Archives of Assessment Psychology, Vol. 14, No. 1, (1-23)

Preschool Classroom Behavior in Context: Development and Validation of the Spanish Form of the Adjustment Scales for Preschool Intervention (ASPI)

Rebecca Jane Bulotsky-Shearer, Ph.D., Paul A. McDermott, Ph.D., Jenna Futterer, Ph.D., Michael López, Ph.D., and Mileidis Gort, Ed.D.

Abstract

The purpose of the study was to validate the context scale scores for the Spanish language form of the Adjustment Scales for Preschool Intervention (ASPI) for use by early childhood teachers from Spanish-speaking backgrounds. Exploratory and confirmatory factor analyses in a diverse sample of preschool children from low-income families (N = 4,077) revealed two dimensions of context scales of behavioral adjustment: Problems in Preschool Classroom Contexts Requiring Disciplined Behavior and Problems in Contexts Requiring Engaged Behavior. ASPI dimension scores across Spanish and English language forms were horizontally equated through item response theory (IRT). Criterion-related validity using multilevel models was established through concurrent associations with teacher ratings, classroom observations, and direct assessments of social, emotional, behavioral, and academic readiness skills. Policy and practice applications and future directions for the use of context scores are discussed.

Introduction

Early childhood is a critical time when foundational social, emotional, and behavioral skills develop to support children's engagement in early learning experiences that prepare them for school (Thompson & Raikes, 2007). Children living in poverty have access to fewer resources that promote healthy development within home, community, and school settings compared to more socioeconomically advantaged peers (National Academies of Sciences, Engineering, and Medicine, 2019; 2023). Consequently, children living in poverty are at greater risk for displaying social-emotional or behavioral problems. Behavior problems, often called challenging behavior, may look like disruptive and inattentive behavior in early childhood settings. Challenging behaviors make it difficult for children to engage in positive and productive ways within the preschool classroom with peers, teachers, and learning tasks (Graziano et al., 2015; Williford et al., 2018).

Early childhood programs such as Head Start are one of the earliest mechanisms for identification and intervention to address the holistic and healthy development of children living in poverty, including all domains of school readiness such as physical and mental health (Office of Head Start, 2022b). From a public health perspective, early intervention programs like Head Start are naturally embedded opportunities located within communities where children and

families reside. Head Start programs are regulated by the Federal Head Start Performance Standards that mandate a holistic focus on all domains of school readiness, as well as procedures for early screening, identification, and prevention of social-emotional, behavioral, and developmental problems (Boyce et al., 2000; U.S. Department of Health and Human Services, 2023).

Validated early childhood measurement tools are essential to enable early screening and equitable identification of children with social-emotional or behavioral needs within programs like Head Start. The most frequently used screening tools for social-emotional and behavioral needs are teacher reports of classroom behavior (McDermott, 1993). However, many of the most commonly available or widely used rating scales rely on reports of mental health symptoms or behavior ratings for use in clinical interventions and were not developed within the early childhood classroom context (U.S. Department of Health and Human Services, 2001). Early childhood programs need access to validated teacher rating scales to identify and intervene to support children's social-emotional and behavioral adjustment within naturalistic classroom settings. Measures that provide practical and contextual information to help early childhood practitioners identify necessary and effective classroom interventions to support children's social-emotional skills are needed.

Reflecting the changing demographics of our nation, early childhood programs such as Head Start serve culturally, ethnically, and linguistically diverse children and families. Recent federal reports indicate that 35% of the children nationally enrolled in Head Start are from Hispanic backgrounds and 20% of all Head Start children enrolled speak Spanish as their primary home language (Office of Head Start, 2022a). Similarly, the early childhood workforce is diverse, with 24% of Head Start teachers identifying as Hispanic (2022). Among teaching staff, 30% indicated that they were proficient in a language other than English (Office of Head Start, 2022a; Paschall et al., 2020). In local community-based Head Start programs teacher demographics often mirror those of the enrolled children and families. For example, in the Southeastern U.S., in some local programs upwards of 50% of the Head Start teachers are from Spanish-speaking immigrant backgrounds from Central or South America (U.S. DHHS, 2017). The ethnic and linguistic diversity of the early childhood workforce calls for measurement tools that are validated in Spanish and for use with diverse, low-income children and families.

To address this national need, we adapted and translated a teacher rating scale of classroom emotional and behavioral adjustment, the Adjustment Scales for Preschool Intervention (ASPI; Lutz et al., 2002). The ASPI is a contextually focused scale originally developed and validated in English for use in a Northeastern U.S. sample of Head Start children. Using a multistep process of adaptation, translation, and rigorous analysis of the items, we adapted the measure for use by Spanish-speaking early childhood teachers in Head Start programs (see Bulotsky-Shearer et al., 2021 for extensive details of the adaptation and measurement development process).

In the initial study, exploratory and confirmatory factor analysis of the adapted and translated Spanish ASPI items resulted in two higher-order scales (overactive and underactive classroom behavior problems; Bulotsky-Shearer et al., 2021). Both scales showed evidence of reliability and concurrent validity. Given the contextual focus of the ASPI, the purpose of the present study was to extend this initial development work of the Spanish ASPI by examining and validating the underlying factor structure of the context scales that measure behavior problems within 22 classroom contexts in which teachers observe children's classroom behavior (e.g., teacher, peer, or learning contexts). In addition, the present study examined concurrent validity support for the resulting Spanish scale scores using a multisource, multimethod approach of social-

emotional, behavioral, and academic skills.

Developmental and Contextual Classroom Assessment Approach

In a developmental and contextual assessment approach, social-emotional and behavioral problems are observed and interpreted within the developmental demands of routine classroom learning contexts (Downer et al., 2010; Sameroff & Fiese, 2000; Sroufe, 1997). This approach highlights the need to assess children's behavior dynamically within the context in which it is observed (Bronfenbrenner & Morris, 1998; Bulotsky-Shearer et al., 2008; Bulotsky-Shearer et al., 2022). In the classroom, several contexts including social interactions with teachers or peers, as well as instructional learning formats (large and small group activities) require a set of self-regulatory and behavioral skills for children to successfully participate and learn (Williford et al., 2018). For example, structured learning activities, such as teacher-directed whole-group instruction, require children to use self-control, inhibition, and listening skills, and pay attention to engage successfully. On the other hand, less structured free-play time with peers requires children to use emotion regulation, communication, empathy, and other social skills. Applying this developmental-contextual model, assessments examine where and why in the classroom children display social-emotional or behavioral problems, such as disruptive, inattentive, or socially withdrawn behavior.

Initial Development and Validation of the English Form of the ASPI

In response to the need for measurement tools that are sensitive to the ecology of the preschool classroom and measured behavior within learning contexts, the ASPI was developed to assess children within early childhood programs. The English form of the ASPI was developed and validated in partnership with early childhood educators, parents, and professional staff of Head Start. The ASPI requires teachers to observe and record children's behavior within the context of 22 routine developmentally appropriate classroom situations (Lutz et al., 2002).

Initial development and validation studies were conducted in a School District Head Start program in the Northeastern U.S. where children were predominantly African American (70%), with 20% Latinx. In the Northeast Head Start sample, three reliable and valid "context" scales were identified, measuring behavior problems in three routine preschool classroom situations: structured learning activities, peer interactions, and teacher interactions (Bulotsky-Shearer et al., 2008) with α= .84, .81, and .75, respectively (Bulotsky-Shearer et al., 2008). Problems in Peer contexts included behaviors surrounding peers such as respecting others' belongings, handling conflict with peers, and overall behavior in the classroom. Problems in Structured Learning contexts included behavior problems within learning tasks or involvement in learning activities. Problems within Teacher contexts included behavior problems observed during interactions with teachers, such as when greeting teachers in the morning, talking, and general manner with their teacher. Criterion-validity evidence was established using measures of social-emotional, behavioral, and academic skills (Bulotsky-Shearer et al., 2008; Bulotsky-Shearer, Bichay-Awadalla et al., 2020; Bulotsky-Shearer, Fernandez et al., 2020).

Further validation work was completed in the National Head Start Impact Study sample of Head Start children (McDermott et al., 2014). The ASPI was included in the battery of assessment tools used in the national study providing an opportunity for validation of the measure's scales in a national sample. Teachers completed the ASPI longitudinally for children across three years, beginning in preschool and then extending through children's kindergarten and first-grade years. In this longitudinal sample, McDermott and colleagues (2014) conducted a series of psychometric

analyses, resulting in a measure they called the Adjustment Scales for Early Transition in Schooling (ASETS; McDermott et al., 2014). Exploratory and confirmatory factor analysis of the ASETS revealed three situational contexts aligned with those identified in the Northeast School District Head Start sample: Peer Context Problems (10 contexts; α = .91), Learning Context Problems (seven contexts; α= .83), and Teacher Context Problems (five contexts; α = .73). Criterion validity results found ASETS context problem scales were positively associated with parent-reported behavior problems and negatively associated with preschool and first-grade academic skills. Additionally, problems in learning contexts are approximately twice as effective when accounting for children's individual differences in academic skills in comparison with teacher or peer context problems (McDermott et al., 2014).

