponents

Another reason for the null relations between classroom organization as typically measured and prekindergarteners' gains may be because classroom organization is typically studied as a *single construct*. However, classroom organization is comprised of several subcomponents that may be differentially predictive of children's gains. Past work using the CLASS has found that subcomponents of classroom organization (e.g., Behavior Management, Productivity, and Instructional Learning Formats) are typically highly correlated with one another in prekindergarten (Downer et al., 2012; La Paro, Pianta, & Stuhlman, 2004) so they are not analyzed separately. However, a child-level duration-based measurement approach as we utilize in this study may capture more nuance with these subcomponents of classroom organization, especially behaviors we can code in individual teacher-child interactions.

Specifically, in this study we investigate four classroom organization subcomponents that are specific to teaching strategies (as opposed to the physical organization of the classroom that cannot be examined at the child level) – teacher use of *verbal directions*, teacher *behavior management,* teacher *modeling of materials and transitions*, and teacher use of *rituals and routine cues*. We focus on these specific teaching strategies because they can be measured in *quantity of exposure* at the child level, and each strategy overlaps with at least one of the three theorized components of classroom organization (Pianta & Hamre, 2009). We focus on quantity of exposure as an alternative to a quality-based measure like the CLASS because of the demonstrated variability in children's *quantity* of certain experiences within the same classroom (Burchinal et al., 2021; Pianta et al., 2020). Also, classroom organization is theorized to impact children's school readiness gains via maximizing time in the classroom devoted to opportunities for learning (Brophy, 1983). Yet, no study to date has looked at quantity of child exposure to different teacher organizational strategies.

The first construct we explore is teacher verbal directions (which aligns with aspects of CLASS Productivity theoretical construct). We hypothesize that time spent in *teacher verbal directions* might support children's language development via additional language dosage (Farran, Aydogan, Kang, & Lipsey, 2006). Verbal directions might also support both language and EF by providing children the opportunity to practice more internal talk to carry out multi-step actions and practice inhibitory control (Bono & Bizri, 2013; Fuhs & Day, 2011; Weiland, Barata, & Yoshikawa, 2014; Winsler, A. De Leon, J.R. Wallace, B.A., Carlton, M.P…Willson-Quayle, A., 2003). The second construct is *teacher behavior management* (which aligns with the CLASS Behavior Management theoretical construct) in which a teacher comments on children's behavior, and provides rules and expectations, may operate

along similar lines – e.g., upping language dosage and supporting development of EF through internal self-talk. The third construct, *teacher modeling* (aligns with aspects of the CLASS Instructional Learning Formats theoretical construct), may provide a higher level of scaffolding than verbal-only interactions (Vukelich, 1991), particularly for children not yet able to remember multiple-step directions (Sainato, 1990). *Modeling* may also be particularly helpful in promoting children's gains in math, as preschool math instruction tends to emphasize use of manipulatives that require extensive modeling to keep children on task and organized during math learning (Rimm-Kaufman, Fan, Chiu, & You, 2007). Finally, *teachers' use of rituals and routine cues* (aligns with aspects of CLASS Productivity theoretical construct) might reduce demands on children's working memory and inhibitory control, thereby helping them stay on task and learn new academic content or attend to new instructions. These cues can also increase overall productivity in the classroom (e.g., clapping cue takes up less time than long-winded reminders and directions), so more time can then be spent on meaningful instructional. To our knowledge, however, empirical evidence on classroom organization measured at this subcomponent level, and how it differentially relates to children's school readiness gains, is currently scant.

Measuring subcomponents of classroom organization at the child level

Testing whether child-level experiences of teacher organization strategies and subcomponents of teacher organization strategies are more predictive of children's learning gains requires measurement at the child level. To our knowledge, the measure used in the present study – the Individualizing Student Instruction measure (ISI; Connor et al., 2009) – is one of the few that measures exposure to teacher organizational strategies at the child level. The ISI is a relatively new assessment of children's individual classroom experiences measuring the duration of time individual children are in different instructional and non-instructional activities, which includes teachers' use of directions. To date, one study of the ISI in 44 first grade classrooms found that classroom organization measured at the child level predicted modest gains in children's spring letter and word reading skills ($\beta = 0.30$; Cameron, Connor, Morrison, & Jewkes, 2008).

Because the original ISI observation system grouped dimensions of classroom organization together, we adapted it to measure four types of teaching strategies for classroom organization in prekindergarten classrooms – exposure to behavior management, verbal directions, modeling of materials and transitions, and rituals and routine cues. We did so because of our hypotheses (explained in the previous section) that these subcomponents may be differentially predictive of prekindergartners' gains. We chose the ISI both because it allows such adaptations and because of its focus on the *duration of time* children are exposed to different organizational strategies. Child-level time measures permit more variability between students within the same classroom versus a Likert-style rating scale used in many classroom observational tools (e.g., 1–5 or 1–7). Further, recording precise amounts of time for each observed child may reduce rater bias; a coder's ratings of a teacher's interactions with one child may be influenced by their interactions with another in a Likert-style rating system. Recording an objective "start" and "end" time of exposure (as with the ISI) can reduce this potential confound in our measurement.

Present study

In the present study, we explore links between child-level classroom organization—specifically exposure to specific teacher organizational strategies—and gains in prekindergarteners' math, language, and EF skills. Our specific research questions are:

Exploring measurement of classroom organization at the child-level

On average, how much time are individual children exposed to teacher organizational strategies in the prekindergarten classroom and does this vary across children in the same classroom?

Exploring the separation of classroom organization into subcomponents

Does the amount of time a child is exposed to *overall* teacher organization and to *specific subcomponents* of teacher organization predict children's math, language, and EF gains in prekindergarten—over and above global classroom organization quality (measured by the CLASS Classroom Organization construct)?

Method

Participants and setting

The sample consists of 263 children from 39 classrooms and 19 public schools who attended the Blinded for peer review prekindergarten program during the 2016–2017 academic year.[1] Thirty-six classrooms included in the current study used the Blinded for peer review Focus on K1 curriculum, which is comprised of an adapted version of the Opening the World of Learning (Schickedanz, Dickinson, and Charlotte-Mecklenburg Schools, 2005) language and literacy curriculum, the Building Blocks (Clements & Sarama, 2007) early mathematics curriculum, and several district-created components (Authors, 2020b).

Full demographic descriptive statistics are included in Table 1. Students were diverse with respect to eligibility for free-or-reduced price lunch, Dual Language Learning status, race/ethnicity, and gender. Teachers were also diverse with respect to race/ethnicity, but most teachers had a Master's degree and all teachers were female.

Procedure

The Institutional Review Board approved the human subjects plan prior to the commencement of study activities (HUM00114067).

School and classroom recruitment

We randomly selected 25 schools to recruit for the study from the 76 schools in the broader district offering the public prekindergarten program. Twenty-one of the targeted schools agreed to participate. Of these schools, one served as a pilot site for testing new measures that could be used in the study schools and was excluded from our research sample. The rest were asked to participate in the study in the fall of 2016. Forty-one classrooms from these schools agreed (Authors, 2020b). However, two teachers in one school declined to be videotaped, resulting in a sample of 39 total classrooms and 19 total schools in the current study.

Student recruitment

Eighty-one percent of children in participating classrooms consented to participate in the study. Of these, we randomly selected 50% (~6–10 per classroom) to participate for a total sample size of 307. For the current study, we excluded students who were absent for filmed observations or were in the one school that did not participate in the filmed observations, for a final sample size to 263. This restricted sample is demographically similar to both the broader study sample (Authors, 2020b) and the school district at large (Authors, 2021b).

