SCHOOL PSYCHOLOGY REVIEW
2020, VOL. 49, NO. 1, 47–59
https://doi.org/10.1080/2372966X.2020.1717372

Forecasting Accuracy of Earliest Assessment Versus Transitional Change in Early Education Classroom Problem Behavior Among Children at Risk

Roland S. Reyes[a], Paul A. McDermott[a], Marley W. Watkins[b], Michael J. Rovine[a], and Jessica L. Chao[a]

[a]University of Pennsylvania; [b]Baylor University

ABSTRACT
This study compared the relative contribution of earliest assessment of preschool children's context-specific problem behaviors with subsequent observations of those behaviors for the prediction of later academic and sociobehavioral performance in first grade. Using a nationally representative sample of low-income children from the Head Start Impact Study ($N = 3{,}827$), children's problem behaviors in 22 classroom situational contexts were assessed annually through 2 years of prekindergarten, kindergarten, and first grade. Results from a two-stage analytical approach support the use of earliest assessment as a suitable strategy for the identification and intervention of children's classroom problem behaviors, where subsequent observations did not increase predictive accuracy over earliest assessment alone. Implications are discussed for assessment theory and practice.

ARTICLE HISTORY
Received December 5, 2018
Accepted October 14, 2019

KEYWORDS
assessment, social–emotional, hierarchical linear modeling, preschool, poverty

ASSOCIATE EDITOR
Timothy Curby

Young children with emergent problem behaviors face increased risk for academic and social–emotional difficulty. Research shows that classroom behavior problems can lead to disruptions in learning and preclude the formation of peer and teacher relationships, all of which have negative implications for later schooling (Campbell, 2002; Raver & Knitzer, 2002; Shonkoff & Phillips, 2000). The issue is especially problematic for children from underresourced families, who evidence suggests tend to have a higher incidence of behavior problems (Qi & Kaiser, 2003; Reiss, 2013) and face high risk for continued academic and social–emotional difficulty (U.S. Department of Health and Human Services [USDHHS], 2010b; Isenberg et al., 2016).

In response, research and policy have argued the importance of early intervention as a strategy for supporting children in need (e.g., Campbell et al., 2016; National Research Council & Institute of Medicine, 2009; Poulou, 2015). However, identifying children at risk for later difficulty can be challenging. First, substantial variation in early childhood development makes it difficult to determine whether early problem behavior will be transient or remain stable over time (Campbell, 2002; National Research Council, 2008; Shonkoff & Phillips, 2000). As a result, earliest assessments of problem behavior administered at a given point in time, while providing an initial indication of future problems, may only capture transitional difficulties unlikely to persist (Campbell et al., 2016; Poulou, 2015). Subsequent observations of the same behavior may augment earliest assessments by informing transitional change over time; however, implementation of such an approach requires increased resources (in terms of time, money, or both) not afforded to all schools and early childcare programs (Epstein, Schweinhart, DeBruin-Parecki, & Robin, 2004; National Research Council, 2008).

Second, though research suggests that individual behavior manifests differently across situations (Mischel, 2004), the most popular teacher rating scales designed to assess social–emotional functioning neglect the classroom-specific situations wherein problem behaviors emerge (e.g., peer versus teacher interactions; McDermott, Watkins, Rovine, & Rikoon, 2014). This shortcoming can limit the predictive validity of the instrument (Neugebauer, 2014; Shaffer & Postlethwaithe, 2012) and make it difficult to determine whether problem behavior generalizes across different classroom contexts or is isolated to specific situations (McDermott, Steinberg, & Angelo, 2005).

In this article, we aim to contrast the forecasting accuracy of earliest assessments versus transitional change approaches for the identification of children at risk for later academic and social–emotional difficulty. We accomplish this using the national sample of children from the Head Start Impact Study (HSIS; USDHHS, 2010a), a longitudinal randomized controlled trial that was designed to assess the impact of Head Start programs on children from underresourced families. The focus of our study

CONTACT Roland S. Reyes roreyes@upenn.edu Graduate School of Education, University of Pennsylvania, 3700 Walnut Street, Philadelphia, PA 19104-6216, USA.

 DOI: 10.1080/2372966X.2020.1717372

sharpens by examining dimensions of problem behaviors that are defined by the classroom situations wherein those behaviors occur. Given the longitudinal nature of the HSIS, we are able to model the trajectory of children's contextually based problem behaviors, where earliest assessments represent the child's initial status in a given behavior problem and transitional change his or her rate and direction of change in that same behavior over time (McDermott, Rovine, Buek, et al., 2018).

Importance of Situational Context

Situations can influence individual behavior (Mischel, 2004). For example, children in the classroom might behave differently in situations involving the teacher than in settings involving their peers. Research suggests that individual behavior is partly a function of the situation in which the behavior is embedded and the individual's disposition (Mischel, Shoda, & Mendoza-Denton, 2002). Thus, a child with a given level of aggression is likely to display varying levels of behavioral disturbance across classroom situations. Moreover, these patterns of behavior can be relatively stable, suggesting meaningful intraindividual variation in one's behavior (Mischel, 2004).

In the classroom, research highlights the importance of peer, teacher, and learning situations because the interactions embedded within these situations can shape the child's academic and social–emotional development (Downer, Booren, Lima, Luckner, & Pianta, 2010; Hamre & Pianta, 2010). Peer situations include free play or resolving conflict with other children (Bulotsky-Shearer, Fantuzzo, & McDermott, 2008; Chen, Fein, Killen, & Tam, 2001; McDermott et al., 2014), whereas teacher interactions may include the manner in which children speak to the teacher and seek the teacher's help (Bulotsky-Shearer et al., 2008; McDermott et al., 2014). Learning situations such as participating in group activities or learning a new task have also been linked to children's school-related outcomes (McClelland & Morrison, 2003; Fantuzzo, Perry, & McDermott, 2004). Positive interactions in these situations are associated with increased academic achievement, social–emotional proficiency, and, with respect to teacher–child relationships, potential protective effects against later negative outcomes (e.g., Baker, Grant, & Morlock, 2008; Buhs & Ladd, 2001; Fantuzzo et al., 2004; McClelland & Morrison, 2003; O'Connor, Dearing, & Collins, 2011; Pianta & Stuhlman, 2004; Spilt, Hughes, Wu, & Kwok, 2012). Conversely, adjustment difficulties in these classroom situations tend to forecast later difficulty (Buhs & Ladd, 2001; Hamre & Pianta, 2001; McDermott, Rovine, Reyes, et al., 2018).

Despite their importance in children's academic development, the most common teacher rating scales used to assess social–emotional functioning neglect the situational contexts wherein problem behaviors occur (McDermott et al., 2005). Thus, situational variation is treated as measurement error and problem behavior scores are effectively averaged across classroom situations (Mischel, 2004; Mischel et al., 2002). Though somewhat informative, this practice has the potential of leading to less accurate inferences of later performance (Mischel et al., 2002). Furthermore, the loss of such information can make it difficult for practitioners to identify the likely motivations behind problem behavior (Mischel, 2004; Zayas, Whitsett, Lee, Wilson, & Shoda, 2008), where knowledge of those situations might assist in the design of appropriate treatment.

Alternatively, contextually based instruments capture, along with the prevalence of problem behavior, the contexts in which those behaviors emerge. This provides a clearer picture of a child's social–emotional functioning across classroom situations that can lead to increased forecasting accuracy of later outcomes (Mischel et al., 2002). For example, research shows that later academic performance has a stronger correlation with disturbances in learning situations than with problems emerging in teacher–child or peer–peer interactions (McDermott et al., 2014). Similarly, Bulotsky-Shearer, Bell, and Domínguez (2012) found that different latent profiles of preschool problem behavior, each including varying manifestations of internalizing, externalizing, and situational problem behaviors, differentially predicted later outcomes. Thus, when compared to global measures of classroom functioning, contextually based instruments should lead to more accurate prediction of later performance (Neugebauer, 2014; Shaffer & Postlethwaithe, 2012) while also capturing meaningful variation in behavior (Mischel, 2004).