In addition to the U.S., the ASETS situational context scales have been validated in international samples. In a sample of kindergarten and first-grade children in the Netherlands, the same three context scales were identified and validated: including behavior problems within peer, teacher, and learning contexts (Hamerslag et al., 2017). Reliability analyses showed all three situational dimensions ranged from sufficient to good (α = .69, .76, and .85 for teacher, learning, and peer contexts, respectively). Validity analyses found that while teacher contexts were not associated with children's outcomes, both peer and learning contexts were negatively associated with children's language and numeracy skills (Hamerslag et al., 2017). Findings from prior studies (Hamerslag et al., 2017; McDermott et al., 2014) emphasize the importance of structured learning contexts in children's early academic development within national and international samples, where problems in learning contexts are strongly related to academic outcomes.

In older children, a similar line of work extends to other international samples of elementary school children (ranging from four to 15 years old) with the Adjustment Scales for Children and Adolescents (ASCA) in Trinidad and Tobago (McDermott et al., 2015). Participating children were representative of Trinidad and Tobago, including 39.9% African, 38.8% East Indian, and 21.7% of a mixed race/ethnicity. Exploratory and confirmatory factor analysis yielded three situational contexts: Peer Context Problems (nine contexts; α = .85), Teacher Context Problems, (five contexts; α = .70), and Learning Context Problems (seven contexts; α = .86). Dimensional properties of all three situational problem contexts of the ASCA are comparable to the ASETS, as outlined above. Criterion validity analyses found ASCA problem context scores were associated with teacher-reported disruptive behavior and learning behaviors, as well as parent-reported attention deficit and hyperactivity disorder and conduct problems. For example, Peer and Learning Context Problems were more strongly associated with children's behavior problems, Learning Context Problems showed stronger relationships with children's learning behaviors (McDermott et al., 2015).

Findings from McDermott and colleagues (2014 and 2015) underscore the contribution of problems within situational contexts to children's social-emotional development (e.g., parent-reported behavior problems, and learning behaviors) in national and international samples. Overall, early assessment of problems within children's situational contexts provides critical insight into their academic and social-emotional adjustment and allows researchers to identify strengths that promote children's adjustment and development.

Study Purpose

Although the three context scales of the English form of the ASPI have been validated with the language diversity and preferences of teaching staff working in early childhood community-based programs in the U.S., there is a need to develop and validate the ASPI in Spanish as well.

Therefore, we adapted and translated the teacher report measure into Spanish using best practices in measurement development. See procedures published in (Bulotsky-Shearer et al., 2021), for the approach used, including a multistep, mixed-method approach to develop, adapt, and translate a Spanish language form of the ASPI for use with Spanish-speaking Head Start teachers in the Southeast U.S. To extend this initial research, we examined the factor structure of the context dimensions of the English and Spanish forms through exploratory and confirmatory factor analysis. Using item response theory, we equated the scores for the resultant scales across language forms and examined criterion-related validity of the ASPI context scale scores through concurrent associations with a set of multimethod assessments of social-emotional and academic readiness skills.

We expected to replicate the three context dimensions as identified in the original English ASPI research in both Head Start (Bulotsky-Shearer et al., 2008) and a national sample (McDermott et al., 2014). In addition, based on prior research, we expected that the scales would be concurrently associated with theoretically similar constructs assessed through teacher ratings, classroom observations, and direct assessments. For example, we expected that teacher ratings of positive peer play interactions, emotion regulation, and observed positive engagement with peers and teachers would be negatively associated with behavior problems in ASPI Peer and Teacher contexts. Given prior research, we expected a negative relation between ASPI Problems in Learning context scores and direct assessments of academic skills.

Method

Participants

Children and teachers were recruited from a large urban Head Start program in the Southeast United States. A larger sample was used for exploratory and confirmatory analyses and a smaller subsample of children was selected randomly, and then stratified by age, sex, and ethnicity, from the larger measurement sample, for external criterion validity purposes.

Overall Measurement Sample of Children

Two cohorts participated in the study over a two-year period during the academic years of 2011 - 2012 and 2012 - 2013). The larger sample included 4,077 Head Start children and 161 lead teachers across 61 centers. The primary cohort (N = 4,077) was drawn over the fall semesters of academic years 2011 - 2012 and 2012 - 2013 and a secondary cohort (N = 3,995) was drawn over the Spring semesters of 2011 - 2012 and 2012 - 2013. The primary cohort was used for scale development and validation and the secondary cohort for validation purposes.

The primary cohort included two groups of children: one whose teachers completed the ASPI in English and a second group, whose teachers completed the ASPI in Spanish. The primary cohort included 2,003 children whose teachers used the English ASPI and 2,074 children whose teachers completed the ASPI in Spanish. The secondary cohort of children included 1,940 children for whom the English ASPI was completed and 2,055 children for whom the Spanish form was completed by their teachers.

Of the overall sample of children, 51.2% were female and ranged in age at school entry, from 32 - 60 months (M = 48.77, SD = 6.78). Children were predominantly Hispanic (65.2%) and 63.0% were from Spanish-speaking households. According to parent reports, children were White (63.2%), Black (35.0%), or of other races including Asian, Mixed, or Native Islander (1.8%). Most children were born in the U.S. (96.5%), with 1.7% in Caribbean countries (including Cuba,

Dominican Republic, and Puerto Rico), 0.5% in Central American countries (Guatemala and Nicaragua), 0.4% in Haiti, 0.2% South American countries (Brazil and Chile), and the remainder in other countries (Nigeria, Italy, the Netherlands, and Mexico).

Validity Subsample of Children. Children (n = 642) from 86 classrooms represented the larger sample, ranging in age from 36 - 60 months (M = 47.08, SD = 6.53 months), with 48.0% boys and 43.7% Black, 55.5% Caucasian, and 1.0% multiracial. In this subsample, 56.7% of children were Hispanic, and 55.0% were from Spanish-speaking households.

Teaching Staff. Most teachers (n = 161) were female (99.4%). The majority reported their ethnicity as Hispanic/Latinx (70.8%), and the remainder were White (57.8%), Black (28.0%), and other (13.7%). With regard to the highest education level, 1.2% reported a high school diploma or GED, 12.4% an associate degree, 6.8% a Child Development Associate® credential, 63.4% a bachelor's degree, and 16.1% a master's degree. Teachers reported working as a preschool teacher for 11.41 years.

Most teachers were born outside of the U.S. (69.6%), with 39.1% from Caribbean countries (Cuba, Dominican Republic, and Puerto Rico), 19.9% from South American countries (Colombia, Brazil, Ecuador, Peru, and Venezuela), 7.5% from Central American countries (Honduras, Nicaragua, El Salvador, and Guatemala), 1.9% from Mexico, and 1.9% from other predominantly Hispanic countries. The majority (68.3%) reported speaking Spanish as their first language. Teachers lived in the U.S. for an average of 26.38 years.

Procedures

Approval to conduct this research project was obtained from the University's Institutional Review Board, from the director of the Head Start programs, and from the Head Start program's Parent Policy Council. Consent to participate was obtained from the directors, teachers, and teacher assistants at each center. Once the center and teachers agreed to participate, parental consent was obtained for all participating children, with the assistance of teaching staff.

Data collection procedures were the same for both cohorts of children during the 2011 - 2012 and 2012 - 2013 school years. Measures used for the current study were collected during the fall. Teachers completed the ASPI on all consented children in their classrooms. Child demographic information was collected through center administrative records. Parents and teachers completed a demographic form. The teacher demographic form included several questions related to country of origin, recency of immigration, preferred language for speaking and writing (at home and at work), and preferred language for completing rating forms. Teachers were given ASPI packets to complete in their preferred language.

Validity measures were collected concurrently in the fall for the validity subsample of children (described above). Measures included teacher ratings of peer social competence and emotion regulation, observations of children's classroom engagement, and direct assessments of academic skills. Scores obtained from the Devereux Early Childhood Assessment (DECA; LeBuffe & Naglieri, 1999), Head Start's program-wide mental-health screening tool, were obtained from the program's administrative records.

Direct assessments of academic skills were conducted on children in the subsample using the Learning Express (LE; McDermott et al., 2009) after children were screened for language proficiency in English using the Preschool Language Assessment Scale (PreLAS2000; Duncan & De Avila, 1998; Rainelli et al., 2017). Direct assessments were administered by trained graduate and undergraduate research assistants in a quiet area outside of the classroom.