[1] We selected this school district as our study site due to a longstanding research practice partnership with the district (Authors, 2021a).

Table 1
Demographic information for study sample and child achievement Descriptive statistics.

Characteristic	Mean/ Percent	SD	Percent Missing
Child characteristics			
Race/ethnicity			
Hispanic	30%	–	0%
White	28%	–	0%
Black	17%	–	0%
Asian	17%	–	0%
Other race	8%	–	0%
Female	52%	–	0%
Eligible for free/reduced-lunch	57%	–	0%
Dual language learner	55%	–	0%
Child age at baseline	4.66	0.29	0%
Difference in ages between baseline and outcome measures	0.54	0.06	0%
Parent education			
High school diploma/GED or less	31%	–	1.9%
Two-year degree or equivalent	23%	–	1.9%
Four-year degree	16%	–	1.9%
Advanced degree	30%	–	1.9%
At least one parent works 35 h per week	89%	–	3.04%
At least one parent attended Head Start or PreK	55%	–	3.42%
Age of mother at first child's birth	27.83	6.96	3.8%
Number of people living in household	4.27	1.24	4.18%
Parents are married/have a partner	70%	–	1.9%
Parent respondent age at baseline	36.77	7.08	5.32%
Mother was respondent	86%	–	2.28%
Father was respondent	13%	–	2.28%
Fall of PreK achievement measures			
PPVT raw	74.26	27.72	1.9%
WAP raw	12.71	5.03	1.52%
Forward Digit Span	3.13	1.06	1.52%
Spring of PreK achievement measures			
PPVT raw	87.48	26.81	1.52%
WAP raw	15.94	4.48	1.90%
Forward Digit Span	3.51	1.02	1.52%
REMA	20.77	9.28	1.52%
ISI measures (child level)			
Minutes observed	184.37	63.88	0%
One observation	16%	–	0%
Time in whole class instruction	80.99	38.62	0%
Teacher disorganization	11.63	7.70	0%
Global organization (classroom level)			
CLASS Organization	5.47	0.60	0%

Notes: $N = 263$. With the exception of minutes observed, ISI measures operationalized as number of minutes observed. Base sample restricted to students with positive total observed time.

Direct assessments

We trained research staff to reliability[2] and then collected direct assessments of children's school readiness skills in the fall of 2016 (October 1st through December 12th) and spring of 2017 (April 5th through June 16th). A master's-level supervisor observed 10% of field assessments to ensure high-quality administration.[3] We used the Preschool Language Assessment Scale (PreLAS; Duncan & DeAvila, 1998) Simon Says and Art Show tests as a warm-up to the assessment battery and to determine the administration language for a subset of assessments (Barrueco, Lopez, Ong, & Lozano, 2012). Of the 263 children in the current study sample, 70 children did not pass the PreLAS screener in either the fall, spring or both time points. 54 of these children were Spanish-speaking and thus took the Spanish version of the assessments, and 16 of these children had a home language other than Spanish and thus still took the assessment in English to retain as many children as possible.

Parent survey

In the fall of 2016, we reached out to the parents of the study students via email and text message to ask them to complete the parent survey. The survey consisted of parental and child demographic information along with questions about educational experiences. Although the majority of parents completed the survey in English, we also translated the survey into Spanish, Vietnamese, and Mandarin. Of the 263 students in our survey, 258 (98%) of parents completed the survey.

Classroom observations

Classrooms were observed across two school days in February and March 2017 (80% of classrooms were observed in these two months; observation timing ranged from January 25th – May 10th). On average, the second observation occurred 15.56 days ($SD = 14.65$) after the first observation. We used two video cameras that captured wide-angle shots of the classroom. One camera was focused primarily on the teacher and the students around her and the other camera focused on the rest of the students in the classroom. Cameras were moved around often to make sure all children in the study were within view. Each observation lasted an average of 1.79 h, ranging from 1.10 to 2.61 h. Coders of these classroom observations participated in multi-day trainings and the double-coding of 20% of the observations. Personnel who coded the ISI measure differed from the CLASS measure.

Measures

Receptive language skills

We used the Peabody Picture Vocabulary Test IV (PPVT IV) to directly assess children's receptive language skills in the fall and spring. The PPVT IV is a nationally normed measure that has been used widely in diverse samples of young children. The test has excellent split-half and test– retest reliability estimates ranging from 0.87 to 0.93, as well as strong qualitative and quantitative validity properties (Dunn & Dunn, 2007). It requires children to choose (verbally or nonverbally) which of four pictures best represents a stimulus word. In our primary analysis, we used the raw score total as our outcome measure. We assessed all children on the PPVT, regardless of whether they passed the PreLAS language screener, in order to describe an equivalent measure of receptive language skills in English across the full sample.

Math skills

We used the Woodcock Johnson Applied Problems III (Woodcock, McGrew, & Mather, 2001) subtest and the Research-Based Early Mathematics Assessment (REMA; Clements, Sarama, & Liu, 2008) to directly assess children's math skills in the fall and spring. We assessed Spanish-speaking children who did not pass the PreLAS language screener using the equivalent Spanish language version of the assessment from the Batería III Woodcock Muñoz (Schickedanz, Dickinson, and Charlotte-Mecklenburg Schools, 2005). The Woodcock Johnson Applied Problems (WJAP) direct assessment is a numeracy and early mathematics measure that requires children to perform relatively simple calculations to analyze and solve arithmetic problems. Its estimated test–retest reliability for 2- to 7- year-old children is 0.90 (Woodcock et al., 2001) and it has been used with diverse populations (Gormley Jr, Gayer, Phillips, & Dawson, 2005; Peisner-Feinberg et al., 2003; Wong et al., 2008). In our primary analyses, we present results using the raw score of the measure. However, we fit models using the age-standardized and W score versions of the Applied Problems as seen in Appendix A.

The REMA is a comprehensive math assessment that captures math

[2] In order ensure internal validity and reliability for the direct child assessments, the research staff participated in a week-long training at the beginning of both the fall of 2016 and spring of 2017. Staff received training on each individual assessment and practiced during the sessions. They then had to give the assessments to an adult role playing as a child, and then they had to assess a student attending a school that was not part of the study sample.

[3] Assessments were in a quiet space outside of the classroom. It took approximately 45 min per child to administer all child assessments, and each child completed all assessments in one day.

skills beyond numeracy, such as geometry, and measurement skills, focusing on hands-on manipulatives and items targeted to the early childhood period. Alpha reliabilities for subscales range from $r = 0.89$ (number) to 0.71 (geometry; Clements, Sarama, & Liu, 2008). Because we collected this assessment only in the spring of prekindergarten, we used the WJAP fall score as a baseline score when using REMA as an outcome. We use raw REMA scores in our primary analysis. Similar to the WJAP, students who did not pass the PreLAS language screener took the Spanish version of the REMA, translated by the original assessment creators with the same psychometrically valid properties (Clements & Sarama, 2007). We include results of models fit using REMA T and IRT scores in Appendix A.