One contextually based measure of classroom problem behavior is the Adjustment Scales for Early Transition in Schooling (ASETS; McDermott, Watkins, Rovine, & Rikoon, 2013, McDermott et al., 2014). The ASETS measures the severity of classroom behavioral disturbance based on its prevalence across a variety of classroom situations involving peers, teachers, and learning activities. The ASETS was selected by the USDHHS (2010a) to measure children's social–emotional functioning in the aforementioned HSIS, a longitudinal study where children were assessed annually through 2 years of prekindergarten, kindergarten, and first grade. Analyses of the ASETS's dimensional structure (McDermott et al., 2013, 2014) show that children's problem behaviors are measured in two ways: through four phenotypes, dimensions of problem behavior defined by their similarity in behavioral expression (Aggression, Attention Seeking, Low Energy, and Reticence/Withdrawal), and three situtypes, problem behavior dimensions defined by the similarity in the situations wherein problems emerge (Peer Context Problems, Teacher Context Problems, and Learning Context Problems). Studies show that all dimensions maintain factorial integrity across the early transition period and are significantly associated with later related outcomes (McDermott et al., 2013, 2014).

CHALLENGES IN EARLY CHILDHOOD IDENTIFICATION

Although research argues the importance of early intervention (National Research Council & Institute of Medicine, 2009; Poulou, 2015), the identification of preschool children in need can be challenging (Shonkoff & Phillips, 2000). This is due in part to substantial variability in the development of young children's problem behaviors, where research indicates varying initial levels and subsequent rates of change in those behaviors across the early childhood period (e.g., Campbell, Spieker, Burchinal, Poe, & The NICHD Early Child Care Network, 2006; McDermott et al., 2019; Wildeboer et al., 2015). Whereas most children demonstrate a decline in behavior problems over time, a small proportion tend to manifest stable or increasing behavioral disturbance. The evidence suggests that higher levels of stable behavior problems are typically associated with an increased risk for later academic and social–emotional difficulty when compared to children with the lowest levels of manifest problem behavior (e.g., Campbell et al., 2006; McDermott et al., 2019; Wildeboer et al., 2015).

Research also suggests that early emergent problem behaviors may not always signal continued difficulty (Campbell et al., 2016; Poulou, 2015). For example, Basten and colleagues (2016) examined stability and change in latent profiles of problem behavior among children assessed at the ages of 1.5, 3, and 6 years old. Four profiles were identified: (a) co-occurring internalizing and externalizing, (b) predominately externalizing, (c) some internalizing, and (d) no problems. Results from transition analysis indicated that while children in the co-occurring internalizing and externalizing profile demonstrated greatest risk for stable problems over time, it was unclear whether those in the externalizing only or in the internalizing only profiles would remain in their respective profiles at the subsequent time point or transition into the profile with no behavior problems.

One way to account for the instability in children's early problem behavior is by using a longitudinal assessment approach (Campbell et al., 2016; Epstein et al., 2004). This would yield growth trajectories that can be used to determine whether problem behavior will persist or decline over time, informing the need for early intervention. However, while longitudinal systems provide a distinct advantage over earliest assessment approaches, their implementation can be challenging. First, instruments designed to capture changes in children's problem behavior must demonstrate their ability to measure the same construct over time (Denham et al. 2008). The second issue relates to schools and early childcare programs themselves, where time or resource constraints may simply preclude the adoption of a longitudinal assessment system (Epstein et al., 2004; National Research Council, 2008). Given these challenges, there is a need to determine an optimal assessment approach for the timely treatment of children's emergent problem behaviors.

Recently, McDermott, Rovine, Buek, et al. (2018) demonstrated a research design for contrasting the predictive utility of earliest assessment and transitional change approaches. Using data from the HSIS, the authors employed a two-stage analytical strategy where the first step involved estimating appropriate growth curve models for each dimension of problem behavior examined. In the second step, distal outcomes at the end of first grade (direct assessments of student achievement, parent assessments of home behavioral adjustments, and teacher assessments of academic proficiency and classroom social adjustment) were regressed onto the estimated individual-level intercept and slope parameters from the first step using a multilevel logistic regression model. Lastly, receiver operating characteristic (ROC) curves were employed to evaluate the accuracy of predictions in the second step. Results for the ASETS phenotypes of problem behavior (Aggression, Attention Seeking, Low Energy, and Reticence/Withdrawal) showed that earliest assessment was sufficient for grounding early intervention.

In this article, we extend prior research by applying the McDermott, Rovine, Buek, et al. (2018) method to contrast earliest assessment versus transitional change approaches for situtypes of problem behavior (Peer Context Problems, Teacher Context Problems, and Learning Context Problems). To this end, we consider the following research questions. First, is there meaningful variation in children's growth trajectories as indicated by statistically significant random effects variance components? The second question relates to the primary goal of this article: What is the relative forecasting accuracy of earliest assessment and transitional change approaches for the prediction of later outcomes?

METHOD

This study applies data from the HSIS, a longitudinal, randomized controlled trial designed to investigate the effectiveness of Head Start and comparable prekindergarten programs. Sampling for HSIS was conducted using a clustered, multistage stratified design in which 84 Head Start grantees were randomly selected from the Northeast, North Central, South, Plains, and West geographic regions of the country. From these grantees, 383 Head Start centers were randomly selected. From among the pool of families applying for enrollment in the selected Head Start centers who were eligible for Head Start enrollment (defined by federal income criteria), children were randomly selected to enroll in Head Start or were permitted to enroll (or not) in a non–Head Start program.

Children's behavioral adjustment was assessed by teachers through 2 years of prekindergarten (PreK 1 and PreK 2), kindergarten (K), and first grade. Sample accretion occurred (i.e., PreK 1 $N=1{,}377$, PreK 2 $N=2{,}764$, K $N=2{,}873$, first grade $N=3{,}077$) because some children selected for prekindergarten enrollment did not enter school in the first year of PreK or were enrolled in noneducational settings and others did not enroll until kindergarten or when first grade attendance was legally required. The effects of missing data are assessed using sensitivity analyses.

The full national sample consisted of 3,827 children, where the mean age at entry was 4.0 years ($SD=0.5$), with 49.6% female, 37.8% Hispanic, 29.5% African American, 32.7% White or other race/ethnicity, 25.7% primarily Spanish-speaking, 12.8% identified with special needs, and 82.7% residing in urban areas. During PreK 1, children attended 540 preschool centers (867 classrooms); during PreK 2, children attended 1,032 centers (1,815 classrooms). During K, children attended 1,469 schools (2,280 classrooms); during first grade, children attended 1,617 schools (2,576 classrooms). During PreK, approximately 80% of classrooms were not affiliated with conventional schools (~60% were day care or other nonschool centers), with about 90% of post-PreK classrooms affiliated with public schools. Because there were no significant effects for Head Start participation on measures of problem behavior or first-grade outcomes, the full national sample was employed (USDHHS, 2010a).

Characteristics of participant teachers were reported from a 2003 survey that framed statistics as they related to the sample of participant children attending Head Start or non–Head Start centers (Research Connections, 2003). At that time, 97.1% of children had female teachers, 24.0% had teachers who were Latino, 47.2% had teachers who were White, and 31.0% had teachers who were African American. On average, children had teachers who had been teaching for 13.0 years ($SD=8.8$). Approximately 31.5% of children had teachers with an associate's degree, 37.3% had teachers with a bachelor's degree or higher, 45.7% had teachers with a state teaching certificate, 60.9% had teachers with a degree in early childhood or related field, and 52.9% had teachers with a child development associate credential.