Children's classroom engagement was individually observed using the Individualized

Classroom Assessment Scoring System (inCLASS; Downer et al., 2010). Trained research assistants completed four 10-minute cycles followed by 5 minutes of real-time coding for each participating child. To become certified observers for the inCLASS, research assistants completed an intensive two-day training by an inCLASS-certified trainer and had to achieve 80% agreement in five video-training master-codes from the University of Virginia's inCLASS team. Interrater reliability was assessed for 20% of the children by double-coding to prevent observer drift.

Criterion-Validity Measures

Peer Social Competence. The Penn Interactive Peer Play Scale (PIPPS-Teacher version; Fantuzzo et al., 1998) is a 32-item rating scale used to measure common play behaviors that promote or interfere with prosocial peer interactions in the classroom. The PIPPS is comprised of three scales: Interactive Play, Disruptive Play, and Disconnected Play (α = .87, .87, and .83, respectively, in the current study sample). External validity is established through direct observations, peer sociometrics, and measures of learning behaviors, temperament, emotion regulation, psychological adjustment, and social skills (Coolahan et al., 2000; Fantuzzo et al., 1998; Mendez et al., 2002).

Emotion Regulation. The Emotion Regulation Checklist (ERC; Shields & Cicchetti, 1997) is a 24-item teaching rating scale comprised of two scales: Emotion Regulation and Lability/Negativity, with α = .81 and .90, respectively (Shields et al., 2001). The Emotion Regulation scale measures a child's ability to regulate positive emotions and affect, show empathy, and display emotional self-awareness. Lability/Negativity measures a child's emotional reactivity (such as mood swings) and the intensity of positive and negative emotions.

Classroom Behavioral Adjustment. The Devereux Early Childhood Assessment (DECA; LeBuffe & Naglieri, 1999) is a 37-item norm-referenced teacher rating scale. It includes four subscales: Initiative, Self-Control, Attachment, and Behavioral Concerns, with high published reliability (α = .90, .90, .85, and .80 respectively). The subscales assess children's ability to take initiative, act independently, control behavior and emotions, form attachments with familiar adults, and communicate their feelings.

Observed Classroom Engagement. The inCLASS (Downer et al., 2010) is a direct observation of individual children's engagement with teachers, peers, and tasks. The inCLASS includes four domains validated for use with Head Start children (Vitiello et al., 2012): Positive Engagement with Teachers, Peers, Tasks, and Conflict Interactions (α = .81, .81, .71, and .72, respectively, in the current sample). Positive associations are documented with measures of peer play, self-regulation, teacher-child relationship quality, behavior problems (Downer et al., 2010) , and language and literacy skills (Bohlman & Downer, 2015).

Academic Readiness Skills. The Learning Express (LE; McDermott et al., 2009) was used to directly assess children's academic skills and includes four scales: Alphabet Knowledge, Vocabulary, Mathematics, and Listening Comprehension, with high internal consistency (α = .98, .96, .96, and .93, respectively). The LE was developed and validated for use with Head Start children. Positive associations are documented with teacher report and direct measures of literacy, vocabulary, mathematics, and science skills (McDermott et al., 2009).

Data Analytic Approach

Analyses for the English and Spanish ASPI forms proceeded in the same way. Factor analysis and IRT equating and calibration each required cross-sectional, mutually exclusive sampling across the two years (i.e., each child appearing only once, averting any within-child

variance sources). Consequently, a computer routine was created that would first identify and include all those children who had been assessed either in 2011 - 2012 or 2012 - 2013 (but not in both years) for the independent English and Spanish portions of the primary cohort, and, second, identify those assessed in both 2011 - 2012 and 2012 - 2013, and randomly include a given child's data from one period only. The resulting English-language equating sample contained 1,863 mutually exclusive members (781 from the fall of 2011, and 1,082 from the fall of 2012) of the primary cohort, and the Spanish-language equating sample contained 1,826 mutually exclusive members (822 from the fall of 2011, and 1,004 from the fall of 2012). The English-language equating sample was randomly partitioned to yield an English-language exploratory subsample (n = 932) and an English-language confirmatory subsample (n = 931). The Spanish-language equating sample was partitioned to form the Spanish-language exploratory subsample (n = 913) and the Spanish-language confirmatory subsample (n = 913).

Exploratory Factor Analyses. Factor analysts have warned of the likely prospect that spurious and unstable factor solutions would emerge whenever categorical or count data are treated as if they were continuous (Bernstein & Teng, 1989; McDonald & Ahlawat, 1974; Mislevy, 1986; Mooijaart, 1983; Muthén, 1987; Waller, 2001). Waller (2001) recommended a viable alternative where categorical data are represented through a polychoric correlation matrix that is smoothed for positive semi-definiteness. We applied two-stage maximum-likelihood estimation (Olsson, 1979) and the respective English- and Spanish-language matrices of 22 behavioral contexts were smoothed for positive semi-definiteness through least-squares approximation of the original matrices (as per Knol & Berger, 1988). The data for each respective language form were submitted for minimum average partialing (MAP; Velicer, 1976) to suggest the number of common factors for retention. The smoothed matrices then were submitted to iterated common (unweighted least squares) factor analysis as advised by Snook and Gorsuch (1989) for instruments containing fewer than 39 variables, with solutions rotated to varimax, equimax, and promax criteria. For each language form, the ideal solution met the following criteria: (a) it approximated simple structure as reflected in maximum hyperplane count (Yates, 1987) and coverage; (b) it had at least four salient behavioral contexts per factor where loadings greater than or equal to .40 indicated salience; (c) it produced internally consistent factors (i.e., $r > .70$); (d) it made theoretical sense in terms of parsimonious coverage of the data and compatibility with prior research (Fabrigar, Wegener, MacCallum, & Strahan, 1999).

Confirmatory Analyses. Employing the respective confirmatory subsamples for the English- and Spanish-language forms, those behavioral contexts that were loading saliently in exploratory analyses were represented in smoothed polychoric matrices. The smoothed polychoric matrices were analyzed through structural equation modeling (SEM) using maximum-likelihood estimation under the Satorra-Bentler scaled difference chi-square for non-normal data (Satorra & Bentler, 2001), seeking acceptable fit where the root-mean-squared error of approximation (RMSEA) was greater than or equal to .06, and the comparative fit index (CFI) was greater than or equal to .95 (Hu & Bentler, 1999).

Equating and Scaling. Final scales were generated by horizontally equating comparable dimensions (i.e., those sharing comparable behavioral contexts) across the language forms. For each qualitatively comparable dimension, linking contexts were identified through multiple-group IRT analysis (Muraki & Bock, 2003) of differential item functioning (DIF). DIF was assessed by χ^2 tests of the residuals (based on expected comparability of context difficulty parameters) for linking contexts across language forms. Behavioral contexts exhibiting statistically significant DIF were precluded as linking contexts. At least one-half the number of contexts comprising a given

dimension were intended as linking contexts to ensure an adequate anchoring of the language forms. Nonequivalent-groups equating was accomplished with the combined English- and Spanish-language equating samples through multiple-group IRT (Muraki & Bock, 2003), using the generalized partial credit model (GPCM; Muraki, 1992). Derived parameters for the behavioral contexts were applied thereafter for the entire primary cohort (all fall assessments) and secondary cohort (all spring assessments), with scores calculated via expected a posteriori (EAP) Bayesian estimation (Thissen & Wainer, 2001), where the scaled score for the English-language form (the reference group) had $M = 50$ and $SD = 10$.

External Validity. Bivariate correlations were used to examine associations between ASPI context scores and concurrent social-emotional, behavioral, and academic readiness skills. Because children were nested within classrooms, to estimate the amount of variance predicted by ASPI scores, a series of hierarchical linear models (Raudenbush et al., 2011) were estimated using SAS 9.3. Each ASPI score (group-mean centered) was used as a predictor in a two-level conditional model, to indicate the percentage of between-children within-teacher-variance accounted for in each outcome by ASPI scores (as per Bliese & Hanges, 2004; Huang, 2018; Waterman et al., 2012).

Results

Dimensionality

MAP for the exploratory analyses with both the English- and Spanish-language forms indicated that 2 factors might be extracted from each of the smoothed correlation matrices. Consequently, 1-3 common factors were extracted for each language form. For both forms, the best solution was the two-factor model, where extraction of a third factor produced an under-identified and unreliable dimension, and extraction of one factor simply collapsed the two viable factors into an uninterpretable dimension. The Goodness-of-Fit Index (GFI; Waller, 2001) for the English-language structure was .991 (Root Mean-Squared Residual [RMSR] = .049) and GFI = .990 and RMSR = .052 for the Spanish-language structure. Per Comrey's (1988) advice, simple structure was enhanced by dropping contexts with salient loadings on both factors for a given solution, with two contexts removed from each English-language dimension and one context removed from both Spanish-language dimensions.

