Published in final edited form as:

Early Child Res Q. 2014 ; 29(4): 682–691. doi:10.1016/j.ecresq.2014.07.007.

Implications of different methods for specifying classroom composition of externalizing behavior and its relationship to social–emotional outcomes

Monica Yudron^{a,*}, Stephanie M. Jones^a, and C. Cybele Raver^b

^aHarvard University, United States

^bNew York University, United States

Abstract

In this paper, we examine common methods for using individual-level data to represent classroom composition by examining exemplary studies that thoughtfully incorporate such measures. Building on these studies, and using data from the Chicago School Readiness Project (CSRP), this paper examines theoretical and analytical implications of a set of different transformations of individual ratings of child externalizing behaviors in order to examine and compare the influence of these representations of classroom composition on Kindergarten internalizing behaviors, social competence, and attention/impulsivity problems. Results indicate that each Kindergarten outcome is influenced by distinct aspects of classroom composition of externalizing behaviors. Kindergarten internalizing behaviors are positively associated with the proportion of children in the Head Start classroom who started with externalizing scores above the 75th percentile regardless of the average value of externalizing behaviors in the classroom. In contrast, Kindergarten social competence is predicted by three aspects of the classroom distribution of externalizing behaviors in the fall of Head Start—the classroom mean, standard deviation, and skew. Finally, Kindergarten attention/impulsivity problems were not associated with any aspect of classroom composition of externalizing behavior examined in this paper.

Keywords

Classroom context; Social-emotional development; Methodology; Intervention; Externalizing behavior; Head Start

Introduction

Researchers in the developmental and prevention sciences are increasingly interested in the effects of classroom composition (DeRosier, Cillessen, Coie, & Dodge, 1994; Dmitrieva, Steinberg, & Belsky, 2007; Justice, Petscher, Schatschneider, & Mashburn, 2011; Mashburn, Justice, Downer, & Pianta, 2009; Moller, Forbes-Jones, Hightower, & Friedman, 2008; Thomas, Bierman, & Powers, 2011). In this paper, we use the phrase *classroom composition*

*Corresponding author at: Harvard University, 14 Appian Way, Office 606, Cambridge, MA 02138, United States. Tel.: +1 512 743 9367. Monica_Yudron@mail.harvard.edu (M. Yudron).

to refer to the characteristics of the peer group with whom a child shares a classroom. We will use it to refer to classroom-level characteristics such as externalizing behavior problems. In the field of economics, studies of classroom characteristics such as classroom size, aggregate language skills, or average socioeconomic status have featured prominently in peer effects studies (Bonesrønning, 2008; Glewwe, 1997; Hanushek, Kain, Markman, & Rivkin, 2003; Lavy & Schlosser, 2011; Neidell & Waldfogel, 2008). Yet, in other fields, classroom composition is understudied despite a long tradition of general theories of human development that posit a central role of context, including micro-contexts such as classrooms, in generating individual child developmental outcomes (Bronfenbrenner, 1979; Bronfenbrenner & Morris, 2006; Sameroff, 2010). Indeed, current theoretical frameworks offer little guidance in how to operationalize features of context such as classroom composition of social–emotional characteristics in empirical work.

In recent years, studies of child development, particularly evaluations of preschool and school-based interventions, have measured and modeled context as a key predictor of a range of child outcomes. These interventions often leverage changes in classroom context to alter child outcomes (Raver, Jones, Li-Grining, Zhai, Metzger, & Solomon, 2009; Raver, Jones, Li-Grining, Zhai, Bub, & Pressler, 2011). While interest in how classroom composition contributes to heterogeneity in child outcomes has increased, the development of a shared set of methods to account for this important environmental characteristic have lagged due to a range of issues including the lack of (1) theory regarding how specific characteristics of classrooms influence individual outcomes, and (2) consensus among researchers regarding what aspects of context matter most and whether this varies from one characteristic to another. In this paper, we examine common methods for using individual-level data to represent classroom composition by examining exemplary studies that thoughtfully incorporated contextual measures. Building on this work, we propose an additional method to transform individual-level ratings of child behavior to the classroom-level. Then, using data from the Chicago School Readiness Project (CSRP) as an illustrative example, we present theoretical and analytical implications of a range of transformations of individual ratings of child externalizing behaviors in order to first determine the implications of each of these transformations and then to unpack the influence of this Head Start classroom compositional measure on important Kindergarten outcomes.

Exemplar investigations of classroom context

Bioecological theories of human development emphasize the importance of multifaceted, dynamic interactions between an individual and her social and physical environment. Bronfenbrenner (1979) set forth a comprehensive theory of human development in which an individual is nested in a series of contexts that move from proximal and therefore direct in their influence, to distal and indirect. For children under five, out-of-home care experiences are highly salient for development and learning as these are the contexts in which children may establish and develop their first non-familial adult and peer relationships. The adults in these settings—the teachers and teacher assistants—are often able to provide an important and accurate depiction of child behaviors and characteristics for several reasons including (1) the frequency and intensity of teachers' interactions with each child as she is interacting with peers and the physical environment and (2) the early childhood development and

learning acumen of teachers built through training and experience with children (Ladd & Profilet, 1996).

Bronfenbrenner's framework and those that incorporate it (Sameroff, 2010) provide important groundwork for investigating the role of environmental composition but offer little guidance regarding how to carry out these investigations. In fact, with few exceptions, researchers who are interested in the role of classroom composition in child outcomes operate without the guidance of broadly tested empirical frameworks regarding how to operationalize classroom composition and link it to child outcomes. In this paper, we examine commonly used methods for operationalizing classroom composition as a means to generate specific hypotheses to be tested in future empirical work.

Using measures of heterogeneity to represent classroom composition—The developmental theories of Piaget and Vygotsky posit specific hypotheses regarding the role of age composition in preschool peer groups. These theories have guided the operationalization of classroom age composition in studies that have investigated the association between this classroom characteristic and individual child outcomes. Researchers have used what they view as a conflict between these theorists to generate testable hypotheses. For example, interpreting Piaget's notion that children learn best when cooperating with same-age peers to construct a shared meaning of the environment, researchers propose that children benefit most when the age spread of a peer group is quite small (Piaget, 1983). However, Vygotsky posits that children benefit most when there is age heterogeneity in a child's peer group since younger children are challenged by older peers and older children learn from modeling and caring for younger peers (Vygotsky, 1987).

Piaget and Vygotsky's theories suggest that the most salient aspect of a classroom's age composition is not the mean age of the group but the degree to which age is homogenous vs. heterogeneous in the classroom. In fact, studies investigating age composition typically operationalize this environmental characteristic by using either the classroom range or standard deviation of the age around the classroom mean. A recent study by Bell, Greenfield and Bulotsky-Shearer (2013) exemplifies this approach. In this study, the authors examined the relationship between classroom composition of age, operationalized as the standard deviation around each classroom mean, and child trajectories of school readiness in Head Start. They found no association between classroom age heterogeneity and children's school readiness outcomes. However, they did find that younger children in classrooms in which the standard deviation around the mean age was high had greater gains in the approaches to learning constructs compared to older children in the same classrooms and to younger children in classrooms with lower standard deviations (Bell et al., 2013).