Longitudinal Measures

The HSIS employed the ASETS (McDermott et al., 2013, 2014), a standardized teacher rating scale for the longitudinal assessment of classroom behavior problems over the 4 years of the study. The instrument includes 134 items embedded in 22 situational contexts that describe the child's behavior over the past month. The 22 contexts cover relationships with the teacher and with peers, as well as coping with classroom expectations. Factor analytic methods identified three scales of children's contextually based problem behaviors: Peer Context Problems, Teacher Context Problems, and Learning Context Problems. These factors are referred to as *situtypes*. Each scale is derived through longitudinal exploratory and confirmatory factor analyses across the four academic years, with Item Response Theory (IRT) calibration under the generalized partial credit model (Muraki, 1992), vertical equating using nonbiased linking items, and scaled scores via expected a posteriori Bayesian estimation where the population scaled score $M=50$ and $SD=10$ at PreK 1, the reference year. Internal consistency as derived from the IRT expected a posteriori scores and their standard errors was .91 for peer context problems, .83 for learning context problems, and .73 for teacher context problems. Substantial evidence for concurrent and predictive criterion validity, along with sensitivity to differential change detection, is described in McDermott et al. (2014).

The Peer Context Problems scale consists of 10 contexts associated with classroom problem behaviors that arise during peer interactions and includes contexts such as "Respecting others' belongings" and "Handling conflicts with children." The Learning Context Problems scale contains seven contexts relating to classroom problem behaviors that emerge when learning and includes situations such as "Paying attention in the classroom" and "Getting involved in classroom activities." Lastly, the Teacher Context Problems scale includes five contexts related to classroom behavior problems that arise during teacher–student interactions. Example contexts include "Talking with the teacher" and "Answering questions." Within each context are three to seven negative problem indicators that are theoretically and empirically reflective of a phenotype (Aggression, Attention Seeking, Reticence/Withdrawal, Low Energy behavior; see McDermott et al., 2013), and most contexts provide one or two positive or healthy behavior choices.

Distal Outcome Measures

Outcomes measured at the end of first grade were regressed onto scores on the ASETS situtype scales. Outcomes were assessed using a variety of approaches (direct assessment, teacher rating, parent rating) and encompass both academic achievement and social–emotional adjustment.

Academic Outcomes

Academic outcomes include direct assessments of achievement and teacher ratings of academic ability. Each measure is described below.

Direct assessments. The Basic Reading Skills cluster (letter and word reading and writing, phonemic and structural analysis) and Mathematics Reasoning cluster (quantitative concepts, counting, problem solving) of the Woodcock-Johnson III Tests of Achievement (WJ; Woodcock, McGrew, & Mather, 2002) were administered. Cluster-level scores were obtained in two steps: (a) the child's raw score on an individual test within a given cluster was converted into a W-Ability score and (b) within-cluster W-Ability scores were averaged to create a cluster-level score (USDHHS, 2010a). HSIS first-grade population internal consistency for Basic Reading Skills was .91 and that for Mathematics Reasoning was .78 (USDHHS, 2010a). Ample validity support has been reported for the two WJ achievement clusters (McGrew, Woodcock, & Schrank, 2007; Salvia, Ysseldyke, & Witmer, 2017).

A shortened version of the Peabody Picture Vocabulary Test, Third Edition (Dunn, Dunn, & Dunn, 1997) was used in the HSIS to assess receptive vocabulary. This version was equated to the full-length test version and adapted by applying three-parameter IRT calibration and Bayesian scoring (USDHHS, 2010a). Scores were estimated based on the child's pattern of correct/incorrect responses (USDHHS, 2010b). Internal consistency for the HSIS sample was .78, with abundant evidence for criterion validity (e.g., Salvia, Ysseldyke, & Bolt, 2007).

Teacher ratings. Teacher report of academic ability was rated at the end of first grade for language and literacy, mathematics, and social science, based on attainment of multiple skills compared to the attainment of peers (USDHHS, 2010a). Initially rated on a 5-point scale from 1 = *far below average* to 5 = *proficient*, the data were collapsed by USDHHS researchers to a simple binary scale (1–2 versus 3–5) to improve parsimony and reliability. Using the amended binary scale, the standard error of the mean is reported here, where SE_M for language and literacy = .008, mathematics = .008, and social science = .007. Evidence of criterion validity is provided for the HSIS population (USDHHS, 2010a).

Social–Emotional Outcomes

Social–emotional outcomes include teacher ratings of the student–teacher relationship and parent ratings of children's behavior problems. Details of each measure are provided below.

Teacher ratings. The Pianta Student–Teacher Relationships Scale (Pianta, 1996) is composed of 15 items, such as "I share an affectionate, warm relationship with this child" and "This child easily becomes angry at me," rated on a 5-point scale ranging from 1 = *definitely does not apply* to 5 = *definitely applies*. These items are broken into two subscales: Closeness (seven items) and Conflict (eight items). Evidence of concurrent and predictive validity is provided in Pianta (2001) and Pianta and Stuhlman (2004). Internal consistency for the HSIS sample = .82 for Closeness and .89 for Conflict (USDHHS, 2010a).

Parent ratings. The Total Behavior Problems scale is a parent report device that rates children's aggressive, defiant, hyperactive, withdrawn, or depressed behavior. The scale is composed of 14 binary items, such as "Is disobedient at home" and "Feels worthless or inferior." Development and validity evidence for the HSIS is provided in USDHHS (2010a), with additional evidence of validity being reported by other researchers (e.g., Ziv, Alva, & Zill, 2010). For the HSIS first-grade sample, internal consistency ranged from .78 to .79.

Data Analytic Strategy

Data analysis proceeded in two steps. First, multilevel individual growth curve modeling was applied to estimate each child's intercept and slope parameters for scores in each of the three situtypes—Peer Context Problems, Teacher Context Problems, and Learning Context Problems—across the four early education years. Second, multilevel generalized linear regression was used to regress the various distal outcomes on those individual child parameters. All models were estimated in SAS 9.4 (SAS, 2013) using the restricted maximum likelihood method.

In the first part of our analysis, we estimated the appropriate multilevel growth model for each situtype. Models incorporated statistically significant linear, quadratic, and cubic fixed effects as per Type 1 sequential *F* tests. The inclusion of higher order fixed effects was examined in order to adequately describe change over time (Singer & Willett, 2003). The initial model also included statistically significant random intercepts at the classroom level and random intercepts and linear slopes at the child level. Higher order child-level random slopes were added based on statistically significant Wald tests for the covariance parameters. Ancillary analyses of model fit were conducted using a series of likelihood ratio tests, contrasting the more complex growth model with a nested model featuring fewer estimated fixed and/or random effects (Singer & Willett, 2003). For these tests, full maximum likelihood estimation was employed to properly account for the

 DOI: 10.1080/2372966X.2020.1717372

addition of fixed effects parameters. General model specification was $\hat{Y}_{ijk} = \gamma_{000} + \gamma_{100}\text{Time}_{ijk} + \gamma_{200}\text{Time}^2_{ijk} + \gamma_{300}\text{Time}^3_{ijk} + (\mu_{00k} + \mu_{10k}\text{Time}_{ijk}) + (\mu_{0jk} + \mu_{1jk}\text{Time}_{ijk}) + r_{ijk}$. Estimated values for the child-level intercepts and slopes were then used as predictors for the second step of the analysis. Moving forward, the intercept is referred to as the child's *level* and the slope as the child's *change*. As such, the former is the estimated level of problem behavior within a specific context at first assessment (i.e., PreK 1) and the latter is the rate and shape of change over the four assessment years (i.e., PreK 1 to first grade).

In the second step, binary distal outcomes were generated and regressed on each child's individual level and change values for problem behavior. Binary outcomes were deemed appropriate because (a) the alternative WJ normal curve equivalent scores and parent and teacher ratings were significantly abnormally distributed and differentially skewed; (b) the WJ item domain representation was relatively sparse below the 25th percentile with punctuated rather than graduated changes in item difficulty (a problem common to commercial tests used with low-income populations; see McDermott et al., 2009); (c) teachers' assessments of child performance in language and literacy, mathematics, and social science had already been bifurcated by the federal government for psychometric efficiency; and (d) dichotomously scaled outcomes would yield *relative probabilities* of desirable versus undesirable outcomes in late first grade as related to children's initial levels and gradual changes in problem behavior. Thus, an outcome reflecting reading proficiency versus nonproficiency was formed from WJ Basic Reading Skills scores where proficiency included performance in the upper four quintiles (scored 0) and nonproficiency included performance in the lowest quintile (scored 1). Correspondingly, a mathematics nonproficiency variable was formed from WJ Mathematics Reasoning (lowest quintile = 1), and similar nonproficiency variables were formed from the Peabody Picture Vocabulary Test, Pianta's Closeness with Teacher, and teacher assessments of language and literacy, mathematics, and social science (lower quintile = 1). A parent home problem behaviors indicator and Pianta's Conflict with Teacher indicator were also formed from the respective scales (upper quintile = 1). Quintiles were preferred because they provided the necessary statistical power for reliable point separation in multilevel generalized (logistic) modeling (Stokes, Davis, & Koch, 2001).