Table 1 displays the dimensional structure for the English-language form and Table 2 the Spanish-language form. The dimensions found for both languages were identically named Indiscipline and Disengagement. The Indiscipline factor included classroom contexts requiring disciplined behaviors. The Disengagement factor included classroom contexts requiring behavioral engagement. Each dimension was adequately reliable with coefficients α for the English-language Indiscipline and Disengagement dimensions being .90 and .78, respectively, and the coefficients for the same dimensions of the Spanish-language form being .87 and .80, respectively.

Confirmatory analyses were based on the saliently loaded markers for each dimension as derived in exploratory analyses. Considering the data for the random confirmatory subsamples, the normalized estimate for multivariate kurtosis was high for both the English-language form (132.91) and the Spanish-language form (221.55). Fit was good for each language model, where for the English-language form the Satorra-Bentler $\chi^2(151) = 514.36$, CFI = .985, RMSEA = .051 (90% CL = .046/.056) and for the Spanish form the Satorra-Bentler $\chi^2(134) = 410.46$, CFI = .985, RMSEA = .048 (90% CL = .042/.053).

Table 1.
Context Dimensional Structure, Properties, and Semantic Equivalence of the English ASPI

Context	Situational context[a]	Factor loadings[b] I	Factor loadings[b] II	h^2	Context/ scale r[c]	Semantic equivalence[d]
Scale I: Problems in Contexts Requiring Disciplined Behavior (Indiscipline); coefficient α = .90[e]						
16	Respecting other people's belongings	*.96* (.87)	-.14 (.45)	.78	.68	
21	Getting along with agemates	*.87* (.86)	-.02 (.52)	.74	.76	*
20	Behaving when standing in line	*.85* (.82)	-.05 (.48)	.68	.67	↔
8	Behaving in the classroom	*.83* (.84)	.02 (.53)	.71	.71	↔
10	Reacting to correction	*.82* (.82)	-.01 (.50)	.67	.69	↔
15	Sitting during teacher directed activities	*.76* (.86)	.17 (.64)	.76	.75	↔
18	Free play/individual choice	*.65* (.78)	.21 (.61)	.64	.70	↔
22	Handling conflicts with other children	*.61* (.70)	.14 (.52)	.51	.62	*
9	Telling the truth	*.55* (.53)	-.04 (.30)	.28	.45	*
12	Coping with new learning tasks	*.54* (.71)	.27 (.61)	.55	.49	
Scale II: Problems in Contexts Requiring Engaged Behavior (Disengagement); coefficient α = .78[e]						
7	General manner with teacher	-.10 (.39)	*.80* (.74)	.55	.55	↔
5	Talking with teacher	-.09 (.38)	*.76* (.70)	.50	.54	*
4	Seeking teacher's help	-.08 (.33)	*.65* (.61)	.37	.39	↔
13	Getting involved in classroom activities	.12 (.52)	*.65* (.73)	.53	.50	↔
3	Answering questions	.12 (.46)	*.55* (.63)	.40	.47	↔
19	Having companions	.10 (.43)	*.53* (.59)	.36	.42	
2	Helping (teacher) with jobs	.22 (.54)	*.52* (.65)	.46	.47	↔
11	Paying attention in the classroom	.37 (.65)	*.45* (.68)	.55	.51	
1	Greeting teacher	.06 (.30)	*.40* (.43)	.19	.35	

Note. N = 932, comprising the random exploratory analysis subsample. [a]Context descriptions are abbreviated for convenient presentation. [b]Nonparenthetical values are promaxian pattern loadings at k = 2, where hyperplane count is maximized. Parenthetical values are structure loadings reflecting the correlation between contexts and latent factors. Salient pattern loadings (>.40) are italicized. [c]Each correlation reflects the relationship between the number of problem behaviors observed within a specific situational context and the sum of problem behaviors observed within all other situational contexts comprising a given scale, where all distributions are standardized to unit-normal form. [d] *indicates a context that has a semantic counterpart on the Spanish-language form of the Adjustment Scales for Preschool Intervention (see Table 2). ↔ indicates a context that has a semantic counterpart on the Spanish-language form *and* that served as a linking context (no Differential Item Functioning) in nonequivalent-groups equating of both language forms. [e]Reliability is based on the full English-language equating sample (N = 1,863), including both exploratory and confirmatory subsamples.

Equating and Scaling

The behavioral contexts associated with each language structure are listed in descending order of pattern and structure loadings for each dimension in Tables 1 and 2. Each context retains its actual ASPI context number (the first column) so that readers may readily find the English and Spanish versions of each context across the tables. The last column in each table indicates whether or not a given context has a semantic counterpart in the opposite language form. Any context displaying the * or ↔ symbol has a semantic counterpart appearing in the corresponding dimension on the opposite language form. Overall, 80.4% of the contexts surviving the exploratory and confirmatory factoring retained a semantic counterpart, with eight of the ten English Indiscipline, eight of the eight Spanish Indiscipline, six of the nine English Disengagement, and six of the eight Spanish Disengagement contexts having counterparts. Focusing then on the behavioral contexts having semantic counterparts, multiple-group DIF analysis identified 5 non-

DIF contexts to link the ten English and nine Spanish Indiscipline contexts (thus linking contexts comprising 50.0% of English and 62.5% of Spanish Indiscipline contexts) and five non-DIF contexts to link the nine English and eight Spanish Disengagement contexts (linking contexts comprising 55.6% of English and 55.6% of Spanish Disengagement contexts). This process generated for equating purposes a final Indiscipline scale containing 13 contexts (five unique to the English-language form, five linking, and three unique to the Spanish-language form) and a Disengagement scale with 14 contexts (four unique English, five linking, five unique Spanish). As an aid to readers, the last columns of Tables 1 and 2 display the ↔ symbol to denote contexts functioning as linking contexts.

Table 2.

Context Dimensional Structure, Properties, and Semantic Equivalence of the Spanish ASPI

Context	Situational context[a]	Factor loadings[b]			Context/ scale r[c]	Semantic equivalence[d]
		I	II	h^2		
Scale I: Problems in Contexts Requiring Disciplined Behavior (Indiscipline); coefficient α = .87[e]						
21	Se lleva con otros niños	*.99* (.87)	-.21 (.37)	.78	.69	*
20	Comportamiento en la fila	*.86* (.84)	-.07 (.44)	.68	.62	↔
10	Reacciona a correcciones	*.82* (.81)	-.01 (.47)	.65	.67	↔
8	Comportamiento en el salón de clase	*.77* (.86)	.15 (.61)	.74	.68	*
15	Comportamiento en actividades con maestra	*.68* (.79)	.19 (.59)	.65	.67	↔
22	Maneja conflictos con otros niños	*.66* (.72)	.10 (.49)	.53	.60	*
9	Es sincero	*.61* (.56)	-.09 (.28)	.32	.38	↔
18	Juego libre o actividades escogidas por el	*.59* (.78)	.33 (.67)	.67	.64	↔
21	Se lleva con otros niños	*.99* (.87)	-.21 (.37)	.78	.69	*
20	Comportamiento en la fila	*.86* (.84)	-.07 (.44)	.68	.62	↔
Scale II: Problems in Contexts Requiring Engaged Behavior (Disengagement); coefficient α = .80[e]						
13	Involucra en actividades del salón	.01 (.49)	*.82* (.83)	.67	.61	↔
2	Ayuda con labores	.02 (.47)	*.77* (.78)	.60	.50	↔
5	Habla con maestro/a	-.05 (.38)	*.74* (.71)	.50	.51	*
4	Pide ayuda	-.27 (.13)	*.69* (.53)	.33	.28	↔
7	Comportamiento general con maestro/a	.06 (.44)	*.65* (.69)	.47	.54	↔
12	Afronta nuevas tareas de aprendizaje	.21 (.57)	*.62* (.74)	.58	.51	
17	Participa en los juegos con otros niños	.38 (.69)	*.52* (.74)	.65	.50	
14	Involucra en las labores manuales	.35 (.64)	*.51* (.71)	.58	.57	
3	Contesta las preguntas	.14 (.40)	*.42* (.50)	.27	.34	↔

Note. N = 913 comprising the random exploratory analysis subsample. [a]Context descriptions are abbreviated for convenient presentation. [b]Nonparenthetical values are promaxian pattern loadings at k = 2, where hyperplane count is maximized. Parenthetical values are structure loadings reflecting the correlation between contexts and latent factors. Salient pattern loadings (>.40) are italicized. [c]Each correlation reflects the relationship between the number of problem behaviors observed within a specific situational context and the sum of problem behaviors observed within all other situational contexts comprising a given scale, where all distributions are standardized to unit-normal form. [d]*indicates a context that has a semantic counterpart on the English-language form of the Adjustment Scales for Preschool Intervention (see Table 1). ↔ indicates a context that has a semantic counterpart on the English-language form *and* that served as a linking context (no Differential Item Functioning) in nonequivalent-groups equating of both language forms. [e]Reliability is based on the full Spanish-language equating sample (N = 1,826), including both exploratory and confirmatory subsamples.