The standard deviation is well-suited to investigations of whether children benefit from homogenous vs. heterogeneous age groupings in classrooms since it represents the degree of similarity or divergence of individual values from the mean within a classroom or other higher-level grouping of individuals. Yet, when only the standard deviation is used to characterize classroom composition, several other aspects of the classroom remain unaccounted for including the average level of the characteristic (e.g., following this example, the average age) in the classroom and whether, and if so, where, children are

clustered in the overall classroom distribution of the characteristic. Additionally, because of the way that it is calculated, the standard deviation is sensitive to the presence of children whose values for the characteristic of interest are extreme compared to the rest of their peers.

For Bell et al. (2013), and other researchers interested in the impact of classroom composition of age, prior theoretical work provided a framework for understanding and making informed choices about this trade-off. Yet, absent theoretical guidance, investigating other aspects of classroom composition, whether it be the composition of behavior problems or cognitive developmental competencies, is a pursuit best undertaken carefully. Frequently, researchers either avoid examining the role of classroom context altogether or default to mathematically simplistic methods for operationalizing classroom characteristics such as the proportion of children above a predetermined threshold or the classroom mean.

Using proportions and counts to represent classroom composition—

Proportions and counts divide the children in a classroom into two groups by comparing their value to a predetermined threshold for a characteristic of interest. A study by Dmitrieva et al. (2007) exemplifies the use of this method to describe Kindergarten classroom composition of extensive preschool experience. They investigated whether the proportion of students in a Kindergarten classroom who had either been placed in preschool before age two or who had prior extensive experience in preschool (greater than or equal to 30 h a week) predicted individual child outcomes in Kindergarten children. Their investigation also examined whether a child's experience of this context differed by whether he, himself, had extensive preschool experience or had been placed in care before age two. Dmitrieva et al. (2007) found that children who had little or no experience in child care had higher levels of externalizing behaviors in Kindergarten when in classrooms where a large proportion of their peers had extensive experience in preschool.

Counts and proportions are derived from the use of a threshold or cut-point of either substantive or statistical interest. Some instruments have an identified threshold. For example, the Center for Epidemiologic Studies-Depression Scale (CES-D) has a threshold, validated within various subgroups, that is used to distinguish between people with a high probability versus a low probability of receiving a clinical diagnosis of major depressive disorder (Radloff, 1977). In cases where an instrument does not have a validated threshold, researchers may use the value of the sample mean, 75th percentile, or 95th percentile to place students in high and low groups. This practice may be well-suited to characteristics for which prior empirical work suggests that the presence of children with values at the extremes (i.e., children with high levels of aggression) has particular salience to individual or group functioning.

When proportions or counts are used to represent classroom composition, information about the distance of each child from the threshold is not considered important for the relationship between classroom composition and the outcome of interest. All children on the same side of the threshold contribute equally to this value of classroom composition regardless of their distance from the threshold.

Using the mean to represent classroom composition—Another commonly used method to transform individual characteristics to the classroom-level is the classroom mean. In a recent paper, Thomas, Bierman, and Powers (2011) used the mean of teacher ratings of individual's level of aggression to represent classroom composition of aggression. Thomas et al. (2011) found that the average classroom composition of aggression at the beginning of first grade predicted individual levels of aggression in second grade.

When used to represent classroom composition, the classroom mean gives the value of each child's characteristic equal weight. In classrooms where the characteristic is evenly distributed across children, the mean reflects the average level of the characteristic in the classroom. Yet, in classrooms where some children have values that are high compared to the rest of their peers, these high scores inflate the value of the mean such that it does not accurately reflect of the presence of children with low values of the classroom characteristic.

Using the mean, standard deviation and skew together to represent classroom composition—One approach to transforming individual-level data into indicators of classroom composition is to fully represent the relevant characteristic's distribution (e.g., the distribution of aggression) within the classroom. For characteristics that do not have a normal distribution within a classroom, this can be done by including three pieces of information. First, including the mean, as noted above, represents the average level of the characteristic in the classroom. Second, including the standard deviation indicates how closely individual child values are clustered around the classroom mean. Third, including the skew indicates the degree to which individual child values are clustered at one end of the classroom distribution. The skew also indicates the end of the distribution in which individual values are concentrated (i.e., data with a positive skew cluster at the low end of the distribution). While no empirical work has utilized this strategy, Glewwe (1997) drew attention to the importance of accounting for distributional properties of classroom characteristics when examining the sensitivity of peer effects studies to the specification of the functional form of the relationship between classroom composition and child outcomes. By simultaneously examining the role of multiple aspects of classroom composition of a characteristic, it may be possible to tease apart (1) the role of the average level of the characteristic in the classroom from the presence of children with extreme values, (2) the role of heterogeneity of the characteristic in the classroom (i.e., the standard deviation), and (3) the role of the concentration of children in certain parts of the distribution (i.e., the skew).

Accounting for the child's relative status within the classroom composition—Using only classroom-level indicators to represent classroom composition assumes all children in the classroom experience the classroom environment in the same way. We know, however, that children make unique contributions to classroom composition and experience their classroom contexts through the lens of their own characteristics (Hanushek et al., 2003). To address this concern, Justice et al. (2011), added an individual measure that indicated the child's reference-group status relative to the classroom mean to models that examined the relationship between classroom composition of language ability at the beginning of preschool as represented by the classroom mean and individual language skills

at the end of the preschool year. A child's reference-group status is a measure of the child's skills or characteristics relative to those of his or her classroom peers (Hanushek et al., 2003). This variable was constructed by subtracting the classroom mean from each child's score such that children who scored higher than the class average received positive values and children who scored lower than the class average received negative values. Justice et al. (2011) found that children with low reference-group status (whose language skills were below the average skill level of their classroom peers) had lower language skills at the end of the preschool year when in classrooms with low average ability when compared to low reference-group status children in classrooms with high average language ability. Children with high reference-group status were not sensitive to the average classroom composition of language skills.

The present study

As the brief review presented above suggests, there has been a recent surge of interest in the role of a range of classroom characteristics in generating individual child outcomes. Efforts at operationalizing classroom composition from child-level data highlight not only the variety of transformations possible but also the need for a decision rubric that provides substantive and analytic guidance to the field. In this study, we explore the implications of four transformation methods for the association between classroom composition and child outcomes by investigating the following research question: What is the relationship between Head Start classroom composition of externalizing behavior and three Kindergarten social-emotional outcomes: internalizing behaviors, social competence, and attention/impulsivity problem behaviors?