General model specification in step 2 was $\text{Nonproficiency}_{\Theta[\text{logit}]ij} = \gamma_{00} + \gamma_{10}\text{Level}_i + \gamma_{20}\text{Change}_i + \mu_{0j}$, where i indexes children and j represents teachers/classrooms. In this model, predicted values for each child's level and change were used as explanatory predictors in a multilevel generalized linear model that estimated the likelihood of each undesirable outcome (mathematics nonproficiency, total behavior problems, etc.). The relative risk increment or reduction for better versus poorer outcomes associated with children's initial problem behavior level and transitional change was estimated through the odds ratio. Finally, the accuracy of predictions was summarized through ROC curve models, which plot the probability of nonproficiency given more problem behaviors (i.e., sensitivity) against the probability of nonproficiency given fewer problem behaviors (i.e., 1 – specificity; Swets, Dawes, & Monahan, 2000).

RESULTS

Parameters for the growth models (pertaining to each form of longitudinal behavior problem) from the first step are presented in Table 1. For Peer Context Problems, a model with a quadratic fixed effect parameter was preferable, whereas growth models with cubic fixed effects provided the best fit for both the Teacher Context Problems and Learning Context Problems scales. All models featured negative linear trends, though Teacher Context Problems and Learning Context Problems each had positive quadratic and negative cubic changes over the early transition years. Results from ancillary analyses of model fit using a series of likelihood ratio tests

Table 1. Properties of Multilevel Individual Growth Curve Models for Early Childhood Problem Contexts Across Four Years

Effect	Peer	Teacher	Learning
Fixed effects parameter estimates (*SE*)[a]			
Intercept	49.87 (0.20)	49.47 (0.22)	50.55 (0.20)
Linear slope	−1.38 (0.24)	−2.50 (0.60)	−3.97 (0.52)
Quadratic slope	0.26 (0.07)	2.11 (0.49)	2.35 (0.42)
Cubic slope		−0.45 (0.11)	−0.40 (0.09)
Random effects parameter estimates (*SE*)[b]			
Variance/covariance			
Classroom intercepts	14.00 (2.49)	8.09 (3.73)	16.91 (2.37)
Child intercepts	30.40 (2.31)	15.86 (2.95)	22.69 (2.51)
Child linear slopes	2.33 (0.45)	13.14 (6.83)	26.20 (5.41)
Child quadratic slopes		1.25 (0.66)	2.46 (0.52)
Child intercepts by linear slopes	−0.82 (0.85)†	−2.42 (3.55)†	−10.55 (2.94)
Child intercepts by quadratic slopes		−0.03 (1.03)†	2.81 (0.85)
Child linear by quadratic slopes		−3.65 (2.08)†	−7.45 (1.63)
Residual	26.90 (2.37)	43.31 (3.81)	20.70 (2.41)
Random effects variance decomposition[c]			
% Variance			
Between classrooms	19.0	10.0	19.0
Between children	44.0	37.0	58.0
Level (viz., intercept)	**92.9**	**54.7**	**43.9**
Change (viz., slope)	**7.1**	**47.6**	**55.8**
Within children	37.0	53.0	23.0

Note. $N = 3{,}827$.
[a]Values are based on Type I (sequential) *F* tests, with statistical significance at $p < .05$. Unreported higher order effects indicate statistical nonsignificance and exclusion from a model.
[b]Values are based on restricted maximum likelihood estimation, with statistical significance at $p < .05$ unless indicated by a dagger (†) for nonsignificance.
[c]For a given area of behavior problems (column), the values for variances between classrooms, between children, and within children sum to 100.0% of all estimated variances. Boldface entries pertain to the focal variance components for the overall study. For a given area of behavior problems (column), the values for level and change sum to 100.0% of the estimated variance between children.

supported these results, with the most preferable models significant at $p < .05$. The variance decomposition for the random effects in the bottom panel of Table 1 is the particular focus of this study. Though all models incorporated linear random effects, only models for the Teacher Context and Learning Context Problems scales included quadratic random effects estimates (higher order cubic random effects were not estimable). Results from the first step of the analysis showed that 44.0% of the variance in the Peer Context Problems, 37.0% in the Teacher Context Problems, and 58.0% in the Learning Context Problems scales is between children. For Peer Context Problems, 92.9% of this variance is associated with differences in level (earliest assessment), with the remaining 7.1% attributable to differences in change (transitional change). For Teacher Context Problems, 54.7% of between-child variance is associated with earliest assessment and 45.3% with change over time. For Learning Context Problems, 43.9% of the variance is associated with level and 55.8% with change. Therefore, the trajectories associated with problem behaviors in teacher contexts and in learning situations are more complex and heterogeneous than they are for behavior problems in situations involving peers.

Level Versus Change

The second step employed a multilevel logistic regression model to estimate distal outcomes at the end of first grade. In this model, each child's estimated level (intercept) and change (linear and, if estimable, quadratic slope) from the first step were included as simultaneous predictors, thereby producing estimates that control for one another. Associated odds ratios and risk increments (%) for a child's inclusion in the least desirable quintile for each distal outcome are presented in Table 2. As illustration, only the odds ratio

Table 2. Multilevel Adjusted Risk Odds for Negative Distal Outcomes as Associated With Level and Subsequent Change in Early Childhood Problem Contexts