Multiple-group IRT was applied for the GPCM model using the combined English- and Spanish-language equating samples. For Indiscipline the slope parameters ranged from 0.59-1.84 (M = 1.05, SD = 0.28 for the English-language form and M = 1.13, SD = 0.27 for the Spanish-language form), whereas Disengagement slopes ranged from 0.67 - 1.33 (M = 0.89, SD = 0.25

for the English form and $M = 1.01$, $SD = 0.23$ for the Spanish form). Total scale information (the inverse of measurement error for Indiscipline was 8.18 and maximum information = 4.69 at θ = 2.07 and total scale information for Disengagement = 6.31 with maximum information = 2.59 at θ = 1.52.

Figure 1(a).

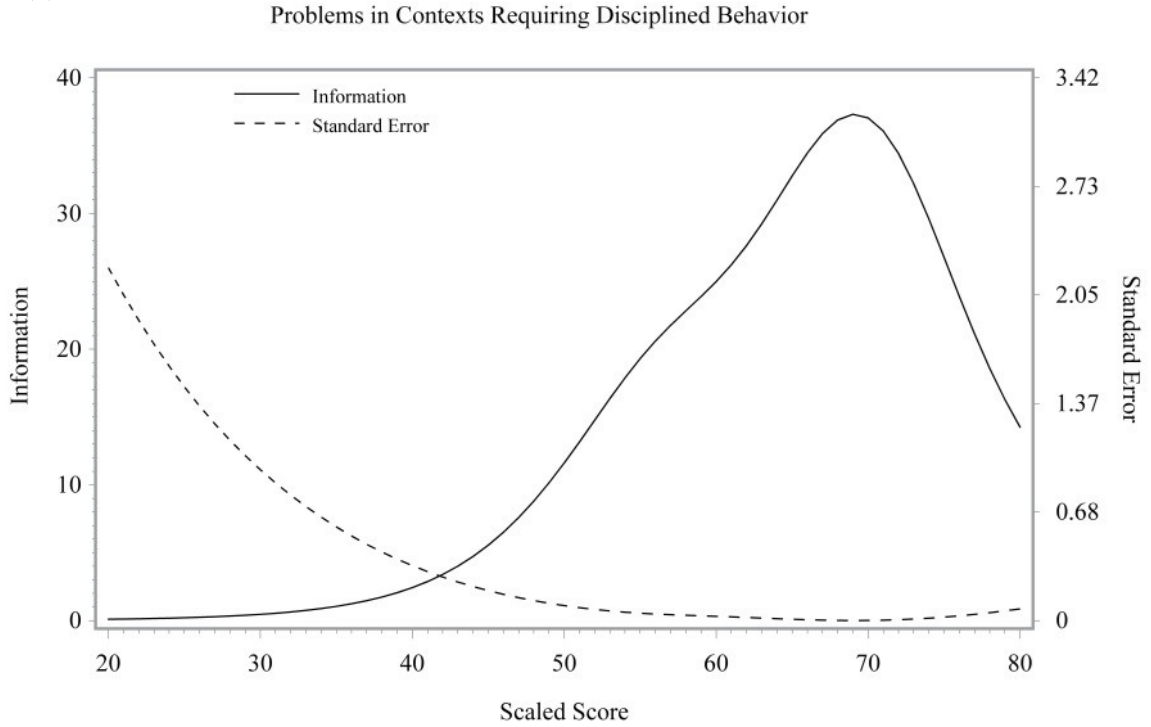
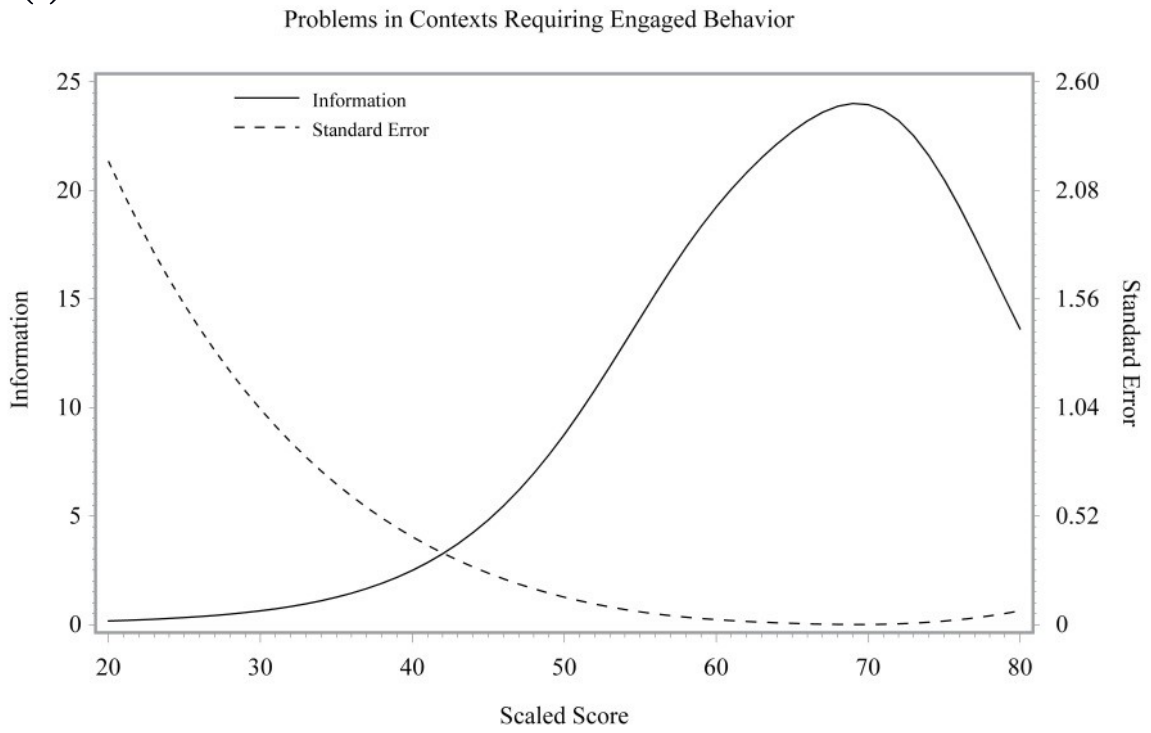


Figure 1(b).



EAP scaled scores were estimated for members of the equating sample (the English-language sample serving as the reference group with population $M = 50.0$ and $SD = 10.0$). Sample scaled scores for English-language participants reported the Indiscipline $M = 49.8$ ($SD = 9.4$), whereas Spanish-language participants had an Indiscipline $M = 45.8$ ($SD = 7.6$). Similarly, sample scaled scores for English-language participants had a Disengagement $M = 50.1$ ($SD = 8.1$), while Spanish-language participants had a Disengagement $M = 46.8$ ($SD = 7.3$). The disparities between English- and Spanish scores are notable, indicating that teachers who use the Spanish-language ASPI form are markedly less inclined to observe and/or report problem behaviors.

Figure 1 illustrates for each equated scale the overlap of total scale information and measurement error. Scores will have practical utility from approximately three-quarters of a standard deviation below the population mean and throughout the highest scaled scores. This is particularly useful since ASPI scaled scores are often used to discriminate classroom behavior that is adjusted or adequately adjusted (< 60), versus at risk (> 60 and < 70), versus maladjusted (> 70).

External Validity

Table 3 displays concurrent associations between ASPI context scores and school readiness measures. All statistically significant correlations were in the expected direction with ASPI scores showing stronger associations with teacher rating measures of social-emotional and behavioral skills, and lower associations with most direct observations of classroom engagement and direct assessments of academic skills. Table 3 presents the multilevel modeling results, indicating the percentage of criterion measure variance that reflects children's actual individual differences, and the parenthetical values indicate how much of that variance is accounted for by each ASPI scale. For example, Table 3's last column entry for PIPPS disruptive scale indicates that 60.7% of the score variance stems from children's individual differences (rather than teacher characteristics), and 48.2% of that variance is predictable from children's ASPI Indiscipline scores and 17.1% is predictable from ASPI Disengagement scores.

