Early Childhood Research Quarterly 29 (2014) 255–267

Contents lists available at ScienceDirect

Early Childhood Research Quarterly

Informing context and change in young children's sociobehavioral development – The national Adjustment Scales for Early Transition in Schooling (ASETS)☆

Paul A. McDermott[a,*], Marley W. Watkins[b], Michael J. Rovine[c], Samuel H. Rikoon[d]

[a] University of Pennsylvania, United States
[b] Baylor University, United States
[c] Pennsylvania State University, United States
[d] Educational Testing Service, United States

ARTICLE INFO

Article history:
Received 23 January 2013
Received in revised form 9 February 2014
Accepted 14 February 2014

Keywords:
Early childhood
Head Start
Behavioral adjustment
Context
Item response theory
Transition

ABSTRACT

This article recounts the design and validity evidence for contextually-specific measures of early childhood social and behavioral adjustment within school using the Adjustment Scales for Early Transition in Schooling (ASETS). Through primary analyses of data from the Head Start Impact Study, a representative nationwide sample ($N = 3077$) of randomly selected children from low-income families was used to inform developmental-transitional stability and change in adjustment across numerous school contexts. Longitudinal exploratory and confirmatory factor analyses yielded reliable and temporally continuous behavioral dimensions assessing the pervasiveness of Peer, Learning, and Teacher Context Problems. Each context dimension was equated vertically through IRT, with Bayesian scoring across two years spanning prekindergarten through 1st grade. Multilevel modeling provided support for the concurrent validity of ASETS contextual scales and their ability to assess future risk of academic and behavioral problems. ASETS scales are also shown to reveal differential, contextually-based, change trajectories across four years of early school transition.

1. Introduction

The past decade has witnessed a phenomenal increase in policy initiatives and research centering on the socio-emotional needs of young children (Campbell, 2001; Egger & Angold, 2006; President's New Freedom Commission on Mental Health, 2003; Rescorla et al., 2007, 2011). Motivation stems largely from the observation that prevalence rates for preschool emotional and behavioral problems approach 20% (Egger & Angold, 2006), with early and untreated problems undermining critical developmental processes and portending more serious and sometimes intractable disorders at later ages (Campbell & James, 2007; Feeney-Kettler, Kratochwill, & Kettler, 2011; Kataoka, Zhang, & Wells, 2002). One understandable response has been a variety of assessment devices to identify and differentiate manifestations of preschool socio-emotional distress (Campbell & James, 2007; Feeney-Kettler et al., 2011; Rescorla et al., 2011), the intention being to clarify the distinct nature of problems in such a way that might lead to preventative or restorative intervention.

Most contemporary instruments for assessment of early emotional and behavioral problems embrace a common formulation. Since young children, given their social, conceptual, and linguistic immaturity, and limited perspectives, are unable to report accurately the relevant symptomatology and incidence of their own distress (Fulmer & Frijters, 2009; Moll & Tomasello, 2012; Norwood, 2007; von Baeyer, Forsyth, Stanford, Watson, & Chambers, 2009), informed adult observers (teachers, parents) are typically asked to respond to rating scales or questionnaires that survey the child's reactions at home or school. The best examples include the Preschool and Kindergarten Behavior Scales (Merrell, 2003), the Devereux Early Childhood Assessment (LeBuffe & Naglieri, 1999), the Behavior Assessment System for Children (Reynolds &

☆ This research was conducted with the cooperation of the U.S. Department of Health and Human Services, Administration for Children and Families, and supported in part by the U.S. Department of Education, Institute of Education Sciences (Grant No. R305C050041-05). The views expressed represent the views of the authors, and not of the Educational Testing Service.