	Peer		Teacher		Learning	
	Odds Ratio (95% CIs)	% Risk Increase	Odds Ratio (95% CIs)	% Risk Increase	Odds Ratio (95% CIs)	% Risk Increase
First grade reading nonproficiency (direct assessment, ICC = .57, *n* = 2,873)						
Level	1.20 (1.14, 1.26)	20.1	1.22 (1.09, 1.36)	21.6	1.32 (1.22, 1.43)	31.9
Linear change	1.07 (0.85, 1.36)†		2.05 (1.55, 2.71)	105.0	1.59 (1.33, 1.92)	59.3
Quadratic change			5.00 (2.11, 11.83)	400.1	2.28 (1.34, 3.88)	128.3
First grade mathematics nonproficiency (direct assessment, ICC = .51, *n* = 2,879)						
Level	1.16 (1.11, 1.22)	16.2	1.16 (1.06, 1.27)	16.4	1.31 (1.22, 1.41)	31.4
Linear change	1.15 (0.92, 1.44)†		1.99 (1.59, 2.50)	99.4	1.63 (1.39, 1.91)	62.5
Quadratic change			4.38 (2.16, 8.87)	338.1	2.47 (1.55, 3.96)	147.4
First grade receptive vocabulary nonproficiency (direct assessment, ICC = .40, *n* = 2,883)						
Level	1.03 (0.99, 1.07)†		1.07 (0.98, 1.16)†		1.11 (1.05, 1.17)	10.6
Linear change	0.96 (0.78, 1.18)†		1.27 (1.04, 1.56)	27.4	1.05 (0.92, 1.19)†	
Quadratic change			1.80 (0.94, 3.43)†		1.06 (0.70, 1.60)†	
First grade language and literacy nonproficiency (teacher rating, ICC = .15, *n* = 3,042)						
Level	1.09 (1.06, 1.12)	8.8	1.22 (1.14, 1.31)	22.1	1.23 (1.17, 1.29)	22.8
Linear change	1.72 (1.45, 2.04)	71.9	2.06 (1.73, 2.46)	106.0	2.11 (1.83, 2.44)	111.4
Quadratic change			11.85 (6.59, 21.31)	>999	10.45 (6.64, 16.46)	945.5
First grade mathematics nonproficiency (teacher assessment, ICC = .20, *n* = 3,029)						
Level	1.12 (1.09, 1.16)	12.4	1.20 (1.12, 1.30)	20.6	1.27 (1.19, 1.34)	26.7
Linear change	1.60 (1.33, 1.92)	59.8	2.15 (1.77, 2.60)	114.6	2.19 (1.85, 2.60)	119.0
Quadratic change			8.69 (4.75, 15.92)	769.3	11.15 (6.55, 18.97)	>999
First grade social studies and science nonproficiency (teacher assessment, ICC = .26, *n* = 3,020)						
Level	1.13 (1.08, 1.18)	12.9	1.29 (1.17, 1.41)	28.7	1.29 (1.19, 1.40)	29.4
Linear change	2.08 (1.63, 2.66)	108.1	2.65 (2.07, 3.39)	164.9	2.71 (2.12, 3.47)	171.1
Quadratic change			24.12 (10.88, 53.45)	>999	23.67 (10.88, 51.50)	>999
First grade home behavior problems (parent rating, ICC = .19, *n* = 2,900)						
Level	1.13 (1.09, 1.16)	12.6	1.24 (1.16, 1.34)	24.2	1.19 (1.13, 1.24)	18.7
Linear change	1.34 (1.12, 1.60)	33.9	1.38 (1.18, 1.62)	38.0	1.34 (1.20, 1.49)	33.8
Quadratic change			3.12 (1.84, 5.28)	211.8	2.23 (1.57, 3.16)	122.9
First grade conflict with teacher (teacher rating, ICC = .19, *n* = 3,050)						
Level	1.31 (1.23, 1.40)	31.0	1.58 (1.42, 1.75)	57.6	1.23 (1.17, 1.31)	23.4
Linear change	15.06 (8.48, 26.73)	1,405.8	4.77 (3.57, 6.37)	377.2	3.03 (2.53, 3.63)	202.9
Quadratic change			189.17 (73.59, 486.30)	>999	32.48 (18.27, 57.73)	>999
First grade lack of closeness with teacher (teacher rating, ICC = .29, *n* = 3,053)						
Level	1.08 (1.05, 1.12)	8.2	1.46 (1.33, 1.62)	46.4	1.20 (1.13, 1.27)	19.9
Linear change	1.69 (1.39, 2.05)	68.8	3.20 (2.49, 4.12)	220.4	1.78 (1.53, 2.06)	77.5
Quadratic change			91.56 (37.37, 224.35)	>999	7.94 (4.82, 13.11)	694.4

Note. CI = confidence interval; ICC = intraclass coefficient. ICC = proportion of between-classroom variance estimated for a given distal outcome variable at the end of first grade. Odds ratios and confidence limits are derived from parameter estimates obtained through multilevel generalized linear modeling applying adaptive quadratures, the Bernoulli response distribution, and logit link function. Percentage risk increase = odds ratio − 1 (100) and expresses increase in risk of future nonproficiency per one-point increase in estimated level, linear change, or quadratic change.
All values are statistically significant at $p < .05$ unless indicated with a dagger (†) for nonsignificance. Significant odds ratios have confidence intervals that do not cross 1.00.

associated with a child's level (earliest assessment) on the Peer Context Problems scale is significant for the prediction of first-grade reading nonproficiency, where a one-point increase in problem behaviors results in a 20.1% increase in risk of nonproficiency. Transitional change on the Peer Context Problems scale did not reliably add to the prediction of reading nonproficiency. For the same outcome, both earliest assessment and transitional linear and quadratic change on the Teacher Context Problems scale reliably predict reading nonproficiency in the first grade. Here, a child's later risk increases by 105.0% for each one-point increase in linear change. Lastly, for the Learning Context Problems scale, both earliest assessment and transitional change (linear and quadratic) reliably predict later reading nonproficiency, where a one-point increase in quadratic change is associated with a 128.3% risk increment. Overall, children's estimated levels and changes reliably predicted most distal outcomes (teacher assessments of achievement, direct assessments of achievement, parent ratings of home adjustment, teacher ratings of classroom adjustment).

ROC Analysis

Predictive utility of the estimated levels and changes were assessed using ROC analyses, where the area under the curve (AUC) is estimated based on the relative ability of levels versus changes to predict a distal outcome. (An AUC of .50 refers to chance accuracy and 1.00 perfect predictive accuracy. Values ≥.90 indicate high accuracy.) Based on those models where level effects were statistically significant (25 models), the AUC M = .946 (SD = .042; range = .854–.996) and change effects (23 models) AUC M = .943 (SD = .038; range = .880–.998), indicating that differences between level and change forecasting accuracy are inconsequential. Joint forecasting accuracy (22 models) AUC M = .941 (SD = .041; range = .856–.997), a figure representing no improvement over the accuracy yielded by either initial assessment or transitional changes in isolation.

Sensitivity Analysis

Given that multilevel individual growth curve models account for missing data (Allison, 2012), all analyses were repeated as based solely on those children who were present at PreK 1. For step 1 multilevel growth models, the sensitivity analysis produced results that were uniformly consistent with results based on the full data set in terms of significant fixed and random effects. Estimates for the step 2 multilevel logistic models were also largely consistent with those from the full data set, with the primary exception isolated to five of the nine quadratic parameter estimates associated with the Teacher Contexts Problems scale. Lastly, the majority of ROC models reproduced results consistent with the full sample with the exception of 8 out of 72 models, related to teacher-graded mathematics, where the AUC dropped below .70. Overall, results supported the assumption that the children in the full sample were observed at random with some data missing at random and unrelated to levels of or changes in the longitudinal variables (Little & Rubin, 2002; Marini, Olsen, & Rubin, 1979).

DISCUSSION

This research examined the relative contribution of early assessment and transitional change in children's contextually based problem behaviors for the prediction of later important outcomes. To accomplish this, a two-stage modeling approach was employed where the first step fit a multilevel growth model for a nationally representative sample of children from low-income families. From this model, each child's estimated initial assessment and growth in those problem behaviors were extracted and applied as predictors in the second step. The second stage of the approach incorporated indices from the first stage in a multilevel generalized linear model to predict later academic and social difficulties. The predictive contributions of initial assessment and transitional change were subsequently compared through a series of ROC analyses.

Fixed effects estimates for each multilevel growth model revealed that problem behaviors decreased nonlinearly throughout the observation period, corresponding with normative declines in behavioral disturbance during the early education period (Campbell, 2002). Substantial between-child variation about the estimated mean level (earliest assessment) and the estimated mean change (transitional change) was also found for each situtype, where positive deviations in problem behavior at either earliest assessment or in transitional change forecast increased risk for later academic and social–emotional difficulty. With respect to the latter, this refers to a slower than average decline in behavioral disturbance or an increase in problem behavior over time. Though the relative risk increments were relatively modest for earliest assessments, increased problem behavior associated with transitional change (linear or quadratic) was much more likely to lead to later undesirable outcomes, which agrees with prior research suggesting that stable problem behavior over time is likely to signal increased risk for later academic and social–emotional difficulty (Campbell et al., 2006).

Differences were also found in the shape of change, where peer context problems followed a decreasing quadratic trend, whereas problems in teacher or learning contexts featured primarily decreasing but more complex cubic curvature. The decline in peer context problems may partly reflect the development of social skills that children can use to resolve peer conflict (Campbell et al., 2016; Chen et al., 2001) as well as preschool teachers' efforts to promote positive social behaviors (Stormont, Lewis, & Beckner, 2005; USDHHS, 2015) and minimize problem behavior (Stormont et al., 2005). For the developmental trajectories of problems arising in teacher and learning contexts, clues regarding their increased complexity may be ascertained through inspection of the associated random effects estimates, which suggest that children may experience highly individualized pathways related to the formation of teacher–child relationships and the acquisition of cognitive skills, respectively. This is further discussed in the later parts of this section.