Executive functioning

We used the Forward Digit Span measure (FDS; Gathercole & Pickering, 2000) to assess children's short-term memory, which is a developmental precursor to the working memory component of executive functioning. We examined short term memory because it exhibits more variability than working memory measures at age 4 (Gathercole, 1999), yet is equally predictive of academic achievement from preschool to age 7 (Bull, Epsy, & Wiebe, 2008). During this task, children are asked to repeat a string of numbers in order. A score of "0" indicates that a child was unable to pass the practice test (child given two attempts to repeat two numbers forwards), a score of "1" indicates that the child passed the practice test but could not complete the first test trial, and a score of "2–6" indicates how many digits the child could repeat in order. FDS demonstrates high correlations with other related memory tasks—both verbal and spatial (Carlson, Moses, & Breton, 2002), as well as good test-retest reliability ($r = 0.73$; Lipsey et al., 2017). As with the math assessments, we administered this assessment in Spanish for students who did not pass the PreLAS screener.

Classroom organization

We measured global Classroom Organization quality using the Classroom Assessment Scoring System (CLASS) PreK (Pianta et al., 2008). Classroom Organization includes measures of behavior management, productivity, and instructional learning formats on a 7-point scale. The CLASS shows good psychometric validity in the literature (Burchinal, Vernon-Feagans, Vitiello, Greenberg, and The Family Life Project Key Investigators, 2014; Hatfield, Burchinal, Pianta, & Sideris, 2016; Leyva et al., 2015; Weiland & Yoshikawa, 2013), and in our study Classroom Organization had good internal consistency ($\alpha = 0.88$) and interrater reliability (Kappa = 0.94). This measure was the only variable used that was measured at the classroom level as the remaining variables were measured at the child level.

Teacher organizational strategies

We used the Individualizing Student Instruction (ISI) Coding System (Connor et al., 2009). The ISI focuses on three dimensions of an individual student's classroom experience: the amount of time students are in specific instructional or non-instructional activities (e.g., classroom management or routines, etc.), how much time students spend in different content areas (e.g., math, language arts, etc.), and how long students spend in different learning situations (e.g., whole-class teacher-managed, individual child self-managed, peer-managed, etc.).

For this study, we focused on the time students were exposed to teacher organizational strategies, or the "planning/organizing/rules" code in the ISI coding system. We separated this into four different types of teacher organizational strategies including: 1) verbal directions; 2) behavior management; 3) modeling of materials and transitions and 4) rituals and routine cues (detailed in Table 2). Each of these codes could be examined at the child-level. For example, the teacher could single out a child during a whole-group activity who is off-task and provide them with a behavior redirection, and the rest of the children would be coded as "waiting," since they are receiving a lull in their instruction. A teacher could also provide direction reminders or additional modeling to a specific child who did not understand the activity the first time around. She could also provide an individual child with a ritual and routine cue (e.g., re-singing a verse of a transition song for a child who was not cleaning up when directed, etc.). In our observations, children were in whole-group for 43% of the time, followed by center-time (35%), transitions (16%), and small-group instruction (6%). As such, there was ample opportunity to capture individual-child interactions with teacher organization strategies.

Table 2
Description of teacher organizational strategies and examples.

Verbal Directions Directions for an upcoming activity *"When you get to your desk, I want you to take out a pencil and your workbook"* Prefacing an activity to orient children (or individual child) *"Remember last week when we read the Three Little Pigs? Well today we are going to read another Three Little Pigs book, but from the perspective of the wolf."* Previewing the schedule *"Today we have a busy schedule. First, we have literacy centers, then we have read aloud, and then we will have a break for snack…"*	**Behavior Management** Commenting on good behavior *"I like the way Jack is sitting, I like the way Katie is sitting…"* Proactively providing behavior expectations and reminders (directed at group of children or individual child) *"I want you to walk over to your tables quietly and not talk to your neighbors when you sit down. Remember we don't touch the materials until I say so, ok? Even if they look fun to play with."* Responding to poor behavior and reorienting (directed at group of children or individual child) *"I'm going to wait for everyone to quiet down before I begin reading the story. I need everyone to sit crisscross applesauce and eyes on me."*
Modeling Modeling an instructional activity *"I'm going to show you how to play the number bingo game. You roll the die like this…"* Modeling suggestions for center time *"At the pretend play area we have a veterinarian clinic where you can weigh your animal like this* (teacher models)*, and then check its temperature like this…"* Modeling transitions *"Watch me. I want you to get up from the carpet like this and put your paper in the basket here, and then you are going to grab your activity bucket and bring it to your table like this. Got it?*	**Rituals & Routine Cues (directed at group of children)** Teacher verbal chant, action, or sound that signals students to orient their attention towards the teacher, refocus, or attend to a specific task *"1,2,3 eyes on me" (students: 1,2, eyes on you)* Movement, song, or other ritualistic activity that signals a transition to a new activity and/or helps orient children's attention and mindset to a new activity *A short song or movement sequence that students know to do when they finish transitioning* e.g., *(hands on their head when they sit on the carpet)*

Notes: To create a teacher organizational strategies composite, we summed the codes above. Each of these strategies could be directed at a group of children or an individual child.

Four members of our research team coded videos using the Noldus Observer XT 13 software for the ISI observation measures (Noldus Information Technology, 2013). We calculated reliability in the Noldus Observer XT software which compares the duration of time (start and end time) of each code and the order/sequence of codes within a 15-s grace window. Our average Kappa was 0.76, a very similar level reliability as past ISI studies (e.g., average of 0.76 in Connor et al., 2009). For the four constructs of interest in the present study, interrater reliabilities were also good: 0.72 for verbal directions, 0.74 for modeling, 0.80 for ritual & routine cues, and 0.95 for behavior management.

For students with two observations (84%, Table 1), we summed their data across both observations days to create an aggregate measure. For the 16% of students with only one observation, we included an indicator variable to control for that characteristic. To operationalize the ISI measures, we used the duration of time spent in minutes in each code in our analytic models.

Child covariates

We used administrative data on child demographics from the school district in the fall of the prekindergarten year to capture student-level covariates. We first created a series of 0/1 indicators to describe children's race/ethnicity (Black, Hispanic, Asian, or Other Race/Ethnicity), with White as the reference group. We also used indicators for children's eligibility for free or reduced-price lunch (FRPL; 1 if eligible; 0 if not), gender (1 = female; 0 = not female) and DLL status (1 if parent reported a non-English language spoken at home; 0 = English spoken at home). Finally, we used age in years at the time of the fall 2016 assessment, as well as amount of time between assessments in the fall of 2016 and the spring of 2017. We chose these covariates to align with previous work done using this sample (Authors, 2020b) and previous work showing their relation to children's outcomes (Choi, Jeon, & Lippard, 2018; Powell, Son, File, & San Juan, 2010; Reardon & Portilla, 2016).

Family covariates

We used parent-reported demographic information as covariates in analyses. We coded variables as 1 if the characteristic described the parent and 0 if not. These variables indicated whether the parent had attended prekindergarten or Head Start, whether there was at least one parent in the home working full-time, whether the parent was married or lived with a partner, and whether the respondent was the child's mother or father (reference group is other relationship). We used continuous variables to describe the age of the child's mother at her first birth, the number of people living in the household, and the parent respondent's age in the fall of the prekindergarten year. We chose these covariates to align with previous work done using this sample (Authors, 2020b) and other prekindergarten studies (Bloom & Weiland, 2015; Puma et al., 2010).

Analytic approach

Missing data

Overall, there was a relatively low amount of missing data across study variables. As illustrated in Table 1, missingness among the assessed outcomes ranged from 1.5% to 1.9% across the PPVT, WJAP, and Digit Span assessments conducted in fall and spring. All parent covariates had less than 5.3% missing. We did not find evidence for systematic differences between the children missing and not missing data (Appendix B Table 3). Given the low levels of missingness, we present results using complete case analysis in our main set of results, resulting in a sample size of $N = 226$ for the regression models. However, as a robustness check, we used multiple imputation with Stata 16 (Graham, 2009) to impute child and parent covariates, fall assessment scores, the number of minutes observed, whether students had one or two observations, time spent in whole class instruction, and classroom organization as measured by the CLASS (see Appendix A for further discussion).