Children's behavior in classrooms has clear consequences for their success in school. For example, research indicates that children who have difficulty regulating behavior and emotions, and who experience high levels of negative emotional arousal, have trouble concentrating in class and recalling things they have learned (Raver, Garner, & Smith-Donald, 2007). In addition, to the extent that the behavior of students in a particular classroom helps or hinders a teacher's efforts to manage the classroom effectively, develop and maintain high quality relationships with children, and deliver high-quality instruction, it may indirectly influence a host of other outcomes (Jones, Brown, & Aber, 2008). We selected externalizing behavior problems as our classroom characteristic of interest because these behaviors are hypothesized to interfere with young children's ability to learn and socialize in preschool and pre-Kindergarten settings and threaten subsequent school performance (Gilliam, 2005; Raver, 2002). Each of the three social-emotional Kindergarten outcomes investigated in this paper has particular salience for young children. They are characteristics which are rapidly developing in young children and which have associations with later school and labor market success (Campbell, Spieker, Burchinal, & Poe, 2006; Ramani, Brownell, & Campbell, 2010).

We hypothesized that each Kindergarten social-emotional outcome would be associated with one rather than all classroom-level indicators of externalizing behavior due to the differential emphasis each method of transformation places on the presence of children with extreme values of externalizing behaviors, the range of behaviors reported in the classroom,

or the clustering of children on the low end versus the high end of the externalizing behaviors scale. Specifically, we hypothesized that children's internalizing behaviors in Kindergarten would be higher for children whose Head Start classrooms have a higher average level classroom externalizing behavior. In an environment characterized by high levels of externalizing behaviors, children may be more apt to withdraw from social interactions rather than engage. We also hypothesized that children's social competence in Kindergarten would be lower when children attended a Head Start classroom with more children whose externalizing behavior was rated above the 75th percentile in the entire sample. Children's ability to practice socially competent interactions may be most curtailed in classrooms where there are more children with high rates of these typically rare—but salient—externalizing behaviors. Finally, we hypothesized that the attention/impulsivity problems of children in Kindergarten would be lower for children whose classroom distribution of externalizing behaviors was skewed toward the low end of the distribution. We posited that a concentration of children at the low end of the distribution of externalizing behaviors would indicate fewer overall disruptions to normal classroom routines.

Method

Participants

We address our research question using data drawn from the Chicago School Readiness Project (CSRP), a preschool intervention that sought to improve behavioral and school readiness outcomes in a population of Head Start students in some of Chicago's lowest resourced Head Start centers. In treatment classrooms, the intervention provided intense professional development in classroom behavior management and the weekly support of a trained mental health specialist to teachers. The study employed a cluster-randomized controlled trial (RCT) design and a pairwise matching procedure to place 18 Head Start centers (35 classrooms in total) into the intervention or control condition (with nine centers in each condition). Each of the 35 classrooms involved in the CSRP contributed data to this study. All classrooms offered full-day care and education services. At least two adults were assigned to each classroom. The average class size was 16 children, approximately half of which were male. In total, 602 students were part of the study sample in Head Start.

A Kindergarten follow-up was conducted in which data were collected from 433 children from the original sample. Children were excluded at the Kindergarten follow-up if they did not attend schools within the Chicago Public School System (for a detailed description of the study design, the intervention, and a summary of its impacts on children's developmental outcomes see Raver, Jones, Li-Grining, Metzger, Smallwood, & Sardin, 2008; Raver et al., 2009, 2011; Zhai, Raver, & Li-Grining, 2011).

The analytic sample used in this study includes 409 of the 433 children followed into Kindergarten. Children were excluded from the analytic sample if they were missing on the outcomes investigated. Table 1 presents basic demographic characteristics of the children in the analytic sample as measured in the fall of Head Start. Children in the CSRP analytic sample, on average, were four years old and roughly half were boys (50.1%). Only 21% of the children lived in homes where their mothers were married. Many CSRP children lived in families where the mother worked fewer than 10 h per week (39%) and/or did not graduate

from high school (28%). Approximately 65% of participating children were non-Hispanic Black, with the majority of the remainder being Hispanic (26%). The remaining 9% of children were reported as belonging to another race/ethnicity group or multiple race/ethnicity groups.

The CSRP data are ideal for this study because they provide data on most children in each Head Start classroom. Data are available on an average of 90% of all children in each classroom. Unlike other papers which have operationalized peer composition from data representing a small proportion of children within each classrooms (Justice et al., 2011; Mashburn, 2008), we are able to mobilize more complete classroom information to understand the classroom composition of externalizing behavior.

Measures

Outcomes—Three Kindergarten outcomes were investigated in this study: internalizing behaviors, social competence, and attention/impulsivity problems. The Kindergarten internalizing behaviors measure was constructed as a sum of scores on the 16 items included in the anxiety and depression subscale ($\alpha = 0.92$) of the child Teacher Rating Form (C-TRF; Achenbach, 1991; Achenbach, Dumeni, & Rescorla, 2002; Ivanova et al., 2007; Leung et al., 2006). The final measure had a range 0–11 with a sample mean of 1.32 ($SD = 1.91$). The Kindergarten measure of social competence was built from the Social Competence and Behavior Evaluation Scale-Short Form (SCBE-30; LaFreniere & Dumas, 1995, 1996). This measure is built from 10 items that ask teachers to rate, on a 6-point scale ($1 = \text{never}$, $6 = \text{always}$) how often each child comforts or assists another child in difficulty and related actions ($\alpha = 0.95$). The scale used in this study was constructed by adding teacher ratings on each of the 10 items to yield an indicator with a range 0–50 ($M = 30.64$, $SD = 11.96$). The Kindergarten attention/impulsivity problems measure is a teacher rating of the instability of each child's attention and cognition. This measure was constructed as a sum of scores on the 26 items included in the attention/impulsivity problems subscale of the C-TRF ($\alpha = 0.81$). In the fall of Kindergarten, teachers were asked to report on each child in their classroom using a three-point scale about how true ($0 = \text{not true}$, $2 = \text{very true}$) statements were about each child. Sample items for the attention/impulsivity problems scale include, the child (1) acts too young for his/her age, (2) hums or makes other odd noises in class, and (3) cannot concentrate, cannot pay attention for long. The final scale had a range 0–17 with a sample mean of 2.95 ($SD = 3.78$). Table 1 contains descriptive information for each of our outcomes, predictors, and covariates.

Predictors: classroom composition of externalizing behaviors—We built our measure of fall Head Start classroom composition of externalizing behaviors from the Behavior Problems Index (BPI) externalizing problem behaviors subscale ($\alpha = 0.92$). This instrument was a teacher report of individual child behaviors adapted from a 28-item parent rating scale (Zill, 1990), 18 of which comprise the externalizing behaviors subscale. It was completed in the fall of the Head Start year by teachers in each Head Start classroom. Each item is rated on a 3-point scale ($0 = \text{not true}$, $2 = \text{very/often true}$). Sample items address how true it is that the child is high strung, cheats or lies, or argues. Child scores on the individual items were added together to create a scale with a sample range 0–30 ($M = 5.77$, $SD = 5.84$).

Within the CSRP sample, 25% of the variation in teacher ratings of child externalizing behaviors exists between classrooms.