The primary focus of this study was the predictive utility associated with earliest assessment and transitional change in children's situationally based problem behavior. Results from a series of ROC analyses demonstrated that both indicators were highly and equally accurate in their prediction of later outcomes and that their combined forecasting power did not enhance predictive accuracy. This was uniformly found for each situtype, suggesting that interventions addressing problem behaviors that arise in classroom-specific situations with peers, teachers, or learning activities can rely on evidence obtained at earliest assessment. These results agree with McDermott, Rovine, Buek, et al. (2018) as well as with leading research that emphasizes treatment at the earliest stages of emergent problem behavior (Bornstein, Hahn, & Suwalsky, 2013; Brassard & Boehm, 2007; Poulou, 2015). However, the present study's findings are unique because they pertain to the classroom-specific situations associated with children's early education problem behaviors (i.e., situtypes) rather than phenotypic definitions of problem behavior.

Our results also showed varying degrees of heterogeneity in children's individual trajectories across situtypes (see Table 1). For trajectories of peer context problems, over 90% of the between-child variability was associated with earliest assessment, suggesting that children entered preschool with different initial levels of peer problem behavior that tended to decrease over time, a finding consistent with research showing normative declines in preschool peer problem behaviors (Raikes, Virmani, Thompson, & Hatton, 2013). Substantial heterogeneity at earliest assessment further suggests that individual differences in peer context problem trajectories were likely related to factors preceding preschool entry. A wealth of research shows that early family experiences, such as the interactions between parent and child, play an important role in children's later social–emotional competence (e.g., Clark & Ladd, 2000; Elicker, Englund, & Sroufe, 1992; Hay, Payne, & Chadwick, 2004; Shaw, Hyde, & Brennan, 2012). For example, the quality of attachment in the parent–child relationship has been implicated as an important predictor of children's later peer-related skills (Elicker et al., 1992; Hay et al., 2004; Raikes et al., 2013).

Unlike trajectories associated with peer context problems, random effects estimates for both teacher and learning context problems demonstrated substantial heterogeneity in children's individual development, where approximately half of the between-child variance was associated with transitional change. With respect to trajectories of teacher context problems, this observed heterogeneity might be partly associated with the quality of teacher–child relationships (Baker et al., 2008). Research suggests that teacher–child relationships can shape children's later academic and social–emotional development (Hamre & Pianta, 2001), where high-quality relationships may afford protection against later negative outcomes partly through increased teacher support (Baker et al., 2008). Furthermore, high-quality teacher–child relationships have been shown to reduce the frequency of internalizing behaviors among children with high levels of socially withdrawn behavior (O'Connor et al., 2011). However, teacher–child relationships marked by conflict or dependency (Hamre & Pianta, 2001) may lead to reinforcement of problematic patterns of behavioral functioning (Doumen et al., 2008). Thus, high variability in the trajectories of teacher context problems might be associated with heterogeneity in the longitudinal patterns of teacher–child relationships.

Studies have found substantial variability in the development of teacher–child relationships across elementary grades. For example, O'Connor et al. (2011) found four trajectories of student–teacher relationships: (a) poor–worsening, (b) poor–improving, (c) strong–worsening, and (d) strong. Their results suggested that students in the strong teacher–child relationship group demonstrated the lowest levels of externalizing problem behaviors throughout the study period. Research by Spilt et al. (2012) examined trajectories associated with two separate dimensions of teacher–child relationships, Warmth and Conflict. Analyses broken out by gender indicated two distinct pathways of Warmth for both boys and girls, whereas Conflict evinced three or four pathways for girls and boys, respectively.

Taken together, these findings suggest a potential relationship between high-quality teacher–child relationships and children's teacher context problems. Therefore, addressing problems that arise in interactions with teachers might be one avenue for enhancing teacher–child relationships. It is important to note, however, that the above studies used data specific to elementary grades, whereas the present study suggests heterogeneity beginning in preschool. Therefore, future research should focus on investigating heterogeneity in the developmental pathways of teacher–child relationships from preschool through the critical transition into formal schooling.

 DOI: 10.1080/2372966X.2020.1717372

Problem behavior trajectories associated with learning contexts were also quite heterogeneous across the early education period. Although research links high-quality teacher–child relationships with more positive teacher perceptions of student classroom engagement (e.g., Baker et al., 2008; Cadima, Doumen, Verschueren, & Buyse, 2015), behavior problems in learning contexts may also be inversely related with a subset of noncognitive skills known as learning behaviors (Fantuzzo et al., 2004), which explain the processes by which children go about the classroom learning process. Recent evidence suggests that among low-income children, there are dominant developmental trajectories of learning behavior that persist from preschool into first grade that differentially predict later academic and sociobehavioral performance (McDermott, Rovine, Reyes, et al., 2018). Notably, learning behaviors declined precipitously for those students who exhibited the most effective learning behaviors in Head Start upon kindergarten entry. In the context of preschool performance fade-out, whereby gains in the cognitive and sociobehavioral skills observed in preschool fade after entry into formal schooling, the authors argued that the timing of such losses in children's learning behaviors may indeed be associated with the later decline in academic and behavioral adjustment skills. Thus, monitoring and addressing problem behaviors that arise in learning contexts may be one approach for sustaining improvements made in preschool programs.

Limitations

This research has certain limitations. The first pertains to the instruments used in the HSIS. Note that because most measures in the HSIS were modified versions of the original instruments, their ability to thoroughly cover the construct of interest may have been compromised. In turn, we recognize that the distal outcomes examined in this study may not be able to capture all relevant facets of a given construct. To address this, we examined outcomes of various types with different raters and methods, and we used only those measures of classroom behavior that were rigorously validated and shown to be reliable.

Second, the generalizability of results is limited due to the characteristics of the study sample. Note that the focus of the HSIS was on children who came from low-income families who were eligible for Head Start entry. Although nationally representative, results of the current study should be extrapolated only to those populations who share demographic and socioeconomic characteristics similar to those the HSIS sample.

The last limitation pertains to the predictors examined in this study, initial assessment and subsequent change. We note that though both were shown to be good predictors of later performance, they do not explain all of the variance related to the outcomes examined. The ROC curves, however, demonstrate the quality of models used in this study.

Implications and Conclusion

In light of these limitations, this research has implications for practice. The results indicated that for each classroom context reviewed in this study, both earliest assessment and transitional change provided high forecasting accuracy and that the addition of repeated measurements did not improve accuracy. As such, the evidence points to the importance of identifying problems early. Children exhibiting high levels of classroom problem behavior might warrant further investigation into the motivation behind those behaviors (Brassard & Boehm, 2007; McDermott et al., 2014). Because contextually based measures of problem behavior capture the attendant situations surrounding those behaviors, this information can serve to inform practitioners' strategies for addressing behavior problems (Mischel, 2004; Zayas et al., 2008). For instance, severe problem behaviors that emerge in peer situations might be partly associated with factors preceding preschool entry (e.g., Elicker et al., 1992; Hay et al., 2004; Raikes et al., 2013), whereas behavior problems in teacher situations may stem in part from difficulties in the teacher–child relationship (Baker et al., 2008; Doumen et al., 2008); problems that arise out of learning situations may point to problems in the focal child's learning behaviors (McDermott, Rovine, Reyes, et al., 2018). But monitoring children's progress can also be important for verifying that interventions are producing the intended outcomes and to detect newly emerging behavior problems.

DISCLOSURE

The authors have no conflicts of interest to report.

ACKNOWLEDGMENTS

This research was conducted with the cooperation of the U.S. Department of Health and Human Services, Administration for Children and Families.

ORCID

Marley W. Watkins http://orcid.org/0000-0001-6352-7174

REFERENCES

Allison, P. (2012). *Handling missing data by maximum-likelihood. SAS Global Forum*. Cary, NC: SAS Institute.

Baker, J. A., Grant, S., & Morlock, L. (2008). The teacher–student relationship as a developmental context for children with internalizing or externalizing behavior problems. *School Psychology Quarterly*, *23*(1), 3–15. doi:10.1037/1045-3830.23.1.3.