* Corresponding author at: School of Education, University of Pennsylvania, 3700 Walnut Street, Philadelphia, PA 19104-6216, United States.
E-mail addresses: drpaul4@verizon.net (P.A. McDermott), rikoon@upenn.edu (S.H. Rikoon).

http://dx.doi.org/10.1016/j.ecresq.2014.02.004
0885-2006/

Kamphaus, 2004), and the Achenbach System of Empirically Based Assessment (Achenbach & Rescorla, 2000). Thus, teachers in the classroom or parents in the home portray a child's adjustment by indicating the presence or frequency of numerous specific symptoms, where the symptoms are manifest either through observable child behaviors or perceived child emotions. In turn, researchers apply factor-analytic procedures to the resultant teacher or parent responses and thereby discover that different symptoms tend to group together and reveal common surface syndromes or dimensions that more or less resemble traditional clinical psychiatric disorders. In this way, a given child can be assessed by a teacher at school and the child's current socio-emotional adjustment may be quantified for a variety of different types of problems (aggression, withdrawal, etc.).

Although such instrumentation has been immensely useful for advancing understanding of early childhood socio-emotional manifestations, and arguably for informing pertinent intervention, it has been our view that most contemporary instruments do not rest on firm theoretical grounds and do not take advantage of the technical capacities available for design and application of survey instruments. Specifically, it is our position that popular rating scales and questionnaires are designed with little or no attention to the distinct contextual frameworks within the school (or home) or to the signature transitional nature of emotional and behavioral problems as children develop. In the introductory section of this article, we discuss the theoretical import of context and transition for studying early socio-emotional adjustment. We then demonstrate through development and application of a new national measure, how context and transition play a central role in advancing understanding of early childhood adjustment.

1.1. Context theory

In the assessment domain, context theory is probably best represented by the work of Mischel (2004). Mischel points to a key factor guiding the construction of traditional measures of personality and adjustment—the assumption that a given score level on a particular trait dimension (e.g., withdrawal) translates to a given disposition for that type of emotion or behavior. But as research and experience dictate, people sharing the same score level on a trait or dimension will, in reality, display a substantial range of dispositions for the anticipated emotions or behaviors. In practice, this makes traditional assessments less accurate and consequently less useful. Alternatively, Mischel, Shoda, and Mendoza-Denton (2002) illustrate how the actual disposition for a given emotion or behavior depends not only on the trait or dimensional score level but on the contextual circumstances wherein the emotions or behaviors are embedded. Thus, children with a high level of a withdrawal trait will not react uniformly in different contexts, such as when approached by a teacher versus when involved in group play versus when confronted by challenging learning tasks. Indeed, knowledge of the situations that give rise to problem behaviors is fundamental to understanding the motivations behind problems and the accurate prediction of future incidence (Zayas, Whitsett, Lee, Wilson, & Shoda, 2008). Traditional instruments tend to ignore contextual frameworks by regarding situational variation as some sort of "noise" or error (Mischel et al., 2002). Such instruments may feature items that either inquire about trait behavior without any reference to specific situations under which it may or may not emerge, or otherwise average scores across all sorts of situations to produce a general composite on trait behavior.

The contextual view is entirely consistent with the developmental–ecological perspective advocated by Sabol and Pianta (2012) for studying contexts that differentially influence teacher–child relationships; by Zayas et al. (2008) and Kagan (2003) who show the role of context for explaining intra-individual variations in behavioral dispositions; by Mian, Wainwright, Briggs-Gowan, and Carter (2011) and Thorsen, Goldberg, Osann, and Spence (2008) who focus on specific situations that invite good versus bad reactions; and by Sameroff (2010) and Bronfenbrenner and Morris (2006), who offer more unified theories to bind natural individual child propensities and contextual frameworks in the broader story of human development. The idea that contextual specificity makes a difference is also supported by emergent empirical literature demonstrating that: (a) young children's withdrawal and emotional regulation vary as a function of classroom context (Buss, 2011; Goldsmith & Davidson, 2004); (b) peer-group contexts affect the aggressiveness of children with special needs (Visser, Kunnen, & van Geert, 2010) and preschool language acquisition (Justice, Petscher, Schatschneider, & Mashburn, 2011); (c) manipulation of classroom structural aspects and learning locations can abate problem behavior (Kern & Clemens, 2007; Wannarka & Ruhl, 2008); (d) alternation of individual and group activities and the amount of teacher involvement affects child classroom engagement (Powell, Burchinal, File, & Kontos, 2008); and (e) planned free-time and classroom transitions affect behavior (Joosten, Bundy, & Einfeld, 2012).