In order to address our research question, we used four methods to operationalize classroom composition from individual BPI externalizing ratings... (Please see supplemental materials for strategies and formulas for calculating composition measures used in this study.) For our first method, we calculated the classroom mean to represent the average level of externalizing behaviors in the classroom. When the classroom means were examined across the classrooms in our sample, the average classroom mean was 5.77 with a standard deviation of 3.16. For our second method, we used the standard deviation around each classroom's mean to represent the amount of heterogeneity of externalizing behaviors at the classroom level. It is important to note that we use heterogeneity to refer to the spread of teacher ratings of child externalizing behaviors rather than to the range of kinds of behaviors. The heterogeneity in our classrooms ranged from 2.65 to 6.77 ($M = 4.71$, $SD = 2.06$). For our third method, we used the classroom mean, standard deviation, and skew ($M = 1.12$, $SD = 0.59$) to fully characterize the distribution of externalizing behaviors in each classroom. Finally, for the fourth method, we used the proportion of children in each classroom that scored above the whole sample 75th percentile to represent the classroom composition of extreme externalizing behaviors. Classrooms varied widely in the proportion of children above the 75th percentile with the lowest proportion being 0.13 and the highest being 0.57 ($M = 0.35$, $SD = 0.22$).

Individual-level measure of BPI externalizing behavior—In each model, we included an individual-level variable to indicate either each child's reference-group status or each child's placement above or below the sample 75th percentile. The child-level variable representing his or her reference-group status was calculated by subtracting the classroom mean from each child's scores. Children whose scores were higher than the average level in their classroom received positive values on this variable. Children whose scores were lower than the average level in their classroom received negative values on this variable.

Covariates—A small set of classroom-, family-, and child-level covariates were included in each final model. These covariates were included because of their hypothesized role in observed heterogeneity of scores on the three outcome measures as well as prior CSRP research. The classroom-level covariates included were treatment status, Fall Head Start classroom emotional climate score as measured on the Classroom Assessment Scoring System (CLASS, Pianta & Hamre, 2009), teacher age, and a dichotomous indicator of whether the lead teacher had a bachelor's degree or not. The family-level covariates indicated whether or not the mother was married, if she worked over 10 h a week, and if the family's income was below the sample median income. Child race, age, and gender were also included as covariates. Additionally, we included individual Fall Head Start scores for the internalizing behaviors and social competence measures to account for any differences in these capacities present at the start of the study. No equivalent measure for the attention/impulsivity problems for Fall of the Head Start year was available.

Analytic strategy

Determining the relationship between classroom context and our outcomes—

After constructing our classroom composition measures, we conducted a descriptive analysis to determine what, if any, systematic relationships exist among the four different aspects of classroom composition of externalizing behaviors. To do this, we examined bivariate correlations among the four measures and created a series of scatterplots to illustrate statistically significant correlations. In addition, in preparation for our examination of the relationship between classroom composition and our three Kindergarten outcomes, we examined the distributional properties of each of the outcomes. We found that none of the three outcomes being examined were normally distributed. In fact, the internalizing behaviors and attention/impulsivity outcomes had a distribution in which child values clustered at or near zero. Because these measures were also counts of low incidence events within a given time frame and there was no reasonable expectation that children could receive a negative score, multilevel Poisson models were fit for these outcomes (de Leeuw & Meijer, 2008). The Kindergarten social competence measure had different distributional properties. While children could not receive a negative score in the measure, child values clustered at the upper end of the measure's scale suggesting that the measure was not able to distinguish between children whose true scores were high versus those who would have received higher scores had the instrument been designed differently. To account for the potential censoring that this ceiling effect suggests, multilevel Tobit models were fit for the social competence Kindergarten outcome (Maddala, 1983). Tobit regression accounts for ceiling and/or floor effects by assuming that individuals at the point of censoring would be distinguishable from one another had the instrument been more appropriately designed and a latent construct is used to make this distinction. Multilevel models were used to examine the relationship between classroom composition and each of our Kindergarten outcomes in order to account for the clustering of children in classrooms. For each outcome, we fit a series of four models, each model testing the association between the outcome of interest and one method for operationalizing classroom composition of externalizing behaviors. For all outcomes, Model 1 included the classroom mean of externalizing behaviors in fall Head Start, Model 2 included the standard deviation, Model 3 included the mean, standard deviation and skew, and Model 4 included the proportion of students in each class above the whole sample 75th percentile. All models included the child's reference status and the set of covariates described above.

Results

Descriptive analysis of different aspects of classroom composition

Table 1 shows the distribution of the different aspects of fall Head Start classroom composition of externalizing behaviors. Classrooms varied widely in their composition of externalizing behaviors but this variation is captured in different ways by the four aspects of composition examined. Analysis of the bivariate correlations between the different aspects of classroom composition revealed a strong and statistically significant relationship between the mean and standard deviation of each classroom ($r = 0.78, p < 0.001$). The scatterplot revealed a positive, linear relationship between these two aspects of classroom composition indicating that high classroom means co-occurred with heterogeneity in externalizing

behaviors. A closer look revealed that high classroom means were driven by the presence of children with values of externalizing behavior that were unusually high compared the sample mean. There were no other significant bivariate correlations.

Predicting individual outcomes using measures of classroom composition

Kindergarten internalizing behaviors—As shown in Table 2, three different aspects of classroom composition of externalizing problems were associated with children's internalizing behaviors in Kindergarten. Kindergarten internalizing behaviors had a positive, small but statistically significant relationship with classroom mean of externalizing behavior (Model 1, $\beta = 0.08$, $p = 0.02$), and standard deviation (Model 2, $\beta = 0.10$, $p = 0.05$). As shown in Model 4, a child's level of internalizing behaviors in Kindergarten was associated with the proportion of children in their Head Start classrooms who scored above the sample 75th percentile on externalizing problems ($\beta = 1.17$, $p < 0.01$), such that a child in a Head Start classroom where only 13% of the children scored above the 75th percentile (representing a classroom with one standard deviation fewer children in fall Head Start classrooms scoring above the full sample 75th percentile compared to sample average proportion of children scoring above the full sample 75th percentile) had an internalizing behaviors score 2.20 points lower ($ES = 2.65$) than a child from a Head Start classroom in which 57% of the children scored above the 75th percentile (representing a classroom with one standard deviation more children in fall Head Start classrooms scoring above the full sample 75th percentile compared to sample average proportion of children scoring above the full sample 75th percentile). Kindergarten internalizing behaviors did not have a statistically significant relationship with the child's reference status.

Kindergarten social competence—Model 1 of Table 3 illustrates the statistically significant relationship between Kindergarten social competence and the child's fall Head Start externalizing behaviors reference status. Children who had externalizing scores that were one unit higher than the mean in their classroom scored, on average, 0.47 points lower on kindergarten social competence ($p < 0.01$, $ES = 0.04$) compared to peers who scored at the classroom mean. This relationship is robust to the inclusion of different classroom compositional measures in Models 2 and 3. Kindergarten social competence did not have a statistically significant relationship to the classroom mean or standard deviation when these operationalization methods were added independently in Models 1 and 2.