Basten, M., Tiemeier, H., Althoff, R. R., van de Schoot, R., Jaddoe, V. W. V., Hofman, A., …, van der Ende, J. (2016). The stability of problem behavior across the preschool years: An empirical approach in the general population. *Journal of Abnormal Child Psychology*, *44*, 393–404. doi:10.1007/s10802-015-9993-y.

Bornstein, M. H., Hahn, C.-S., & Suwalsky, J. T. D. (2013). Developmental pathways among adaptive functioning and externalizing and internalizing behavioral problems: Cascades from childhood into adolescence. *Applied Developmental Science*, *17*(2), 76–87. doi:10.1080/10888691.2013.774875.

Brassard, M. R., & Boehm, A. E. (2007). *Preschool assessment: Principles and practice*. New York, NY: Guilford.

Buhs, E. S., & Ladd, G. W. (2001). Peer rejection as an antecedent of young children's school adjustment: An examination of mediating processes. *Developmental Psychology*, *37*(4), 550–560. doi:10.1037/0012-1649.37.4.550.

Bulotsky-Shearer, R. J., Bell, E. R., & Domínguez, X. (2012). Latent profiles of problem behavior within learning, peer, and teacher contexts: Identifying subgroups of children at academic risk across the preschool year. *Journal of School Psychology*, *50*(6), 775–798. doi:10.1016/j.jsp.2012.08.001.

Bulotsky-Shearer, R. J., Fantuzzo, J. W., & 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. doi:10.1037/0012-1649.44.1.139.

Cadima, J., Doumen, S., Verschueren, K., & Buyse, E. (2015). Child engagement in the transition to school: Contributions of self-regulation, teacher–child relationships, and classroom climate. *Early Childhood Research Quarterly*, *32*, 1–12. doi:10.1016/j.ecresq.2015.01.008.

Campbell, S. B. (2002). *Behavior problems in preschool children*. New York, NY: Guilford.

Campbell, S. B., Denham, S. A., Howarth, G. Z., Jones, S. M., Whittaker, J. V., Williford, A. P., … Darling-Churchill, K. (2016). Commentary on the review of measures of early childhood social and emotional development: Conceptualization, critique, and recommendations. *Journal of Applied Developmental Psychology*, *45*, 19–41. doi:10.1016/j.appdev.2016.01.008.

Campbell, S. B., Spieker, S., Burchinal, M., & Poe, M. D., & The NICHD Early Child Care Network (ECCRN). (2006). Trajectories of aggression from toddlerhood to age 9 predict academic and social functioning through age 12. *Journal of Child Psychology and Psychiatry*, *47*(8), 791–800. doi:10.1111/j.1469-7610.2006.01636.x.

Chen, D. W., Fein, G. G., Killen, M., & Tam, H.-P. (2001). Peer conflicts of preschool children: Issues, resolution, incidence, and age-related patterns. *Early Education & Development*, *12*(4), 523–544. doi:10.1207/s15566935eed1204_3.

Clark, K. E., & Ladd, G. W. (2000). Connectedness and autonomy support in parent–child relationships: Links to children's socioemotional orientation and peer relationships. *Developmental Psychology*, *36*(4), 485–498. doi:10.1037/0012-1649.36.4.485.

Denham, S. A., Wyatt, T. M., Bassett, H. H., Echeverria, D., & Knox, S. S. (2008). Assessing social–emotional development in children from a longitudinal perspective. *Journal of Epidemiology & Community Health*, *63*(Suppl 1), i37–i52. doi:10.1136/jech.2007.070797.

Doumen, S., Verschueren, K., Buyse, E., Germijs, V., Luyckx, K., & Soenens, B. (2008). Reciprocal relations between teacher–child conflict and aggressive behavior in kindergarten: A three-wave longitudinal study. *Journal of Clinical Child & Adolescent Psychology*, *37*, 588–599. doi:10.1080/15374410802148079.

Downer, J. T., Booren, L. M., Lima, O. K., Luckner, A. E., & Pianta, R. C. (2010). The Individualized Classroom Assessment Scoring System (inCLASS): Preliminary reliability and validity of a system for observing preschoolers' competence in classroom interactions. *Early Childhood Research Quarterly*, *25*(1), 1–16. doi:10.1016/j.ecresq.2009.08.004.

Dunn, L. M., Dunn, L. L., & Dunn, D. M. (1997). *Peabody picture vocabulary test* (3rd ed.). Circle Pines, MN: American Guidance.

Elicker, J., Englund, M., & Sroufe, L. A. (1992). Predicting peer competence and peer relationships in childhood from early parent–child relationships. In R. D. Parke & G. W. Ladd (Eds.), *Family-Peer relationships.* (pp. 77–106). Hillsdale, NJ: Lawrence Erlbaum.

Epstein, A. S., Schweinhart, L. J., DeBruin-Parecki, A., & Robin, K. B. (2004). Preschool assessment: A guide to developing a balanced approach (Issue Brief No. 7). Retrieved from http://nieer.org/.

Fantuzzo, J., Perry, M. A., & McDermott, P. (2004). Preschool approaches to learning and their relationship to other relevant classroom competencies for low-income children. *School Psychology Quarterly*, *19*(3), 212–230. doi:10.1521/scpq.19.3.212.40276.

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. doi:10.1111/1467-8624.00301.

Hamre, B. K., & Pianta, R. C. (2010). Classroom environments and developmental processes. In J. L. Meece & J. S. Eccles (Eds.), *Handbook of research on schools, schooling, and human development.* (pp. 25–41). New York, NY: Routledge.

Hay, D. F., Payne, A., & Chadwick, A. (2004). Peer relations in childhood. *Journal of Child Psychology and Psychiatry*, *45*(1), 84–108. doi:10.1046/j.0021-9630.2003.00308.x.

Isenberg, E., Max, J., Gleason, J., Johnson, M., Deutsch, J., Hansen, M., & Angelo, L. (2016). Do low-income students have equal access to effective teachers? Evidence from 26 districts (Report No. NCEE 2017-4007). Retrieved from https://ies.ed.gov/ncee/pubs/20174008/pdf/20174007.pdf.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data.* (2nd ed.). New York, NY: Wiley.

Marini, M. M., Olsen, A. R., & Rubin, D. B. (1979). Maximum likelihood estimation in panel studies with missing data. In K. Schuessler (Ed.), *Sociological methodology 1980*. San Francisco, CA: Jossey Bass.

McClelland, M. M., & Morrison, F. J. (2003). The emergent of learning-related social skills in preschool children. *Early Childhood Research Quarterly*, *18*(2), 206–224. doi:10.1016/S0885-2006(03)00026-7.

McDermott, P. A., Fantuzzo, J. W., Waterman, C., Angelo, L. E., Warley, H. P., Gadsden, V. L., & Zhang, X. (2009). Measuring preschool cognitive growth while it's still happening: The

Learning Express. *Journal of School Psychology*, *47*(5), 337–366. doi:10.1016/j.jsp.2009.07.002.

McDermott, P. A., Rovine, M. J., Buek, K. W., Reyes, R. S., Chao, J. L., & Watkins, M. W. (2018). Initial assessment versus gradual changes in early childhood education—which better foretells the future?*Psychology in the Schools*, *55*(9), 1071–1085. doi:10.1002/pits.22150.

McDermott, P. A., Rovine, M. J., Reyes, R. S., Chao, J. L., Scruggs, R., Buek, K., & Fantuzzo, J. W. (2018). Trajectories or early education learning behaviors among children at risk: A growth mixture modeling approach. *Psychology in the Schools*, *55*(10), 1205–1223.

McDermott, P. A., Steinberg, C. M., & Angelo, L. E. (2005). Situational specificity makes the difference in assessment of youth behavior disorders. *Psychology in the Schools*, *42*(2), 121–136. doi:10.1002/pits.20050.

McDermott, P. A., Watkins, M. W., Rovine, M. J., & Rikoon, S. H. (2013). Assessing changes in socioemotional adjustment across the early school transitions—New national scales for children at risk. *Journal of School Psychology*, *51*(1), 97–115. doi:10.1016/j.jsp.2012.10.002.