RQ 1: Descriptive analysis

To address our first research aim – examining the extent to which students spent time in types of organizational strategies – we calculated the average amount of time spent in each type of instruction and its associated standard deviation across the full sample. We also calculated intraclass correlations (ICCs) at the child, classroom, and school levels to describe the amount of between-classroom variation in organizational instruction compared to the amount of within-classroom organizational instruction.[4] We calculated the ICCs with random effects conditional on features of our ISI data collection – number of minutes observed, whether students had one observation, the amount of time spent in whole class instruction, and the amount of time spent in teacher disorganization.

Table 3
Descriptive statistics for classroom organizational strategies.

Organizational Strategy	Mean	SD	Minimum	Maximum	Percent Missing
Teacher organizational strategies total	**23.49**	**11.06**	**1.30**	**53.25**	**0%**
Verbal directions	16.52	8.26	1.30	41.07	0%
Modeling	2.53	3.08	0.00	19.62	0%
Behavior management	2.64	2.86	0.00	13.33	0%
Ritual & routine cues	1.78	2.25	0.00	15.77	0%

Notes: $N = 263$. Base sample restricted to students with positive total observed time.

RQ 2: Multi-level modeling

We used multi-level modeling to answer research question 2, exploring whether time spent in organizational instruction was associated with children's gains in receptive vocabulary, math, and executive functioning across the prekindergarten year. Because students ($N = 263$) in our sample were nested within classrooms ($N = 39$) nested within schools ($N = 19$), we first fit unconditional random effects for classrooms and schools. This process allowed us to calculate intraclass correlations (ICCs) differentiating within- and between-classroom variation in the outcomes and to examine the extent to which observations were non-independent. Based on the variation we observed, we used a three-level model with random intercepts for classrooms and schools (Snijders & Bosker, 2012).

For each outcome, we present three models. In model one, the key predictor was our measure of global classroom quality (CLASS organization), controlling for the baseline level of the outcome and child- and family-level covariates. In model two, we regressed each outcome on either the total amount of time spent in organizational instruction or the amount of time spent in the four types of organizational instruction entered into the model together. In our model two regressions, we controlled for the baseline level of the outcome, features of our ISI data collection (the total number of minutes observed, whether a student had only one observation, the amount of time spent in whole class instruction, and teacher disorganization), and child- and family-level covariates. In model three, we added a covariate capturing global classroom quality (CLASS Organization) to our second model and examined the stability of our point estimates for the organization strategies predictors. Given that our work is exploratory in nature rather than confirmatory, we do not adjust for multiple comparisons (Schochet, 2009). All of the variables in these models were measured at the child level, except for the global classroom quality (CLASS organization) which was measured at the classroom level.

Results

RQ 1: child-level measurement of teacher organizational strategies

As shown in Table 3, we found that students spent an average of 23.49 (standard deviation = 11.06) minutes of their total observed time in some form of teacher organizational strategies. The majority of this time (16.52 min, $SD = 8.26$) was spent in verbal directions while was the rest of time was spent in behavior management (2.64 min, $SD = 2.86$), modeling (2.53 min, $SD = 3.08$), and ritual & routine cues (1.78 min, $SD = 2.25$).

Additionally, as illustrated in Table 4, we found that most of the

[4] The formula provided for calculating the ICC as discussed in Snijders and Bosker (2012) and the resulting simulations as discussed in Bujang and Baharum (2017) indicate that an average of 7 students per classroom is sufficient to calculate intraclass correlations because there are 39 classrooms. Although the variance and resulting ICCs estimates are related to the sample size in general, 39 classrooms is sufficient to ensure accurate ICC estimates for the range that we find in our results, even with an average of 7 students per classroom.

Table 4
ICCs for ISI Codes.

	Total Organization	Verbal Directions	Modeling	Behavior Management	Ritual & Routine Cues
Child Level	0.17	0.22	0.16	0.11	0.06
Classroom Level	0.64	0.78	0.84	0.89	0.43
School Level	0.19	0.00	0.00	0.00	0.51

Note: Estimates represent percent of the total variation. Models fit controlling for total number of minutes coded, whether students had one or two observations, the amount of time spent in whole class instruction, and the amount of time spent in teacher disorganization.

variation in total organization was at the classroom level (64%), followed by the school (19%) and child (17%) levels. Across the types of organization, most of the variance in verbal directions (78%), modeling (84%), and behavior management (89%) was at the classroom level while the remaining variation (11%–22%) was at the child level. Most of the variance in ritual and routine cues was at the school level (51%), with 43% at the classroom level and just 6% at the child level. These findings show that our student-level observation measure captured a non-trivial amount of difference in student experiences, with some variation across type.

RQ 2: associations between subcomponents of teacher organization and gains in language, math, and executive functioning skills during the prekindergarten year

We found small, and some significant, correlations between the four types of teacher organization (ranging from $r = 0.01$–0.27), and between the four types of teacher organization and the CLASS classroom organization construct ($r = 0.00$–0.23). We present the child gains results in Table 5. For each outcome, Model 1 controls for baseline test scores, child- and parent-level covariates, and includes our question predictor, global classroom quality (CLASS). Model 2 includes controls for baseline test scores, ISI observational features (the number of minutes observed, whether a student had one or two observations, the amount of time spent in a whole class setting, and the amount of observed time spent in teacher disorganization), child-level covariates, and parent-level covariates, as well as our key questions predictors, ISI teacher organizational strategies. Model 3 adds the global measure of classroom organization (CLASS) to model 2 to assess stability of estimates across model specification. Our unstandardized estimates results are presented in Table 5, and we calculated the Beta standardized coefficients in order to more easily compare across measures in our text discussion.

As shown in Table 5, we found that classroom organization (CLASS) was not statistically significantly related to any of our outcomes (Panels A and B, Models 1, 4, 7, 10). The magnitude of the Beta standardized coefficients for classroom organization ranged from −0.01 (receptive vocabulary) to 0.02 (math measured by the WJAP). When considering the teacher organizational strategies, the relation between our total organizational strategies composite and gains in all outcomes were also null (Panel A, Models 2, 5, 8, 11). These results were unchanged when we combined the total organizational composite measure and classroom organization into the same model (Panel A, Models 3, 6, 9, 12).

When examining the four types of organizational strategies, relations between most types of organizational strategies and child gains likewise were null (Panel B, Models 2, 5, 8, 11), with one exception. Specifically, we found evidence of a statistically significant positive linear relationship between the time exposed to modeling and gains in math as measured by the WJAP ($\gamma = 0.21$, $SE = 0.07$, $p = 0.003$, $\beta = 0.15$, Panel B, Model 5). This association was stable when controlling for CLASS classroom organization ($\gamma = 0.22$, $SE = 0.07$, $p = 0.003$, $\beta = 0.16$, Panel B, Model 6). The other models combining the teacher organizational strategies and classroom organization had null results (Panel B, Models 3, 6, 9, 12). Full regression results for the statistically significant findings are presented in Appendix B Table 4, and the remaining full regression results are available upon request from the authors.

Robustness checks

We conducted multiple sensitivity analyses to address eleven different threats to the internal validity of our results and to determine

Table 5
Linear associations between types of organizational strategies and gains in children's skills.