Discussion

The purpose of this study was to expand the availability of validated Spanish language teacher-rating scales of preschool classroom behavior by examining the psychometric evidence for a recently adapted and translated measure, the ASPI. Exploratory and confirmatory analyses provided evidence for two underlying factors: Behavior Problems in Contexts Requiring Disciplined Behavior, and Behavior Problems in Contexts Requiring Engaged Behavior. We found that children's scores on both context scales were associated concurrently with social-emotional, behavioral, and academic skills using a multisource, multimethod measurement approach. Following, we contextualize the findings within the broader literature and include implications for the use of Spanish measures of classroom behavior in early childhood settings, such as Head Start.

Exploratory and confirmatory factor analyses revealed a reliable, two-factor structure across both English and Spanish forms. The two scales included Behavior Problems in Contexts Requiring Disciplined Behavior and Behavior Problems in Contexts Requiring Engaged Behavior. The first scale measured preschool behavior problems in classroom contexts that require self-regulation of behavior, including behaving in structured learning contexts (e.g., standing in line, sitting during teacher-directed activities), less structured contexts with peers (e.g., free play, handling conflicts with peers) and dealing with frustrating or challenging learning situations (e.g., reaction to teacher correction, coping with new learning tasks). The second scale included behavior problems in classroom situations that require children's initiation and engagement with teachers

and learning tasks, such as talking to the teacher, seeking the teacher's help, getting involved in classroom activities, having friends, helping the teacher with jobs, greeting teacher, answering questions, and paying attention in the classroom.

Table 3.

Associations between fall ASPI Context Scores and Concurrent Social and Academic Measures

Context/Domain	*n*	ASPI dimensions [a]		% Explainable variance[b]
		Indiscipline	Disengagement	
Interactive Play	744	.37** (13.9)	.41** (23.8)	57.9
Disruptive Play	745	.56** (48.2)	.34** (17.1)	60.7
Disconnected Play	742	.45** (16.0)	.50** (26.1)	56.1
Devereux Early Childhood Assessment				
Initiative	3431	.22** (9.9)	.30** (16.1)	65.0
Attachment		.12** (8.5)	.15** (10.8)	53.9
Self-control		.39** (28.5)	.30** (16.8)	62.7
Behavior Concerns		.41** (24.7)	.30** (13.7)	57.0
Emotion Regulation Checklist				
Emotion Regulation	671	.22** (11.6)	.25** (18.4)	41.3
Lability/Negativity	669	.49** (37.2)	.32** (10.3)	58.4
Learning Express				
Vocabulary	699	.03 (0.0)	.06 (0.1)	98.7
Mathematics		.10* (2.2)	.20** (3.4)	96.5
Listening comprehension		.10* (0.8)	.15** (3.4)	97.1
Alphabet knowledge		.09 (0.2)	.16** (2.3)	95.6
InCLASS				
Teacher positive engagement	521	.00 (1.4)	.01 (0.3)	83.9
Peer positive engagement		.07 (0.9)	.19** (4.5)	92.8
Task positive engagement		.19** (3.4)	.22** (7.0)	82.1
Conflict engagement		.29** (9.7)	.10 (5.8)	96.9

Note. $^{**}p < .001$. $^{*}p < .01$. [a] Nonparenthetical entries are Pearson product moment correlations. Parenthetical entries indicate the percentage of variance in the respective criterion measure scores between children within classrooms that is accounted for by each ASPI scale score. Values equal the proportional reduction in the residual variance (100) as estimated via hierarchical linear modeling. Each two-level random coefficients model entered a given ASPI scale as the covariate. ASPI= Adjustment Scales for Preschool Intervention. InCLASS=Individualized Classroom Scoring System. [b]Total percentage of potentially explainable variance between children within classrooms. Values equal 1-intraclass correlation (100), estimated via hierarchical linear modeling. Each two-level, unconditional means model applied random intercepts for classrooms, where the random effect was significant at $p < .001$.

Contrary to our expectations, we validated a two-factor structure, rather than a three-factor structure as was previously identified using the English version of the ASPI in local and national samples (Bulotsky-Shearer et al., 2008; McDermott et al., 2014). For example, in the national Head Start Impact Study sample, McDermott and colleagues (2014) found a three-factor structure for the context scales (Behavior Problems in Structured Learning Tasks, Teacher Contexts, and Peer

Contexts). However, other studies conducted with the ASPI in international samples have found a similar two-factor structure. Prior research conducted in Trinidad and Tobago, using a parent-reported version of the ASCA (the elementary school version of the ASPI) validated two comparable "situtypic" dimensions (Drogalis et al., 2017).

It is unclear why, in our local Head Start sample, we were not able to validate a three-factor solution. The third factor that was extracted in the dimensional analyses did not meet the criteria for retention. In part, the difference in factor structure could be due to differences in the way that teachers interpreted or responded to the items. Perhaps, it was attributable to cultural and language diversity, or English literacy of the teacher respondents in our sample, many of whom were first-generation immigrants from Latin or Central America and the Caribbean. Future studies are needed to disentangle some of the issues related to item endorsement that might relate to teacher demographics and other background characteristics.

External Validity

We found evidence for concurrent validity, with associations between the two ASPI context scale scores and children's scores on a multi-method, multi-source battery of school readiness measures assessed in the fall of the preschool year. Aligned with our hypotheses and prior research, associations were strongest for measures assessing social-emotional constructs, such as peer play interactions, emotional lability, and observed behavior (Bulotsky-Shearer et al., 2008, 2021). Indiscipline scores were positively associated with teacher-reported disruptive and disconnected play, behavior concerns, emotional lability, and observed conflict within the classroom. Children whose teachers rated them with higher disruptive and conflictual behaviors within the context of peer and teacher interactions on the ASPI also scored higher on measures of emotional lability and reactivity, general classroom behavior concerns, and greater peer play disruption in the classroom. Children's Disengagement scores were positively associated with teacher-reported disconnected play with peers, overall classroom behavior concerns, and emotional lability. Additionally, ASPI Disengagement scores were negatively associated with academic skills such as mathematics, listening comprehension, and alphabet knowledge.

Findings align with prior early childhood research. For example, disruptive or externalizing behavior within social contexts, such as with peers and teachers, has been found to interfere with positive peer and teacher relationships and to be associated with higher ratings of peer play disruption and emotional lability in the classroom (Bulotsky-Shearer, Fernandez, et al., 2020). In our study, ASPI Indiscipline scores were negatively associated with mathematics and language skills, mirroring previous research showing negative relationships between preschool problem behaviors and mathematics (Doctoroff et al., 2016) and language skills (Bulotsky-Shearer et al., 2008). ASPI Disengagement scores were negatively associated with mathematics, language, and literacy skills, suggesting socially withdrawn behaviors in preschool may influence children's positive engagement within learning situations that foster skill development across multiple academic domains (Bulotsky-Shearer et al., 2008; Lonigan et al., 2016; Sims et al., 2016).

Limitations and Future Directions

While this study extends prior research by validating the Spanish form of the ASPI and contributes to the availability of early childhood measurement tools for use in programs by teaching staff who prefer to use Spanish assessments, there are limitations of note. First, the current study focused on initial concurrent validity evidence for the ASPI scores and is limited to the fall of the children's Head Start year. Future research could employ a longitudinal approach including multiple time points to examine the trajectories of children's classroom behavior problems, as was conducted by McDermott et al. (2019) in the national Head Start Impact Study sample. Examining predictors of variability in children's behavioral trajectories over time, as well as social-emotional and academic outcomes associated with these trajectories in preschool and in the transition to kindergarten, would be important next steps.

Additionally, future studies could extend this initial variable-centered study by applying a person-centered model, such as latent profile analysis or growth mixture modeling. Applying a person-centered approach would enable us to examine whether there are certain subgroups of children who display similar patterns of behavior problems, and examine predictors and outcomes associated with subgroups of children (Howard & Hoffman, 2018). This approach would help in education settings to identify children who may display relative risk for poor school adjustment and could benefit from tailored interventions by early childhood teachers within the preschool classroom setting (McDermott et al., 2019).

Finally, while our study examined relationships between classroom behavior problems and child outcomes, future research could examine other factors influencing children's behavior, such as classroom quality or family engagement. For example, previous research identifies classroom quality as an important influence on children's behavior and school readiness (Bulotsky-Shearer, Fernandez, et al., 2020). Family engagement and parents' participation in children's learning also promote positive social-emotional and academic outcomes, especially in Spanish-speaking families (McWayne et al., 2016). Researchers can extend the current study using the ASPI to examine the relationship between children's behavior and important contextual factors, such as classroom quality and family engagement, in Spanish-speaking communities.