McDermott, P. A., Watkins, M. W., Rovine, M. J., & Rikoon, S. H. (2014). Informing context and change in young children's sociobehavioral development—The national adjustment scales for early transition in schooling (ASETS). *Early Childhood Research Quarterly*, *29*(3), 255–267. doi:10.1016/j.ecresq.2014.02.004.

McDermott, P. A., Watkins, M. W., Rovine, M. J., Rikoon, S. H., Irwin, C. W., Reyes, R., & Chao, J. L. (2019). Emergent growth patterns of early education self-control problems among children from underresourced families. *Early Childhood Research Quarterly*, *48*, 1–13.

McGrew, K. S., Woodcock, R. W., & Schrank, K. A. (2007). *Woodcock–Johnson III normative update technical manual*. Itasca, IL: Riverside.

Mischel, W. (2004). Toward an integrative science of the person. *Annual Review of Psychology*, *55*(1), 1–22. doi:10.1146/annurev.psych.55.042902.130709.

Mischel, W., Shoda, Y., & Mendoza-Denton, R. (2002). Situation-behavior profiles as a locus of consistency in personality. *Current Directions in Psychological Science*, *11*, 50–54. doi:10.1111/1467-8721.00166.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, *16*(2), 159–176. doi:10.1177/014662169201600206.

National Research Council (NRC). (2008). *Early childhood assessment: Why, what, and how*. Washington, DC: National Academies Press. Retrieved from: https://www.nap.edu/.

National Research Council and Institute of Medicine (NRCIM). (2009). *Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities.* Washington, DC: The National Academies Press. doi:10.17226/12480.

Neugebauer, S. R. (2014). Context-specific motivations to read for adolescent struggling readers: Does the motivation for reading questionnaire tell the full story?*Reading Psychology*, *35*(2), 160–194. doi:10.1080/02702711.2012.679171.

O'Connor, E. E., Dearing, E., & Collins, B. A. (2011). Teacher–child relationship and behavior problem trajectories in elementary school. *American Educational Research Journal*, *48*, 120–162. doi:10.3102/0002831210365008.

Pianta, R. G. (1996). *Student-Teacher Relationship Scale*. Charlottesville, VA: University of Virginia.

Pianta, R. C. (2001). *Student-Teacher Relationship Scale professional manual*. Lutz, FL: Psychological Assessment Resources.

Pianta, R. C., & Stuhlman, M. W. (2004). Teacher–child relationships and children's success in the first years of school. *School Psychology Review*, *33*, 444–458.

Poulou, M. S. (2015). Emotional and behavioural difficulties in preschool. *Journal of Child and Family Studies*, *24*(2), 225–236. doi:10.1007/s10826-013-9828-9.

Qi, C. H., & Kaiser, A. P. (2003). Behavior problems of preschool children from low-income families: Review of the literature. *Topics in Early Childhood Special Education*, *23*, 188–216. doi:10.1177%2F02711214030230040201.

Raikes, H. A., Virmani, E. A., Thompson, R. A., & Hatton, H. (2013). Declines in peer conflict from preschool through first grade: Influences from early attachment and social information processing. *Attachment & Human Development*, *15*, 65–82. doi:10.1080/14616734.2012.728381.

Raver, C. C., & Knitzer, J. (2002). *Ready to enter: What research tells policymakers about strategies to promote social and emotional school readiness among three- and four-year old children*. New York, NY: National Center for Children in Poverty, Mailman School of Public Health, Columbia University. doi:10.7916/D82V2QVX.

Reiss, F. (2013). Socioeconomic inequalities and mental health problems in children and adolescents: A systematic review. *Social Science & Medicine*, *90*, 24–31. doi:10.1016/j.socscimed.2013.04.026.

Research Connections. (2003). *Head Start Impact Study Spring 2003 teacher survey data codebook*. Washington DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. Retrieved from: https://www.icpsr.umich.edu/.

Salvia, J., Ysseldyke, J., & Bolt, S. (2007). *Assessment in special and inclusive education.* (11th ed.). Belmont, CA: Wadsworth.

Salvia, J., Ysseldyke, J. E., & Witmer, S. (2017). *Assessment in special and inclusive education.* (13th ed.). Boston, MA: Cengage Learning.

SAS. (2013). *Statistical analysis system (version 9.4)*. Cary, NC: SAS Institute Inc. [Software].

Shaffer, J. A., & Postlethwaite, B. E. (2012). A matter of context: A meta-analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Personnel Psychology*, *65*(3), 445–494. doi:10.1111/j.1744-6570.2012.01250.x.

Shaw, D. S., Hyde, L. W., & Brennan, L. M. (2012). Early predictors of boys' antisocial trajectories. *Development and Psychopathology*, *24*(3), 871–888. doi:10.1017/S0954579412000429.

Shonkoff, J. P., & Phillips, D. (2000). *From neurons to neighborhoods: The science of early child development*. Washington, DC: National Academic Press.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.

Spilt, J. L., Hughes, J. N., Wu, J., & Kwok, O. (2012). Dynamics of teacher–student relationships: Stability and change across elementary school and the influence of children's academic success. *Child Development*, *83*(4), 1180–1195. doi:10.1111/j.1467-8624.2012.01761.x.

Stokes, M. E., Davis, C. S., & Koch, G. G. (2001). *Categorical data analysis using the SAS system.* (2nd ed.). Cary, NC: SAS Institute.

Stormont, M., Lewis, T. J., & Beckner, R. (2005). Positive behavior support systems: Applying key features in preschool settings. *TEACHING Exceptional Children*, *37*(6), 42–49. doi:10.1177/004005990503700605.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, *1*(1), 1–24. doi:10.1111/1529-1006.001.

U.S. Department of Health and Human Services (USDHHS). (2010a). Head Start Impact Study: Final report. Retrieved from http://files.eric.ed.gov/fulltext/ED507845.pdf.

U.S. Department of Health and Human Services (USDHHS). (2010b). Head Start Impact Study technical report. Washington DC: Administration for Children and Families. Retrieved from https://www.acf.hhs.gov/sites/default/files/opre/hs_impact_study_tech_rpt.pdf.

U.S. Department of Health and Human Services (USDHHS). (2015). Head Start early learning outcomes framework: Ages birth to five. Retrieved from: https://eclkc.ohs.acf.hhs.gov/school-readiness/article/head-start-early-learning-outcomes-framework.

Wildeboer, A., Thijssen, S., van IJzendoorn, M. H., van der Ende, J., Jaddoe, V. W. V., Verhulst, F. C., … Bakermans-Kranenburg, M. J. (2015). Early childhood aggression trajectories: Associations with teacher-report behaviour. *International Journal of Behavioral Development*, *39*(3), 221–234. doi:10.1177/0165025414562239.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2002). *Woodcock-Johnson III Tests of Achievement*. Itasca, IL: Riverside.

Zayas, V., Whitsett, D. D., Lee, J. J. Y., Wilson, N., & Shoda, Y. (2008). From situation assessment to personality: Building a social-cognitive model of a person. In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *Handbook of personality theory and assessment: Personality measurement and testing.* (Vol. 2; pp. 377–401). Thousand Oaks, CA: SAGE.

Ziv, Y., Alva, S., & Zill, N. (2010). Understanding Head Start children's problem behaviors in the context of arrest or incarceration of household members. *Early Childhood Research Quarterly*, *25*(3), 396–408. doi:10.1016/j.ecresq.2.

AUTHOR BIOGRAPHICAL STATEMENTS

Roland S. Reyes, MS, is a graduate student in the Quantitative Methods program at the University of Pennsylvania, Graduate School of Education.

Paul A. McDermott, PhD, is a Professor of Quantitative Methods at the University of Pennsylvania, Graduate School of Education.

Marley W. Watkins, PhD, is a Nonresident Scholar in the Department of Educational Psychology at Baylor University.

Michael J. Rovine, PhD, is a Senior Fellow in Quantitative Methods at the University of Pennsylvania, Graduate School of Education.

Jessica L. Chao, PhD, is a recent graduate of the Quantitative Methods program at the University of Pennsylvania, Graduate School of Education.