	Receptive Vocabulary (PPVT Raw Score)			Math (WJAP Raw Score)			Math (REMA Raw Score)			Executive Functioning (Digit Span Categorical Score)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A. Total organization construct												
Predictors												
Total organization		0.09	0.09		0.02	0.02		0.02	0.01		−0.01	−0.01
		(0.15)	(0.15)		(0.03)	(0.03)		(0.06)	(0.06)		(0.01)	(0.01)
Classroom organization	−0.18		−0.46	0.12		−0.01	0.20		0.16	0.00		0.02
	(1.88)		(2.35)	(0.39)		(0.43)	(0.71)		(0.85)	(0.11)		(0.13)
Panel B. Type of organizational strategies												
Predictors												
Verbal directions		0.19	0.19		−0.03	−0.03		0.02	0.02		−0.01	−0.01
		(0.19)	(0.20)		(0.04)	(0.04)		(0.07)	(0.07)		(0.01)	(0.01)
Modeling		0.09	0.10		0.21**	0.22**		0.16	0.17		0.02	0.02
		(0.40)	(0.41)		(0.07)	(0.07)		(0.15)	(0.15)		(0.02)	(0.02)
Behavior management		0.49	0.51		0.10	0.10		−0.13	−0.13		−0.03	−0.03
		(0.48)	(0.48)		(0.09)	(0.09)		(0.18)	(0.18)		(0.03)	(0.03)
Routine & ritual cues		−1.21	−1.20		−0.04	−0.04		−0.16	−0.16		−0.02	−0.02
		(0.68)	(0.68)		(0.12)	(0.12)		(0.26)	(0.26)		(0.04)	(0.04)
Classroom organization	−0.18		−0.46	0.12		−0.18	0.20		−0.04	0.00		−0.02
	(1.88)		(2.33)	(0.39)		(0.43)	(0.71)		(0.86)	(0.11)		(0.14)

Notes: Standard errors in parentheses. $N = 226$. Statistical significance levels are indicated as *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Results are unstandardized estimates. All models control for child and family covariate conceptual blocks and the baseline measure of the outcome. Additionally, the models with teacher organization also control for the number of minutes observed, whether a student had one or two observations, the amount of time spent in whole class, and the amount of time spent in teacher disorganization.

the robustness of our child gains findings. Specifically, we fit models to test robustness to possible multicollinearity among the teacher organizational strategies; missing data decisions; using raw scores versus other versions of academic achievement; time spent in teacher- and child-managed instruction; time spent in off-task behavior; a nonlinear relationship between teacher organizational strategies and child gains; the error structure of models; differences by student subgroups; analytic decisions for students who did not pass the fall PreLAS screener; the operationalization of the teacher organization variables; and aggregating the ISI teacher organization measures to the classroom level in order to compare more directly to the CLASS measure. A full description of these checks is included in Appendix A. All of our results were robust except for the models fit using the WJAP W score and the REMA T and IRT scores, and the classroom-level aggregated ISI measures. The results from these alternative outcome models were not robust in terms of either statistical significance or the magnitude of the standardized coefficient. Overall, most of our robustness checks were stable in terms of magnitude and statistical significance, and the models in which we explored robustness to different error structures using teacher fixed effects imply larger magnitude between time spent in modeling and math gains (β = 0.15 to 0.24). These results show some sensitivity to the operationalization of math achievement and should be interpreted with caution.

Discussion

This is one of the first studies to closely examine prekindergarteners' individual exposure to teacher organizational strategies such as directions, cues, modeling, and behavior management, and the first study (to our knowledge) to compare this approach to a global measure of classroom organization. Supporting the efficacy of measuring multiple dimensions of teacher organizational strategies at the child level, we found that children's time spent exposed to teacher organizational strategies varied between classrooms and individual children, and that this variation was larger for some organization subcomponents than others. In a context in which global classroom organization did not predict gains in children's school readiness skills (Authors, 2020), we also found that our composite teacher organization strategies construct was only modestly statistically significantly correlated with the composite CLASS organization measure (r = 0.26; see Appendix B Table 2). Controlling for CLASS organization, one child-level subcomponent of teacher organization (modeling) statistically significantly predicted children's gains in math skills.

Despite our largely null findings, this study offers several important implications for measuring children's classroom experiences and their contribution to school readiness skills. For example, researchers have widely found that CLASS classroom organization does not consistently predict preschoolers' gains, despite its wide use in policy accountability systems (e.g., Authors, 2020). As the field accordingly turns to potential new directions, measurement at the child level is one of the leading suggestions for a path forward (see Burchinal, 2018; Weiland, 2018). Our study thus represents path-breaking work that shows that this approach has some merit and some advantages over the CLASS but it unlikely to solve the problem entirely.

Exploring measurement of classroom organization at the child level

In this study we were interested in exploring whether individual children experienced different amounts of exposure to teacher organizational support in prekindergarten. Our results highlight that teachers provided more organizational support to some children more than others, but only with specific organizational supports. For example, a key contribution of our work highlights the significant variation in the amount of time individual children spent receiving *verbal directions* from teachers. This variation may be due to the fact that *verbal directions* comprised the majority of teacher organizational time children were exposed to (16 of 24 min) so there was more "opportunity" for children to experience different amounts. Furthermore, some children may have asked for more directions from the teacher, or the teacher could have been responding to children's individual needs in the classroom. Thus, it may be important to continue measuring this aspect of classroom organization at the child level, but examine nuance in why and when teachers are providing more *verbal directions* to some children than others. Further breaking down this construct into subtypes of verbal directions may explain more variation in children's outcomes, especially given the prevalence of this teacher organization technique in early childhood settings (U.S Department of Education, 2002; Wen, Elicker, & McMullen, 2011).

Less time was spent in — and less within classroom variation was present in — the other three organizational techniques (e.g., *behavior management*, *rituals & routines*, and *modeling*), because it was often the case that these strategies did not consume a lot of time regardless of its frequency. However, the short durations of time noted in these constructs may be explained by prior work on how time spent in classroom organization more broadly should decrease throughout the school year (Cameron et al., 2008). Considering that our study collected observation data in the winter and spring of prekindergarten—when classroom expectations should be established, and children should be beginning to internalize supports and self-regulating—it is not surprising that children only experienced short amounts of time in behavior management and teachers' modeling materials and modeling transitions. This may represent a positive finding in our contexts because with less time spent in behavior management, there is hypothetically more room for meaningful instruction (Burchinal, 2018). The limited within-class variation on these three teacher organizational strategies suggests that pursuing more detailed and tedious child-level measurement of these phenomena is not necessary.

Exploring the separation of classroom organization into subcomponents

A second aim of our study was to explore child exposure to different components of teacher organization support. Our results demonstrate that a global, single construct of teacher organization did not predict children's gains in math, language and executive functioning skills, but specific subcomponents did. For example, teacher *modeling* materials before an activity or a transition predicted gains in children's math skills. The magnitude of this association – about 0.15 standard deviations—was even larger in some alternative models that we fit as part of our sensitivity analysis (Appendix A). However, it is important to note that this was not significantly robust to all forms of scoring on this math measure, and that given the number of models performed and the exploratory nature of our study, findings should be interpreted with caution. It is possible that this result may support classroom management practices structured to first show children how to use materials before an instructional activity begins. Doing so can help ensure that children are organized in handling the materials on their own and are thus better equipped at achieving the learning goal of the activity (Darch & Kame'enui, 2003; Hemmeter, Ostrosky, Artman, & Kinder, 2008). In our specific language diverse sample (55% dual-language learners), modeling behaviors from the teacher (showing vs. just telling) may be particularly beneficial for this population of students who may need more support beyond verbal input (Espinosa, 2013).