As shown in Model 3 of Table 3, Kindergarten social competence is associated with the overall distribution of fall Head Start classroom composition of externalizing behaviors (column 3) such that for every one unit increase in the mean level of classroom externalizing behaviors in Head Start a child scored 1.03 points higher in Kindergarten social competence ($p = 0.04$) but only if the standard deviation and skew of the classroom composition were held constant. Interestingly, when children were in Head Start classrooms where the heterogeneity of externalizing behavior was high (as represented by classroom standard deviations one unit higher than the average standard deviation across classrooms) children scored 1.28 points lower ($p = 0.03$) than children in classrooms where the characteristic was more homogeneously distributed among their peers. The positive and statistically significant association of Kindergarten social competence to the classroom skew ($\beta = 2.94$, $p = 0.04$)

indicates that children had higher Kindergarten social competence scores when they had been in Head Start classrooms where the scores of the majority of the children clustered near the low end of the scale (indicating low overall externalizing behaviors). Similarly, as Model 4 shows, Kindergarten social competence was not predicted by the classroom proportion of students above the whole sample 75th percentile.

Kindergarten attention/impulsivity problems—As shown in Models 1–4 of Table 4, Kindergarten attention/impulsivity problems was predicted by child reference status ($\beta = 0.05$, $p < 0.001$) and whether or not the child was above the whole sample 75th percentile of fall Head Start externalizing behaviors ($\beta = 2.94$, $p < 0.001$). Children who were rated one unit higher on fall Head Start externalizing behaviors than the classroom mean scored 1.05 points higher on kindergarten attention/impulsivity problems ($ES = 0.28$). However, if a child's externalizing behaviors were above the sample 75th percentile, they scored 1.72 points ($ES = 0.45$) higher than peers whose rating fell below this cut point. This outcome was not associated with any aspect of fall Head Start classroom composition of externalizing behaviors investigated in this study.

Discussion

In recent years, the relationship between child outcomes and the classroom composition of a range of characteristics has been studied using one of three methods to transform individual-level data to the classroom level. These methods are the classroom mean, the standard deviation around the mean and the proportion of children above or below a pre-specified threshold. This study sought to better understand the implications of different methods of transforming individual-level data into classroom-level indicators of externalizing behavior within a single sample to predict three different Kindergarten outcomes.

Before conducting this study, we hypothesized that children's internalizing behaviors in Kindergarten would be higher if the average level of externalizing behaviors in their Head Start classroom was high. Indeed, a positive and statistically significant relationship was found between the classroom mean of externalizing behavior and Kindergarten internalizing behaviors. Positive, statistically significant relationships were also found between Kindergarten internalizing behavior problems and (1) the standard deviation around the classroom mean of externalizing behaviors in Head Start as well as (2) the proportion of children in the Head Start classroom whose externalizing behaviors were rated above the sample 75th percentile. Importantly, the average level of externalizing behaviors in a child's Head Start classroom did not have a statistically significant relationship with Kindergarten internalizing behavior when the standard deviation and skew of the distribution were held constant. That is, if you consider two Head Start classrooms in which individual child values of externalizing behaviors were equally distributed around the mean (equivalent standard deviations) and in which the distribution was perfectly symmetrical (i.e., skew = 0) differences in the mean value of externalizing behaviors in those Head Start classrooms would not be associated with Kindergarten internalizing behaviors. An average child from each of these classrooms would be expected to have equivalent Kindergarten internalizing behaviors even if the classroom means differed. Taken together and combined with our knowledge that there was a positive, strong and statistically significant correlation between

classroom mean of externalizing behaviors and the standard deviation around the mean (classrooms with higher means also were composed of children whose individual values were, on average, farther away from the mean) suggests that the presence of children with extreme values of externalizing behaviors in Head Start may be driving this relationship. It may be that in classrooms in which children's externalizing behaviors are not uniformly distributed and in which there are children with high levels of externalizing behaviors, teachers' attention may be too diverted to ensure that quiet and withdrawn children have ample opportunities to appropriately interact with peers.

We also hypothesized that Kindergarten social competence would be lower if children experienced Head Start classroom contexts with a high proportion of children with externalizing behaviors above the 75th percentile of this characteristic in Head Start. While we did not find that the presence of individuals with extreme values of externalizing behavior—absent other characteristics of the composition of externalizing behaviors—related to lower values of social competence, we did find evidence that the saturation of externalizing behaviors in the Head Start classroom context related to Kindergarten social competence. Kindergarten social competence had statistically significant relationships with the classroom mean, standard deviation around the mean, and skew of externalizing behaviors in Head Start when they were entered into the same model. The inclusion of these variables in the same model was meant to more fully characterize the distributional qualities of externalizing behaviors in the Head Start classrooms. Let us return to the example in the paragraph above to interpret these findings.

When two average children from different Head Start classrooms with the same standard deviation of externalizing behaviors and the same symmetry of distribution of individual ratings around the classroom mean are compared in Kindergarten, the child who had been in the Head Start classroom with the higher average level of externalizing behaviors is predicted to have higher social competence. Our interpretation of this finding hinges on unpacking the relationships between this outcome and the standard deviation and skew of externalizing behaviors. Our interpretation of this finding is also refined when we remember that the behaviors captured by the externalizing behaviors sub-scale of the BPI (e.g., Bullies or is cruel/mean to others) are low incidence behaviors; therefore it is common to have many children in a classroom whose teacher-ratings are low on this construct. We would expect that children who experienced Head Start classrooms in which the children's values clustered near the classroom mean (small standard deviation) and were clustered near the low end of the distribution of externalizing behaviors (positive skew) would have, on average, higher ratings of social competence. If these conditions were met, the only way that the mean level of externalizing behaviors in Head Start would be high would be if there were just a few children with extreme values of externalizing behaviors pulling the mean upward. One possible way a child's social competence in Kindergarten may be positively influenced in a classroom such as this is through positive modeling by the teacher of responses to and redirection of externalizing behaviors. However, it is important to note that children in these classrooms with externalizing scores above the classroom mean have lower social competence scores in Kindergarten. Therefore, it might be that the teacher may be modeling positive behaviors for the majority of the children in the classroom but in a way that is not

effective for children with higher-than-average externalizing behaviors. This interpretation ought to be explored in future work.

Our final hypothesis stated that children in Head Start classrooms whose classroom composition of externalizing behaviors was skewed toward the low end of the distribution would have fewer attention and impulsivity problems in Kindergarten. In our analysis we found no statistically significant relationship between Kindergarten attention/impulsivity problems and classroom-level variables describing the Head Start composition of externalizing behaviors. Yet, we did find that an individual child's Head Start rating relative to the classroom mean of externalizing behaviors did have a positive, statistically significant relationship with Kindergarten attention/impulsivity problems. Children who had higher externalizing behavior problems than the average level of externalizing problems in their Head Start classroom were reported to have higher attention/impulsivity problems in Kindergarten. This finding mirrors part of the Kindergarten social competence findings in that children with higher-than-average externalizing behaviors in their Head Start classroom have poorer behavioral outcomes in Kindergarten.