Implications for Mental Health Policy and Assessment Practices

Implementing appropriate mental health and social-emotional screening measures within early childhood programs is one of the greatest challenges facing programs nationally, particularly for larger urban programs that serve culturally and linguistically diverse families. Often, teaching staff working within community-based programs such as Head Start are from similar backgrounds and communities as the children and families participating in the program. There is therefore a need for programs and teaching staff to access Spanish-language measures that all staff can use comfortably. Our study extends the availability of such tools by validating the context scales of the Spanish form of the ASPI. Programs can use these scores to help identify children in need of additional observation, assessment, and early intervention support within the classroom.

Importantly, the ASPI was developed to reflect a developmental-contextual assessment approach, which focuses on observations of children's behavior within the context of developmentally appropriate preschool classroom contexts (Bulotsky-Shearer et al., 2008; Lutz et al., 2002). Scores can be used by mental health support specialists within early childhood programs to identify what behavior is of concern, and in what context. For example, mental health consultants (in collaboration with the teacher) can use the contextual scores on the ASPI in conjunction with other data, such as direct observation of children's behavioral engagement within

the contexts of teachers, peers, and learning tasks (e.g., inCLASS; Downer et al., 2010), to identify appropriate strategies that match a child's needs within specific contexts. Strategies could be selected that best support the child in developing skills to be more successful within classroom contexts, such as in large-group time, or in unstructured play with peers, such as implemented through the Learning to Objectively Observe Kids consultation model (LOOK; Downer et al., 2018; Williford et al., 2018).

About the Authors

Rebecca Jane Bulotsky-Shearer, Ph.D., is a Professor in the Department of Psychology Child Division at the University of Miami in Coral Gables, FL (https://orcid.org/0000-0002-7974-0487). Correspondence concerning this article should be addressed to Rebecca J. Bulotsky-Shearer, Department of Psychology, University of Miami, Flipse Building, 5665 Ponce de Leon Blvd, Coral Gables, FL 33146. Email: rshearer(at)miami.edu.

Paul A. McDermott, Ph.D. is a Professor in the Human Development and Quantitative Methods Division at the Graduate School of Education of the University of Pennsylvania in Philadelphia, PA.

Michael Lopez, Ph.D., is the Vice President of Education and Child Development with NORC at the University of Chicago in Bethesda, MD.

Mileidis Gort, Ed.D., is the Associate Dean of Students and a Professor of Equity, Bilingualism and Biliteracy at the School of Education of the University of Colorado in Boulder, CO (https://orcid.org/0000-0002-9581-567X).

Jenna N. Futterer, Ph.D., is a Research Analyst with the County of Orange Social Services Agency in Orange, CA.

Acknowledgements

This research project was funded by the U.S. Department of Education, Institute for Education Sciences, Measurement Development Research Grant, Early Childhood Programs and Policies (R305A1100233), awarded to the first, second, fourth, and fifth authors.

A very special thank you to the Miami-Dade County Human Services Action Agency Head Start/Early Head Start Program for their collaboration in this project.

References

Bernstein, I. H., & Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence for multidimensionality due to item categorization. *Psychological Bulletin*, *105*(3), 467-477. https://doi.org/10.1037/0033-2909.105.3.467

Bliese, P. D., & Hanges, O. J. (2004). Being both too liberal and too conservative: The results of treating grouped data as though they were independent. *Organizational Research Methods, 7*(4), 400-417. https://doi.org/10.1177/1094428104268542

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. https://doi.org/10.1037/0012-1649.44.1.139

Bulotsky-Shearer, R. J., López, L. M., & Mendez, J. L. (2016). The validity of interactive peer play competencies for Latino preschool children from low-income households. *Early Childhood Research Quarterly, 34*(1), 78-91. https://doi.org/10.1016/j.ecresq.2015.09.002

Bulotsky-Shearer, R., Bichay-Awadalla, K., Bailey, J., Futterer, J., & Huaqing Qi, C. (2020). Teacher-child interaction quality buffers negative associations between challenging behaviors in preschool classroom contexts and language and literacy skills. *Topics in Early Childhood Special Education, 40(3),* 159-171. https://doi.org/10.1177/0271121420947155

Bulotsky-Shearer, R. J., Fernandez, V., Bichay-Awadalla, K., Bailey, J., Futterer, J., & Huaqing Qi, C. (2020). Teacher-child interaction quality moderates social risks associated with problem behavior in preschool classroom contexts. *Journal of Applied Developmental Psychology, 67, 101103.* https://doi.org/10.1016/j.appdev.2019.101103

Bulotsky-Shearer, R., Carter, T. M., Williford, A.P., Alamos, P., & Hasbrouck, S. (2022). Making the invisible visible: Using a contextual assessment approach to identify children with social-emotional and behavioral needs in preschool classrooms. *Topics in Early Childhood Special Education, 42*(4), 344-356. https://doi.org/10.1177%2F02711214221111396

Bulotsky-Shearer, R. J., McDermott, P. A., Lopez, M., Gort, M., Bouza, J., Fernandez, V., Bichay-Awadalla, K. (2021). Development and initial validation of the Spanish form of the adjustment scales for preschool intervention (ASPI). *Journal of School Psychology, 84*, 124-142. https://doi.org/10.1016/j.jsp.2020.11.003

Boyce, C. A., Hoagwood, K., Lopez, M. L., & Tarullo, L. B. (2000). The Head Start mental health research consortium: New directions for research partnerships. *Behavioral Disorders*, *26*(1), 7–12. https://doi.org/10.1177/019874290002600101

Bronfenbrenner, U. & Morris, P.A. (1998). The ecology of developmental processes. In W. Damon & R. M. Lerner (Eds.) *Handbook of child psychology: Theoretical models of human development* (5th edition). Wiley & Sons.

Castro, M., Mendez, J. L., & Fantuzzo, J. (2002). A validation study of the Penn interactive peer play scale with urban Hispanic and African American preschool children. *School Psychology Quarterly, 17*, 109–127. https://doi.org/10.1521/scpq.17.2.109.20856

Comrey, A. L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology*, *56*(5), 754-761. https://doi.org/10.1037/0022-006X.56.5.754

Coolahan, K., Fantuzzo, J., Mendez, J., & McDermott, P. (2000). Preschool peer interactions and readiness to learn: Relationships between classroom peer play and learning behaviors and conduct. *Journal of Educational Psychology, 92*(3), 458–465. https://doi.org/10.1037/0022-0663.92.3.458

Doctoroff, G. L., Fisher, P. H., Burrows, B. M., & Edman, M. T. (2016). Preschool children's interest, social–emotional skills, and emergent mathematics skills. *Psychology in the Schools, 53*(4), 390-403. https://doi.org/10.1002/pits.21912

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. https://doi.org/10.1016/j.ecresq.2009.08.004

Drogalis, A. R., McDermott, P. A., Watkins, M. W., Chao, J. L., Worrell, F. C., & Hall, T. E. (2017). Parent and teacher perspectives on psychological adjustment: A national measurement study in Trinidad and Tobago. *International Journal of School and Educational Psychology, 5*(2), 74-87. http://dx.doi.org/10.1080/21683603.2016.1191398

Duncan, S. E., & De Avila, E. A. (1998). *Pre-Las 2000*. McGraw-Hill.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272-299. https://doi.org/10.1037/1082-989X.4.3.272

Fantuzzo, J., Coolahan, K., Mendez, J., McDermott, P., & Sutton-Smith, B. (1998). Contextually-relevant validation of peer play constructs with African American Head Start children: Penn Interactive Peer Play Scale. *Early Childhood Research Quarterly*, *13*(3), 411-431. https://doi.org/10.1016/S0885-2006(99)80048-9

Graziano, P., Slavez, J., Ros, R., Garb, L., Hart, K., & Garcia, A. (2015). Self-regulation assessment among preschoolers with externalizing behavior problems. *Psychological Assessment, 27*(4), 1337–1348. https://doi.org/10.1037/pas0000113

Hamerslag, R., Oostdam, R., & Tavecchio, L. (2017). Inside school readiness: The role of socioemotional and behavioral factors in relation to school, teachers, peers and academic outcome in kindergarten and first-grade. *European Early Childhood Education Research Journal, 26*(1), 80-96. https://doi.org/10.1080/1350293X.2018.1412035

Howard, M. C., & Hoffman, M. E. (2018). Variable-centered, person-centered, and person-specific approaches: Where theory meets the method. *Organizational Research Methods*, *21*(4), 846-876. https://doi.org/10.1177/1094428117744021

Hu., L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Huang, F. L. (2018). Multilevel modeling myths. *School Psychology Quarterly, 33*(3), 492–499. https://doi.org/10.1037/spq0000272

Knol, D. L., & Berger, M. P. F. (1991). Empirical comparison between factor analysis and multidimensional item response models. *Multivariate Behavioral Research*, *26*(3), 457-477. https://doi.org/10.1207/s15327906mbr2603_5

LeBuffe, P. A., & Naglieri, J. A. (1999). *The Devereux Early Childhood Assessment*. Kaplan Press.