Past work in early elementary schools has also linked teacher modeling to more behavioral engagement during math lessons specifically (Lan et al., 2009). This modeling is particularly important for math learning in the prekindergarten context because these manipulatives can be easily viewed as toys, so careful modeling of the appropriate use of these materials can keep children more on-task, especially during center time when they may be working independently or with a peer (McLennan, 2014). The children in our sample were also exposed to the Building Blocks math curriculum (Clements & Sarama, 2007), which emphasizes the use of math manipulatives. However, it appears that

gains in math were not simply an effect of exposure to this curriculum as results remained robust when controlled for time spent in math activities. Additionally, when we took a closer look at when teacher modeling and organizing of materials occurred, only 36% of child exposure to this construct happened before a math activity. The rest of teacher modeling occurred prior to center time (e.g., mixed content; 40%), transitions (13%) and literacy activities (11%). Future research should continue to study the influence of teacher modeling and organization of materials on children's learning.

It is important to highlight, too, that only 16% of variability in modeling was attributable to within classroom differences and the rest was due to variation at the teacher level. Thus, our only statistically significant finding appears to be a measure that is more teacher-driven than child-elicited. In fact, when we aggregated modeling to the classroom level we found it remained a robust predictor of children's math skills gains. This finding suggests that it may be more beneficial for the field to create global measures of teacher modeling materials and transitions for children than to focus on child-level measurement of this construct specifically.

Although not statistically significant at the child level, *classroom-level* exposure to ritual and routine cues (e.g., "1, 2, 3, eyes on me!") negatively predicted children's language gains. Interestingly, a majority of variation appeared at the school level which suggests that this teacher organizational strategy may be learned or emphasized between teachers within a school. To date, no study has examined the effects of these song-like cues on children's outcomes, so it is difficult to interpret and connect our finding to the literature. However, considering teachers rarely use verbal language during routine cues (and instead use short phrases, gestures, or music), it may be that the decreased verbal input is explaining the negative relation to children's language gains. It is important to highlight, too, that although statistically significantly correlated with duration of time, there was far more variability when we analyzed *frequency count* of ritual cues ($M = 7.4$ times; $SD = 22.8$). Thus, future work may benefit from examining different conceptualizations of measuring children's "quantity" of exposure to organization strategies that appear in short spurts throughout the school day (e.g., duration vs. frequency counts).

The other two types of teacher organizational strategies—*behavior management*, and *verbal directions*—did not predict gains (examined at either the child or classroom level) in any child outcomes. We originally hypothesized that *behavior management* would support children's EF skills given theoretical evidence on external behavior supports helping the development of children's internal behavior supports (Vygotsky, 1978). It is possible, however, that our null finding was because we did not differentiate between reactive and proactive behavior management techniques. For example, more reactive behavior management techniques (e.g., "you are too loud, please use your inside voices") would be elicited by the child(ren), whereas proactive behavior management is more driven by the teacher (e.g., "remember to use your inside voices when we step into the hall"). Past work has shown that proactive behavior reminders may be more likely to scaffold children's EF skills, whereas reactive strategies may be more indicative of a teacher simply responding to poor self-regulation with consequences which does not appropriately scaffold the development of self-regulation (Clunies-Ross, Little, & Kienhuis, 2008; Park & Lynch, 2014). For *verbal directions*, we hypothesized that this facet of classroom organization might support children's language skills because of the verbal input, but this was also not the case ($\beta = 0.06$). Notably, *verbal directions* was perhaps the broadest category and encompassed a number of different strategies such as previewing the schedule of the day, providing step-by-step directions to students, providing general directions that do not detail every step, or just giving blanket statements such as "time to come to the carpet" or "let's line up." It may be useful for future work to differentiate this strategy type further and examine predictions to gains in child outcomes, especially since a majority of organizational time is dedicated to this strategy and we found moderate within-classroom variation (22%).

Limitations

Our study has several important limitations. First, there is a limitation of our observational measure. Although comprehensive and reliable, the ISI observation measure is time-intensive to code because of the continuous focus on each study child in the classroom, and as a result, it can often take 8 h to code 2.5 h of classroom footage. As such, we could only feasibly focus on a sample of students per classroom, which poses issues of power and the generalizability of our findings. Additionally, it was only feasible for us to focus on a subset of constructs related to classroom organization, which also limits the generalizability of our findings to our four specific teacher organizational strategy codes. The measure also requires teachers to consent to videotaping, which can create barriers for data collection. Consequently, this measure may not be particularly appealing to other researchers or practitioners in the field of ECE and is certainly not feasible to live-code. Recent efforts have been made to adapt the ISI observation measure for practical use in ECE settings (Optimizing Student Learning Opportunities (OLOS); Connor et al., 2020). However, the OLOS is not yet validated and the current version of that tool does not distinguish between types of teacher organizational strategies. As with most classroom observation studies, it is also difficult to tease apart each child from the broader classroom context. As such, we do not know the extent to which individual children elicit certain organizational strategies from the teacher, or if teacher organizational techniques are more proactively organizational rather than reactive. Teachers' use of organizational strategies may also differ across the year and may be more influential on child outcomes at the beginning of the year. For example, "more" behavior management and verbal direction in winter/spring of the school year may not be a good thing, as most behavior expectations should be established in the beginning of the year. Other organizational techniques, like modeling and routine cues, may be less likely to change in frequency and duration across the school year. Further research may consider this hypothesis with data collected in the fall of the academic year.

Another limitation of our study is our choice of child direct assessments. One of the most surprising findings of our paper was the null relation between organizational strategies and children's EF skills. It is possible that this finding reflects our choice of EF outcome measure (Digit Span), which is a precursor skill to the working memory domain of EF, the least malleable dimension of EF (Redick, Shipstead, Wiemers, Melby-Lervag, & Hulme, 2015). We also did not have a fall score for the REMA (and instead used the WJ Applied Problem score in its place). The lack of robustness of our modeling findings across all versions of REMA scores might have been due to this issue. Additionally, the relations between time spent in modeling and gains in math as measured by the WJAP is confounded by the Focus on K1 curriculum, despite our robustness checks to eliminate this possibility.

Lastly, our study is correlational and not causal. Due to our exploratory aims, we did not adjust for multiple tests because we examined four outcomes (receptive vocabulary, two measures of math, and executive functioning) with six primary predictors (CLASS, total organization, verbal directions, modeling, behavior management, and routine and ritual cues), follow Schochet (2009). Future confirmatory work should adjust for multiple comparisons.

Conclusion

Despite these limitations, our findings have implications for the ongoing search for the active ingredients in prekindergarten that promote children's school readiness and for a next generation of measurement work in early childhood settings (Burchinal, 2018; Weiland, 2018). From a policy perspective, it is important to isolate the key factors of children's classroom experiences that matter for child outcomes, as well as highlight any issues of equity that may be occurring between children

within the same classroom. We found that teachers' use of specific types of organization strategies varied across classrooms and across individual children. We also found that one particular facet of classroom organization, measured at the child level, predicts gains in children's math skills. Importantly, our context is one in which multiple past studies have found that *global quality* measures do not predict child gains in language, math, and executive function (Authors, 2013; Authors, 2020). Together, our findings point to possibly fruitful new directions in early childhood measurement particularly when considering the role of classroom organization – and the specific dimensions that make up this domain of classroom quality – in supporting young children's development.