While non-significant findings are rarely reported, our null findings also have important implications in that they reveal ways in which classroom composition of externalizing behaviors may not matter for classroom functioning and subsequently for later child outcomes. For example, the mean level of Head Start externalizing behavior alone is not associated with Kindergarten social competence or attention/impulsivity problems. This may indicate either that the average level of externalizing behaviors is not representative of a consistent and salient feature of classroom contexts for individual student functioning or that the impact of the average level of externalizing behavior operates through the unique manner in which each student perceives that milieu. Support for the latter can be seen in the robust association between a child's relative status in Head Start externalizing behaviors and Kindergarten social competence and attention/impulsivity problems. Neither the proportion nor overall number of children scoring at or higher than the sample 75th percentile in externalizing behaviors were associated with Kindergarten social competence and attention/impulsivity problems. This may indicate that just being around children who are perceived to be high in externalizing behavior may not be enough to influence these domains of functioning.

In preschool classroom contexts, children may be directly or indirectly influenced by the classroom composition of externalizing behaviors (Bierman, 2011). Our operationalization techniques suggest that Kindergarten outcomes that are associated with child relative status of externalizing behaviors in Head Start may be directly influenced by a child's participation in a class in which her own level of externalizing behaviors is meaningful in comparison to the average classroom level. Possibly, children who have high levels of externalizing behavior compared to the class average have fewer or lower quality interactions with peers regardless of the absolute level of their problem behaviors. However, because classroom composition of externalizing behavior was derived from a teacher-rating of this behavior, it may also be that children whose individual externalizing behaviors are high relative to the classroom average are perceived as troublemakers by a teacher even when their absolute level of externalizing behaviors may be quite low. Future work needs to be done to

disentangle these influences including replication studies with observations of these behaviors. Additionally, these assumptions about externalizing behavior cannot be said to extend to other classroom composition characteristics and they also must be tested in other data to ensure that the null findings are not specific to the CSRP sample. Relatedly, since our compositional characteristic was derived from teacher-ratings of individual children's externalizing behaviors, the ratings may have been influenced in some way by the children with whom the child was in class. Certainly, teacher-ratings of child behavior, while sophisticated and often based upon systematic knowledge of child knowledge, may be biased by the full set of children with whom the teacher has worked throughout her career (Ladd & Profilet, 1996).

Taken together, the findings of this analysis indicate that there are important theoretical implications to consider when selecting how to transform child ratings or characteristics to the classroom level. This study highlights four challenges that ought to be addressed. First, theory and substantive concerns must inform responsible hypothesis formation about the function of the characteristic, behavior, or perception at the classroom-level. Once this hypothesis has been determined, one must consider three methodological issues related to the transformation. First, what kind of summary information about the collection of individual ratings is most relevant given the hypothesized function of the phenomenon at the classroom-level? Second, does the placement of the individual in relationship to the overall group matter given the hypothesis and the anticipated relationship between the classroom-level construct and the individual outcome? And finally, given the restrictions set by addressing the first two methodological challenges, what is the most parsimonious mathematical transformation of the individual-level data that is consistent with the underlying theory and enables the investigator to address the substantive questions of interest?

This study used Head Start teacher ratings of externalizing behavior to represent the classroom composition of externalizing behavior. Importantly, our outcomes of interest were three important social-emotional skills derived from a different rater—the child's Kindergarten teacher. This, along with the availability of data on an average of 90% of all children in each Head Start classroom, is a strength of the analytic approach used in this study. However, there are several avenues for improvement and future work. In particular, the threat to validity posed by the use of statistics (i.e., mean, standard deviation, and skew) as predictors in a regression model must be explored in detail. Using statistics as independent variables in a regression model violates the assumption that all independent variables are measured without error. Further, experiments in which children are randomly assigned to classrooms would reduce the degree to which teacher skill is confounded with the characteristics and skills of the children assigned to her classroom.

As interest in the variability of impact of early education grows, so does interest in the role that classroom composition plays in this variability. By seeking consensus in our understanding of how to best transform individual-level data to classroom-level characteristics, we seek to refine our tools for understanding variability. This paper summarizes and extends our tools for understanding variability due to classroom composition.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ecresq.2014.07.007>.

References

- Achenbach, TM. Integrative guide for the 1991 CBCL/4-18, YSR, and TRF profiles. Burlington, VT: University of Vermont, Department of Psychiatry; 1991.
- Achenbach TM, Dumenci L, Rescorla LA. Ten year comparisons of problems and competencies for national samples of youth: Self, parent, and teacher reports. *Journal of Emotional and Behavioral Disorders*. 2002; 10(4):194–203. <http://dx.doi.org/10.1177/10634266020100040101>.
- Bell ER, Greenfield DB, Bulotsky-Shearer RJ. Classroom age composition and rates of change in school readiness for children enrolled in Head Start. *Early Childhood Research Quarterly*. 2013; 28(1):1–10. <http://dx.doi.org/10.1016/j.ecresq.2012.06.002>.
- Bierman KL. The promise and potential of studying the “invisible hand” of teacher influence on peer relations and student outcomes: A commentary. *Journal of Applied Developmental Psychology*. 2011; 32:297–303. <http://dx.doi.org/10.1016/j.appdev.2011.04.004>.
- Bonesrønning H. Peer group effects in education production: Is it about congestion? *Journal of Socio-Economics*. 2008; 37(1):328–342. <http://dx.doi.org/10.1016/j.socsec.2006.12.044>.
- Bronfenbrenner, U. *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press; 1979.
- Bronfenbrenner, U., Morris, P. The bioecological model of human development. In: Lerner, RM., editor. *Handbook of child psychology*. 6th. Vol. 1. New York, NY: Wiley; 2006. p. 793–828. <http://dx.doi.org/10.1002/9780470147658.chpsy0114>
- Campbell SB, Spieker S, Burchinal M, Poe MD. Trajectories of aggression from toddlerhood to age 9 predict academic and social functioning through age 12. *Journal of Child Psychology and Psychiatry*. 2006; 47(8):791–800. <http://dx.doi.org/10.1111/j.1469-7610.2006.01636.x>. [PubMed: 16898993]
- de Leeuw, J., Meijer, E. *Handbook of multilevel analysis*. New York, NY: Springer Press; 2008. <http://dx.doi.org/10.1007/978-0-387-73186-5>
- DeRosier ME, Cillessen AHN, Coie JD, Dodge K. Group social context and children’s aggressive behavior. *Child Development*. 1994; 65(4):1068–1079. <http://dx.doi.org/10.2307/1131305>. [PubMed: 7956465]
- Dmitrieva J, Steinberg L, Belsky J. Child-care history, classroom composition, and children’s functioning in kindergarten. *Psychological Science*. 2007; 18(12):1032–1039. <http://dx.doi.org/10.1111/j.1467-9280.2007.02021.x>. [PubMed: 18031408]
- Gilliam, WS. *Prekindergartners left behind: Expulsion rates in state prekindergarten systems*. New Haven, CT: Yale Child Studies Center; 2005.
- Glewwe P. Estimating the impact of peer group effects on socioeconomic outcomes: Does the distribution of peer group characteristics matter? *Economics of Education Review*. 1997; 16(1): 39–43. [http://dx.doi.org/10.1016/S0272-7757\(96\)00017-9](http://dx.doi.org/10.1016/S0272-7757(96)00017-9).
- Hanushek EA, Kain JF, Markman JM, Rivkin SG. Does peer ability affect student achievement? *Journal of Applied Econometrics*. 2003; 18(5):527–544. <http://dx.doi.org/10.1002/jae.741>.
- Ivanova MY, Achenbach TM, Rescorla LA, Dumenci L, Almqvist F, Bathiche M, et al. Testing the Teacher’s Report Form syndromes in 20 societies. *School Psychology Review*. 2007; 36(3):468–483.