Lonigan, C. J., Lerner, M. D., Goodrich, J. M., Farrington, A. L., & Allan, D. M. (2016). Executive function of Spanish-speaking language-minority preschoolers: Structure and relations with early literacy skills and behavioral outcomes. *Journal of Experimental Child Psychology*, *144*(1), 46-65. https://doi.org/10.1016/j.jecp.2015.11.003

Lutz, M. N., Fantuzzo, J., & McDermott, P. (2002). Multidimensional assessment of emotional and behavioral adjustment problems of low-income preschool children: Development and initial validation. *Early Childhood Research Quarterly*, *17*(3), 338-355. https://doi.org/10.1016/S0885-2006(02)00168-0

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. https://doi.org/10.1016/j.jsp.2009.07.002

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. https://doi.org/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. https://doi.org/10.1016/j.ecresq.2018.08.010

McDermott, P. A., Watkins, M. W., Rhoad, A. M., Chao, J. L., Worrell, F. C., & Hall, T. E. (2015). Trinidad and Tobago national standardization of the Adjustment Scales for Children and Adolescents. *International Journal of School & Educational Psychology, 3*(4), 278-292. https://doi.org/10.1080/21683603.2015.1067873

McDermott, P. A. (1993). National standardization of uniform multisituational measures of child and adolescent behavior pathology. *Psychological Assessment, 5*, 413-424.

McDonald, R. P., & Ahlawat, K. S. (1974). Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology*, *27*(1), 82-99. https://doi.org/10.1111/j.2044-8317.1974.tb00530.x

McWayne, C. M., Melzi, G., Limlingan, M. C., & Schick, A. (2016). Ecocultural patterns of family engagement among low-income Latino families of preschool children. *Developmental Psychology, 52*(7), 1088-1102. https://doi.org/10.1037/a0040343

Mendez, J. L., McDermott, P., & Fantuzzo, J. (2002). Identifying and promoting social competence with African American preschool children: Developmental and contextual considerations. *Psychology in the Schools*, *39*(1), 111-123. https://doi.org/10.1002/pits.10039

Mislevy, R. J. (1986). Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics*, *11*(1), 3-31. https://doi.org/10.3102/10769986011001003

Mooijaart, A. (1983). Two kinds of factor analysis for ordered categorical variables. *Multivariate Behavioral Research*, *18*(4), 423-441. https://doi.org/10.1207/s15327906mbr1804_5

Muraki, E., & Bock, D. (2003). *PARSCALE* (Ver. 4.1) [Computer program]. Scientific Software International.

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

Muthén, B. (1987). *LISCOMP: Analysis of linear structural equations with a comprehensive measurement model*. Scientific Software International.

National Academies of Sciences, Engineering, and Medicine. 2019. *Vibrant and healthy kids: Aligning science, practice, and policy to advance health equity.* The National Academies Press. https://doi.org/10.17226/25466.

National Academies of Sciences, Engineering, and Medicine. (2023). *Closing the opportunity gap for young children.* The National Academies Press. https://doi.org/10.17226/26743.

Office of Head Start (2022a). *Head Start services snapshot (2021-2022)*. https://eclkc.ohs.acf.hhs.gov/sites/default/files/pdf/no-search/service-snapshot-hs-2021-2022.pdf

Office of Head Start (2022b). *Head Start Early Learning Outcomes Framework*. https://eclkc.ohs.acf.hhs.gov/school-readiness/article/head-start-early-learning-outcomes-framework

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Pychometrika*, *44*, 443-460. https://doi.org/10.1007/BF02296207

Paschall, K., Madill, R., & Halle, T. (2020). *Demographic characteristics of the early care and education workforce: Comparisons with child and community characteristics.* OPRE Report #2020-108. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/sites/default/files/documents/opre/demographic-characteristics-ECE-dec-2020.pdf

Puma, M., Bell, S., Cook, R., Heid, C., & Lopez, M. (2005). Head Start impact study: First year findings. *Administration for Children & Families*.

Rainelli, S., Bulotsky-Shearer, R., Fernandez, V., Lopez, M., & Greenfield, D. (2017). Examining the validity of the PreLAS2000 subtests as a language routing procedure for low-income, bilingual Spanish-speaking preschool children. *Early Childhood Research Quarterly*, *38*(1), 10-22. https://doi.org/10.1016/j.ecresq.2016.08.001

Raudenbush, S. W., Bryk, A. S., Congdon, R. (2011). *Hierarchical linear and nonlinear modeling (version 7).* Scientific Software International.

Sameroff, A., & Fiese, B. H. (2000). Models of development and developmental risk. In C. H. J. Zeanah (Ed.), *Handbook of infant mental health* (2nd ed., pp. 3-19). Guilford Press.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*, 507-514. https://doi.org/10.1007/BF02296192

Shields, A., & Cicchetti, D. (1997). Emotion regulation among school-age children: The development and validation of a new criterion Q-sort scale. *Developmental Psychology, 33*(6), 906–916. https://doi.org/10.1037/0012-1649.33.6.906

Shields, A., Dickstein, S., Seifer, R., Giusti, L., Dodge Magee, K., & Spritz, B. (2001). Emotional competence and early school adjustment: A study of preschoolers at risk. *Early Education and Development*, *12*(1), 73-96. https://doi.org/10.1207/s15566935eed1201_5

Sims, D. M., Purpura, D. J., & Lonigan, C. J. (2016). The relation between inattentive and hyperactive/impulsive behaviors and early mathematics skills. *Journal of Attention Disorders*, *20*(8), 704-714. https://doi.org/10.1177/1087054712464390

Snook, S. C., & Gorsuch, R. L. (1989). Component analysis versus common factor analysis: A Monte Carlo study. *Psychological Bulletin*, *106*(1), 148-154. https://doi.org/10.1037/0033-2909.106.1.148

Sroufe, L. A. (1997). Psychopathology as an outcome of development. *Development & Psychopathology, 9*, 251-268.

Thissen, D., & Wainer, H. (2001). *Test scoring*. Erlbaum.

Thompson, R. A., & Raikes, H. A. (2007). The social and emotional foundations of school readiness. In D. F. Perry, R. K. Kaufmann, & J. Knitzer (Eds.), *Social and emotional health in early childhood: Building bridges between services and systems* (pp. 13–35). Brookes.

U.S. Department of Health and Human Services. (2017). *Office of Head Start Program Information Report – Head Start Program Facts: Fiscal Year 2017.* https://eclkc.ohs.acf.hhs.gov/sites/default/files/pdf/hs-program-fact-sheet-2017_0.pdf

U.S. Department of Health and Human Services. (2023). *Head Start Program Performance Standards*, 45 CFR Chapter XIII Part 1304, *Federal Register, 61,* 57186-57227. Washington, DC: U.S GPO. https://eclkc.ohs.acf.hhs.gov/policy/45-cfr-chap-xiii

U.S. Department of Health and Human Services. (2001). *Report of the surgeon General's conference on children's mental health: A national action agenda*. www.surgeongeneral.gov/cmh/

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*, 321-327. https://doi.org/10.1007/BF02293557

Vogel, C., Aikens, N., Atkins-Burnett, S., Martin, E. S., Caspe, M., Sprachman, S., & Love, J. M. (2008). *Reliability and validity of child outcome measures with culturally and linguistically diverse preschoolers: the first 5 LA universal preschool child outcomes study spring 2007 pilot study.* Mathematica Policy Research, Inc.

Waller, N. G. (2001). *MicroFACT 2.0: A microcomputer factor analysis program for ordered polytomous data and mainframe size problems*. Assessment Systems.

Waterman, C., McDermott, P. A., Fantuzzo, J. W., & Gadsden, V. L. (2012). The matter of assessor variance in early childhood education—Or whose score is it anyway? *Early Childhood Research Quarterly*, *27*(1), 46-54. https://doi.org/10.1016/j.ecresq.2011.06.003

Williford, A. P., Bulotsky-Shearer, R., Bichay, K., Reilly, S., & Downer, J. T. (2018). Adapting assessments of child engagement to develop an early childhood consultation model. *Journal of Applied Developmental Psychology*, *56*, 67-78. https://doi.org/10.1016/j.appdev.2018.01.005

Yates, A. (1987). *Multivariate exploratory data analysis: A prospective on exploratory factor analysis*. State University of New York Press.