Author statement

Lillie Moffett: Conceptualization, Writing-Original draft. Amanda Weissman: Methodology, Formal analysis, Writing-Original draft. Christina Weiland: Writing-Review & editing, Supervision, Investigation, Funding acquisition. Meghan McCormick: Writing-Review & editing, Supervision, Investigation, Funding acquisition. JoAnn Hsueh: Investigation, Funding acquisition. Catherine Snow: Funding acquisition. Jason Sachs: Investigation, Project administration

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.appdev.2021.101346.

References

Authors (2013).
Authors (2020).
Authors (2020b).
Authors (2021a).
Authors (2021b).

Barrueco, S., Lopez, M., Ong, C., & Lozano, P. (2012). *Assessing Spanish-English bilingual preschoolers: A guide to best approaches and measures*. Baltimore, MD: Paul H Brookes Publishing.

Bloom, H. S., & Weiland, C. (2015). Quantifying Variation in Head Start Effects on Young Children's Cognitive and Social-Emotional Skills Using Data from the National Head Start Impact Study. Available at SSRN https://ssrn.com/abstract=2594430. https://doi.org/10.2139/ssrn.2594430.

Bono, K. E., & Bizri, R. (2013). The role of language and private speech in preschoolers' self-regulation. *Early Child Development and Care, 184*(5), 658–670.

Brophy, J. E. (1979). Teacher behavior and its effects. *Journal of Educational Psychology, 71*, 733–750.

Brophy, J. E. (1983). Classroom organization and management. *The Elementary School Journal, 83*(4), 264–285.

Bujang, M. A., & Baharum, N. (2017). A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: A review. *Archives of Orofacial Sciences, 12*(1), 1–11.

Bull, R., Epsy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology, 33*(3), 205–228.

Bulotsky-Shearer, R. J., Fantuzzo, J., & McDermott, P. A. (2008). An investigation of classroom situational dimensions of emotional and behavioral adjustment and cognitive and social outcomes for head start children. *Developmental Psychology, 44*(1), 139–154.

Burchinal, M. (2018). Measuring early care and education quality. *Child Development Perspectives, 12*(1), 3–9.

Burchinal, M., Garber, K., Foster, T., Bratsch-Hines, M., Franco, X., & Peisner-Feinberg, E. (2021). Relating early care and education quality to preschool outcomes: The same or different models for different outcomes? *Early Childhood Research Quarterly, 55*(1), 35–51.

Burchinal, M., Vernon-Feagans, L., Vitiello, V., Greenberg, M., & The Family Life Project Key Investigators. (2014). Thresholds in the association between child care quality and child outcomes in rural preschool children. *Early Childhood Research Quarterly, 29*(1), 41–51.

Cameron, C. E., Connor, C. M., Morrison, F. J., & Jewkes, A. M. (2008). Effects of classroom organization on letter-word reading in first grade. *Journal of School Psychology, 46*(2), 173–192.

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development, 11*(2), 73–92.

Chien, N. C., Howes, C., Burchinal, M., Pianta, R. C., Ritchie, S., Bryant, D. M., … Barbarin, O. A. (2010). Children's classroom engagement and school readiness gains in prekindergarten. *Child Development, 81*(5), 1534–1549.

Choi, J. Y., Castle, S., Williamson, A. C., Young, E., Worley, L., & Long, M…& Horm, D. M.. (2016). Teacher-child interactions and the development of executive function in preschool-age children attending head start. *Early Education and Development, 27*(6), 751–769.

Choi, J. Y., Jeon, S., & Lippard, C. (2018). Dual language learning, inhibitory control, and math achievement in head start and kindergarten. *Early Childhood Research Quarterly, 42*, 66–78.

Clements, D. H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the building blocks project. *Journal for Research in Mathematics Education*, 136–163.

Clements, D. H., Sarama, J. H., & Liu, X. H. (2008). Development of a Measure of Mathematics Achievement Using the Rasch Model: The Research-Based Early Maths Assessment. *Journal of Educational Psychology, 28*(4), 457–482.

Clunies-Ross, P., Little, E., & Kienhuis, M. (2008). Self-reported and actual use of proactive and reactive classroom management strategies and their relationship with teacher stress and student behavior. *Educational Psychology, 28*(6), 693–710.

Connor, C. M., Adams, A., Zargar, E., Wood, T. S., Hernandez, C., & Vandell, D. L. (2020). Observing individual children in early childhood classrooms using optimizing learning opportunities for students (OLOS): A feasibility study. *Early Childhood Research Quarterly, 52B*(3), 74–89.

Connor, C. M., Morrison, F. J., Fishman, B. J., Ponitz, C. C., Glasney, S., Underwood, P. S., Piasta, S. B., Crowe, E. C., & Schatschneider, C. (2009). The ISI classroom observation system: Examining the literacy instruction provided to individual students. *Educational Researcher, 38*(2), 85–99.

Connor, C. M., Morrison, F. J., & Slominski, L. (2006). Preschool instruction and Children's emergent literacy growth. *Journal of Educational Psychology, 98*(4), 665–689.

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

Downer, J. T., Lopez, M. L., Grimm, K. J., Hamagami, A., Pianta, R. C., & Howes, C. (2012). Observation of teacher-child interactions in classrooms serving Latinos and dual language learners: Applicability of the classroom assessment scoring system in diverse settings. *Early Childhood Research Quarterly, 27*(1), 21–32.

Duncan, S. E., & DeAvila, E. A. (1998). *PreLAS*. Monterey, CA: CBT McGraw Hill.

Dunn, L. M., & Dunn, L. M. (2007). *PPVT-IV: Peabody picture vocabulary test*. American Guidance Service.

Espinosa, L. M. (2013). *Early education for dual language learners: Promoting school readiness and early school success*. Washington, DC: Migration Policy Institute.

Farran, D. C., Aydogan, C., Kang, S. J., & Lipsey, M. W. (2006). Preschool classroom environments and the quantity and quality of children's literacy and language behaviors. *Handbook of early literacy research, 2*, 257–268.

Fisher, C., Berliner, D., Filby, N., Marliave, R., Cahen, L., & Dishaw, M. (1980). Teaching behavior, academic learning time, and student achievement: An overview. In C. Denham, & A. Lieberman (Eds.), *Time to learn*. Washington, D.C: National Institute of Education.

Fuhs, M. W., & Day, J. D. (2011). Verbal ability and executive functioning development in preschoolers at head start. *Developmental Psychology, 47*(2), 404–416.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences, 3*(11), 410–419.

Gathercole, S. E., & Pickering, S. J. (2000). Assessment of Working Memory in Six- and Seven-Year-Old Children. *Journal of Educational Psychology, 92*(2), 377–390.

Gormley, W. T., Jr., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-K on cognitive development. *Developmental Psychology, 41*(6), 872.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549–576.

Hamre, B., Hatfield, B., Pianta, R., & Jamil, F. (2014). Evidence for general and domain-specific elements of teacher child interactions: Associations with preschool Children's development. *Child Development, 85*(3), 1257–1274.

Hatfield, B. E., Burchinal, M. R., Pianta, R. C., & Sideris, J. (2016). Thresholds in the association between quality of teacher-child interactions and preschool children's school readiness skills. *Early Childhood Research Quarterly, 26*(3), 561–571.

Hemmeter, M. L., Ostrosky, M. M., Artman, K. M., & Kinder, K. A. (2008). Moving Right Along . . . Planning Transitions to Prevent Challenging Behavior. *Young Children, 63*(3), 18–25.