- Jones, SM., Brown, JL., Aber, JL. Classroom settings as targets of intervention and research. In: Shinn, M., Yoshikawa, H., editors. *The power of social settings: Transforming schools and community organizations to enhance youth development*. New York, NY: Oxford; 2008. p. 58-77.
- Justice LM, Petscher Y, Schatschneider C, Mashburn A. Peer effects in preschool classrooms: Is children's language growth associated with their classmates' skills? *Child Development*. 2011; 82(6):1768–1777. <http://dx.doi.org/10.1111/j.1467-8624.2011.01665.x>. [PubMed: 22026480]
- Ladd GW, Profilet SM. The child behavior scale: A teacher-report measure of young children's aggressive, withdrawn, and prosocial behaviors. *Developmental Psychology*. 1996; 32(6):1008–1024. <http://dx.doi.org/10.1037/0012-1649.32.6.1008>.
- Lavy V, Schlosser A. Mechanisms and impacts of gender peer effects at school. *American Economic Journal*. 2011; 3:1–33. <http://dx.doi.org/10.3386/w13292>. [PubMed: 22199993]
- LaFreniere, PJ., Dumas, JE. *Social Competence and Behavior Evaluation: Preschool edition (SCBE)*. Los Angeles, CA: Western Psychological Services; 1995.
- LaFreniere PJ, Dumas JE. Social competence and behavior evaluation in children ages 3 to 6 years: The short form (SCBE-30). *Psychological Assessment*. 1996; 8(4):369–377.
- Leung PW, Kwong SL, Tang CP, Ho TP, Hung SF, Lee CC, et al. Test–retest reliability and criterion validity of the Chinese version of CBCL, TRF, and YSR. *Journal of Child Psychology and Psychiatry*. 2006; 47(9):970–973. <http://dx.doi.org/10.1111/j.1469-7610.2005.01570.x>. [PubMed: 16930392]
- Maddala, GS. *Limited-dependent and qualitative variables in econometrics*. Cambridge, MA: Cambridge University Press; 1983.
- Mashburn AJ. Quality of social and physical environments in and children's development of academic, language, and literacy skills. *Applied Developmental Science*. 2008; 12(3):113–127. <http://dx.doi.org/10.1080/10888690802199392>.
- Mashburn AJ, Justice LM, Downer JT, Pianta RC. Peer effects on children's language achievement during pre-kindergarten. *Child Development*. 2009; 80(3):686–702. <http://dx.doi.org/10.1111/j.1467-8624.2009.01291.x>. [PubMed: 19489897]
- Moller AC, Forbes-Jones E, Hightower AD, Friedman R. The developmental influence of sex composition in preschool classrooms: Boys fare worse in preschool classrooms with more boys. *Early Childhood Research Quarterly*. 2008; 23(3):409–418. <http://dx.doi.org/10.1037/a0013099>.
- Neidell, M., Waldfogel, J. Cognitive and non-cognitive peer effects in early education. Cambridge, MA: NBER Working Paper Series; 2008. <http://dx.doi.org/10.3386/w14277>
- Piaget, J. Piaget's theory. In: Mussen, PH., editor. *Handbook of child psychology*. 4th. Vol. I. Vol. I. Wiley; New York, NY: 1983. p. 294-356.(Original work published in 1970)
- Pianta RC, Hamre BK. Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*. 2009; 38:109–119. <http://dx.doi.org/10.3102/0013189x09332374>.
- Radloff LS. The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*. 1977; 1:385–401. <http://dx.doi.org/10.1177/014662167700100306>.
- Ramani GB, Brownell CA, Campbell SB. Positive and negative peer interaction in 3- and 4-year-olds in relation to regulation and dysregulation. *Journal of Genetic Psychology*. 2010; 171(3):218–250. <http://dx.doi.org/10.1080/00221320903300353>. [PubMed: 20836431]
- Raver CC. Emotions matter: Making the case for the role of young children's emotional development for early school readiness. *Social Policy Report*. 2002; XVI(3):1–30.
- Raver, CC., Garner, P., Smith-Donald, R. The roles of emotion regulation and emotion knowledge for children's academic readiness: Are the links causal?. In: Pianta, RC.Cox, MJ., Snow, KL., editors. *School readiness and the transition to kindergarten in the era of accountability*. Baltimore, MD: Paul H. Brookes; 2007. p. 121-147.
- Raver CC, Jones SM, Li-Grining CP, Metzger M, Smallwood K, Sardin L. Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in head start settings. *Early Childhood Research Quarterly*. 2008; 63(3):253–255. <http://dx.doi.org/10.1016/j.ecresq.2007.09.001>. [PubMed: 18364994]

- Raver CC, Jones SM, Li-Grining C, Zhai F, Metzger MW, Solomon B. Targeting children's behavior problems in preschool classrooms: A cluster-randomized controlled trial. *Journal of Consulting and Clinical Psychology*. 2009; 77(2):302–316. <http://dx.doi.org/10.1037/a0015302>. [PubMed: 19309189]
- Raver CC, Jones SM, Li-Grining C, Zhai F, Bub K, Pressler E. CSRP's impact on low-income preschoolers' preacademic skills: Self-regulation as a mediating mechanism. *Child Development*. 2011; 82(1):362–378. <http://dx.doi.org/10.1111/j.1467-8624.2010.01561.x>. [PubMed: 21291447]
- Sameroff A. A unified theory of development: A dialectic integration of nature and nurture. *Child Development*. 2010; 81(1):6–22. <http://dx.doi.org/10.1111/j.1467-8624.2009.01378.x>. [PubMed: 20331651]
- Thomas DE, Bierman KL, Powers CJ. The influence of classroom aggression and classroom climate on aggressive-disruptive behavior. *Child Development*. 2011; 82(3):751–757. <http://dx.doi.org/10.1111/j.1467-8624.2011.01586.x>. [PubMed: 21434887]
- Vygotsky, L.S. *The collected works of L S Vygotsky: Volume 1 Problems of General Psychology*. Reiber, RW., Carton, AS., editors. New York, NY: Plenum Press; 1987.
- Zhai F, Raver CC, Li-Grining C. Classroom-based interventions and teachers' perceived job stressors and confidence: Evidence from a randomized trial in Head Start settings. *Early Childhood Research Quarterly*. 2011; 26:442–452. <http://dx.doi.org/10.1016/j.ecresq.2011.03.003>. [PubMed: 21927538]
- Zill, N. *Behavior problem index based on parent report: National health interview survey–child health supplement*. Washington, DC: National Center for Health Statistics; 1990.