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*(5), 409–426.

Lan, X., Ponitz, C. C., Miller, K. F., Li, S., Cortina, K., Perry, M., & Fang, G. (2009). Keeping their attention: Classroom practices associated with behavioral engagement in first grade mathematics classes in China and the United States. *Early Childhood Research Quarterly, 24*, 198–211.

Levya, D., Weiland, C., Barata, M., Yoshikawa, H., Snow, C., Trevino, E., & Rolla, A. (2015). Teacher-Child. *Interactions in Chile and their Associations with Prekindergarten Outcomes, 86*(3), 781–799.

Lipsey, M. W., Nesbitt, K. T., Farran, D. C., Dong, N., Fuhs, M. W., & Wilson, S. J. (2017). Learning-related cognitive self-regulation measures for prekindergarten children: A comparative evaluation of the educational relevance of selected measures. *Journal of Educational Psychology, 109*, 1084–1102.

McLennan, D. P. (2014). Making math meaningful for young children. *Teaching Young Children, 8*(1).

NAEYC. (2009). *Developmentally appropriate practice in early childhood programs serving children from birth through age 8*.

Noldus Information Technology. (). *Noldus Observer XT. [Software]*. Available at: http://www.noldus.com/human-behavior-research/products/the-observer-xt.

Park, H. S. L., & Lynch, S. A. (2014). Evidence-based practices for addressing classroom behavior problems. *Young Exceptional Children, 17*(3), 33–47.
Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., & Tazejian, N. (2003). The relation of preschool child-care quality to Children's cognitive and social developmental trajectories through second grade. *Child Development, 72*(3), 1534–1553.
Perlman, M., Falenchuk, O., Fletcher, B., McMullen, E., Beyene, J., & Shah, P. S. (2016). A systematic review and meta-analysis of a measure of staff/child interaction quality (the classroom assessment scoring system) in early childhood education and care settings and child outcomes. *PLoS One, 11*(12), Article e0167660.
Hamre, B. K., & Pianta, R. C. (2001). Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development, 72*(2), 625–638.
Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*(2), 109–119.
Pianta, R. C., LaParo, K., & Hamre, B. (2008). *Classroom assessment scoring system (CLASS)*. Charlottesville, VA: University of Virginia.
Pianta, R. C., Whittaker, J. E., Vitiello, V., Ruzek, E., Ansari, A., Hofkens, T., & DeCoster, J. (2020). Children's school readiness skills across the pre-k year: Associations with teacher-student interactions, teacher practices, and exposure to academic content. *Journal of Applied Developmental Psychology, 66*, 1–10.
Powell, D. R., Son, S. H., File, N., & San Juan, R. R. (2010). Parent–school relationships and children's academic and social outcomes in public school pre-kindergarten. *Journal of School Psychology, 48*(4), 269–292.
Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P., Jenkins, F., Fletcher, P., Quinn, L., Friedman, J., Ciarico, J., Rohacek, M., Adams, G., & Spier, E. (2010). Head Start Impact Study. *Final Report*.
Reardon, S. F., & Portilla, X. A. (2016). Recent trends in income, racial, and ethnic school readiness gaps at kindergarten entry. *AERA Open, 2*(3), 1–18.
Redick, T. S., Shipstead, Z., Wiemers, E. A., Melby-Lervag, M., & Hulme, C. (2015). What's working in working memory training? An educational perspective. *Educational Psychology Review, 27*, 617–633.
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*(4), 958–972.
Rimm-Kaufman, S. E., Fan, X., Chiu, Y., & You, W. (2007). The contribution of the responsive classroom approach on children's academic achievement: Results from a three-year longitudinal study. *Journal of School Psychology, 45*(4), 401–421.
Sainato, D. M. (1990). Classroom transitions: Organizing environments to promote independent performance in preschool children with disabilities. *Education and Treatment of Children, 13*, 288–297.
Schickedanz, J. A., Dickinson, D. K., & Charlotte-Mecklenburg Schools. (2005). *Opening the world of learning: A comprehensive early literacy program*. Pearson: Parsippany.
Schochet, P. Z. (2009). An approach for addressing the multiple testing problem in social policy impact evaluations. *Evaluation Review, 33*(6), 539–567.
Snijders, T. A. B., & Bosker, R. J. (2012). Discrete dependent variables. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*, 304–307.
Tout, K., Magnuson, K., Lipscomb, S., Karoly, L., Starr, R., Quick, H., & Wenner, J. (2017). *Validation of the Quality Ratings Used in Quality Rating and Improvement Systems (QRIS): A Synthesis of State Studies. OPRE Report #2017-92. Washington, DC: Office of Planning, Research and evaluation, Administration for Children and Families*. U. S. Department of Health and Human Services.
U.S Department of Education. (2002). *The Early Childhood-Head Start Task Force. Chapter 4 "Creative a Learning Environment for Young Children"*.
Vitiello, V. E., Booren, L. M., Downer, J. T., & Williford, A. P. (2012). Variation in children's classroom engagement throughout a day in preschool: Relations to classroom and child factors. *Early Childhood Research Quarterly, 27*(2), 210–220.
Vukelich, C. (1991). *Materials and modeling: Promoting literacy during play*. In James F. Christie's Play and Early Literacy Development: SUNY Press.
Vygotsky, L. S. (1978). Interaction between learning and development (M. Lopez-Morillas, trans.). In M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds.), *Mind in society: The development of higher psychological processes* (pp. 79–91). Cambridge, MA: Harvard University Press.
Webster-Stratton, C., Reid, J. M., & Stoolmiller, M. (2008). Preventing conduct problems and improving school readiness: Evaluation of the incredible years teacher and child training programs in high-risk schools. *Child Psychology and Psychiatry, 49*(5), 471–488.
Weiland, C. (2018). Commentary: Pivoting to the "how": Moving preschool policy, practice, and research forward. *Early Childhood Research Quarterly, 45*, 188–192.
Weiland, C., Barata, M. C., & Yoshikawa, H. (2014). The co-occurring development of executive function skills and receptive vocabulary in preschool- aged children: A look at the direction of the developmental pathways. *Infant and Child Development, 23*, 4–21.
Weiland, W., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive functioning, and emotional skills. *Child Development, 84*(6), 2112–2130.
Wen, X., Elicker, J. G., & McMullen, M. B. (2011). Early Childhood Teachers' Curriculum Beliefs: Are They Consistent With Observed Classroom Practices? *Early Education & Development, 22*(6), 945–969.
Williford, A. P., Maier, M. F., Downer, J. T., Pianta, R. C., & Howes, C. (2013). Understanding how children's engagement and teachers' interactions combine to predict school readiness. *Journal of Applied Developmental Psychology, 34*(6), 299–309.
Winsler, A. De Leon, J.R. Wallace, B.A., Carlton, M.P…Willson-Quayle, A.. (2003). Private speech in preschool children: Developmental stability and change across-task consistency, and relations with classroom behavior. *Journal of Child Language, 30*(3), 583–608.
Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *The Woodcock-Johnson Tests of Achievement—Third Edition (WJ—III)*. Itasca, IL: Riverside Publishing Company.
Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An Effectiveness-Based Evaluation of Five State Pre-Kindergarten Programs. *Journal of Policy Analysis and Management, 27*(1), 122–154.
Xu, F., Han, W., Sabbagh, M. A., Wang, T., Ren, X., & Li, C. (2013). Development Differences in the Structure of Executive Function in Middle Childhood and Adolescence. *PLoS One, 8*(10), Article e77770.