Table 1Descriptive statistics for the kindergarten Chicago School Readiness Project sample ($N_{\text{children}} = 409$).

	<i>M</i>	<i>SD</i>	Range
<i>Head start classroom covariates</i>			
CLASS: Emotional climate score	16.03	2.69	9.0–20.2
Lead teacher has a BA/BS ^a	0.63	0.48	0.0–1.0
Lead teacher age	40.36	11.31	22.0–65.0
Treatment ^a	0.51	0.50	0.0–1.0
Average number of children in classroom	16.55	2.64	9.0–20.0
<i>Head start family covariates</i>			
Mother is married ^a	0.21	0.41	0.0–1.0
Mother did not graduate from high school ^a	0.28	0.44	0.0–1.0
Mother works fewer than 10 h ^a	0.39	0.49	0.0–1.0
Family has less than sample median income ^a	0.35	0.46	0.0–1.0
<i>Head start child covariates</i>			
Child is male ^a	0.50	0.50	0.0–1.0
Child age	48.84	7.43	25.8–72.9
Child is non-Hispanic African American ^a	0.65	0.48	0.0–1.0
<i>Head start fall</i>			
Social competence	25.84	8.18	0.0–50.0
Internalizing behaviors	1.15	1.32	0.0–6.5
Individual BPI externalizing behaviors	5.77	5.84	0.0–30.0
<i>Head start fall classroom composition</i>			
Classroom mean BPI externalizing behaviors	5.77	3.16	0.3–13.5
Classroom standard deviation of BPI externalizing behaviors	4.71	2.06	0.5–10.8
Classroom skew of BPI externalizing behaviors	1.12	0.59	–0.09–2.7
Proportion of children in classroom scoring above sample 75th percentile	0.35	0.22	0.0–0.8
<i>Kindergarten outcomes</i>			
Social competence	30.64	11.96	0.0–50.0
Internalizing behaviors	1.32	1.91	0.0–11.0
Attention and impulsivity problems	2.95	3.78	0.0–17.0

^aThe mean value of dichotomous variables indicates proportion of sample with a value of 1.

Table 2

Multilevel Poisson modeling results for final kindergarten internalizing behaviors models ($N_{children} = 409$, $N_{classrooms} = 35$).

	1 β (SE)	2 β (SE)	3 β (SE)	4 β (SE)
<i>Intercept</i>	-1.88* (0.89)	-1.65 (0.90)	-2.76** (1.07)	-1.27*** (0.39)
Fall HS internalizing behaviors	-0.02 (0.05)	-0.01 (0.04)	-0.02 (0.05)	0.01 (0.04)
<i>Child-level fall HS BPI externalizing behaviors</i>				
Child reference status	0.02 (0.01)	0.01 (0.01)	0.02 (0.01)	
Child above 75th percentile				0.04 (0.13)
<i>Fall HS classroom composition of BPI externalizing behaviors</i>				
Classroom mean	0.08* (0.03)		0.10 (0.06)	
Standard deviation		0.10* (0.04)	-0.002 (0.07)	
Skew			0.23 (0.16)	
Proportion of students above 75th percentile				1.17*** (0.39)
Model fit (-2LL)	-730.47	-731.31	-729.39	-729.33

Notes: HS = Head Start; -2LL model fit statistics can be compared to one another to determine whether one model represents an improvement over another model. The difference between the -2LL of two models being compared is considered a likelihood ratio. This likelihood ratio can be compared to a χ^2 distribution with degrees of freedom equal to the difference in number of parameters in the models.

* $p < 0.05$.

** $p < 0.01$.

*** $p < 0.001$.

Table 3

Multilevel Tobit modeling results for final kindergarten social competence models ($N_{children} = 409$, $N_{classrooms} = 35$).

	1 β (SE)	2 β (SE)	3 β (SE)	4 β (SE)
<i>Intercept</i>	36.14 *** (8.87)	40.21 *** (8.51)	26.22 ** (10.02)	40.04 *** (9.14)
Fall HS social competence	0.27 ** (0.10)	0.27 ** (0.10)	0.29 ** (0.09)	0.38 *** (0.09)
<i>Child-level fall HS BPI externalizing behaviors</i>				
Child reference status	-0.47 ** (0.16)	-0.47 ** (0.16)	-0.42 ** (0.16)	
Child above 75th percentile				-1.92 (1.70)
<i>Fall HS classroom composition of BPI externalizing behaviors</i>				
Classroom mean	-0.06 (0.28)		1.03 * (0.49)	
Standard deviation		-0.41 (0.38)	-1.28 * (0.60)	
Skew			2.94 * (1.41)	
Proportion of students above 75th percentile				-1.52 (4.20)
Model fit (-2LL)	-1541.2	-1540.6	-1537.9	-1544.7

Notes: HS = Head Start; -2LL model fit statistics can be compared to one another to determine whether one model represents an improvement over another model.

The difference between the -2LL of two models being compared is considered a likelihood ratio. This likelihood ratio can be compared to a χ^2 distribution with degrees of freedom equal to the difference in number of parameters in the models.

* $p < 0.05$.

** $p < 0.01$.

*** $p < 0.001$.

Table 4

Multilevel Poisson modeling results for final kindergarten attention/impulsivity problems models ($N_{children} = 409$, $N_{classrooms} = 35$).

	1 β (SE)	2 β (SE)	3 β (SE)	4 β (SE)
<i>Intercept</i>	-0.27 (0.75)	-0.38 (0.75)	-0.07 (0.90)	-0.42 (0.77)
<i>Child-level fall HS BPI externalizing behaviors</i>				
Child reference status	0.05 *** (0.01)	0.05 *** (0.01)	0.05 *** (0.01)	
Child above 75th percentile				0.54 *** (0.08)
<i>Fall HS classroom composition of BPI externalizing behaviors</i>				
Classroom mean	0.03 (0.03)		-0.01 (0.05)	
Standard deviation		0.05 (0.04)	0.05 (0.06)	
Skew			-0.10 (0.14)	
Proportion of students above 75th percentile				-0.05 (0.39)
Model fit (-2LL)	-1139.5	-1137.2	-1137.0	-1155.1

Notes: HS = Head Start; -2LL model fit statistics can be compared to one another to determine whether one model represents an improvement over another model. The difference between the -2LL of two models being compared is considered a likelihood ratio. This likelihood ratio can be compared to a χ^2 distribution with degrees of freedom equal to the difference in number of parameters in the models.

*
 $p < 0.05$.

**
 $p < 0.01$.

 $p < 0.001$.