1.2. Transition theory

Early childhood transition theory emerges from the work of Entwisle and Alexander (1993) and Entwisle, Alexander, and Olson, (2005), with seminal connections to Piagetian and Eriksonian concepts of stage theory. Transition theory essentially argues that children's developmental status is multifaceted and constantly changing in response to ontogenetic and environmental influences. It holds that children's more or less successful adaptations to those influences set the template for future capacities to adapt and that, as pertains to long-term acquisition of coping mechanisms and cognitive achievements, the most critical developmental periods are those proximate to major transitions. As researchers point out, such transition periods in early childhood education include movement into and through preschool and progression into regular kindergarten and finally first grade (Benner & Crosnoe, 2011; Buss, 2011; Goldsmith & Davidson, 2004; Hemmeter & Ostrosky, 2006; Pianta, Cox, & Snow, 2007; von Suchodoletz, Trommsdorff, Heikamp, Wieber, & Gollwitzer, 2009).

Thus, whereas prekindergarten entry will often provide a child's first exposure to part- or full day schedules organized around group meals and naps and individual or companion play, kindergarten and first-grade activities begin to supplant discovery learning with more deliberate and structured activities emphasizing group-centered common curricula that encourage self-reliance and competition. Eventually, desks replace play circles, vocality becomes imperative, literacy becomes fundamental to what is transpiring in the classroom, and academic failure or retention become real prospects. Consequently, the causal centrality of early school transition to long-term child development has essentially risen to a meta-theoretical level that regards early transitions in schooling as a major developmental milestone (Eivers, Brendgen, & Borge, 2010).

1.3. Innovative instrumentation

The instrumentation and methods for assessing the contextual nature of children's school socio-behavioral adjustment was first suggested by Stott (1966) and implemented fully in the development and national standardization of the Adjustment Scales for Children and Adolescents (ASCA; McDermott, 1993; McDermott, Steinberg, & Angelo, 2006). In contrast to the traditional format of classroom rating scales, where teachers are asked to respond to lists of symptomatic behaviors by indicating their general intensity or frequency, the ASCA presents 122 problem behavior items embedded in 22 different contexts involving peer interactions, teacher, learning tasks, group activity, organized games, and free play. The teacher indicates whether each problem behavior typifies the child's behavior over the past month, where a set of multiple possible behaviors appears within each context. Also unlike traditional rating scales, ASCA embeds one or two healthy or commonplace behaviors within each context. Thus, teachers are provided alternative normal behavioral variants as a measure to avert response bias induced by exclusive presentation of problem behavior choices. The nature and severity of children's adjustment problems are measured in two ways: First, the problem behaviors are grouped by factor analyses to reveal multiple dimensions whose component behaviors indicate common surface syndromes (such as oppositionality, diffidence, and avoidance). These dimensions are called *phenotypes* because they embody behaviors that share similar appearance and function, although unlike other rating scales, they actually reflect specific types of problematic behaviors that are pervasive across multiple different contexts. In this way, maladjustment is defined by its generality across school contexts rather that its emergence in isolated circumstances (the latter phenomenon actually suggesting a transient reactive problem rather than a more general maladjustment; Horn, Wagner, & Ialongo, 1989). Second, the contexts themselves are grouped by factor analyses to reveal the situations that give rise to the phenotypic problem behavior. These contextual dimensions are called *situtypes* and indicate whether problems emerge in contexts involving the teacher, classmates, or structured learning.

The ASCA is designed for children in kindergarten through 12th grade. As a means to accommodate preschool children, ASCA was revised and validated for application with Head Start children (Bulotsky-Shearer, Fantuzzo, & McDermott, 2008; Noone-Lutz, Fantuzzo, & McDermott, 2002) and named the Adjustment Scales for Preschool Intervention (ASPI). As with ASCA, ASPI yields scores on phenotype dimensions that indicate *what* types of problem behavior exist and situtype dimensions that inform *when and where* they exist, thus availing important clues as to motivation and potential intervention. More recently, the technology to assess children's transition from preschool into formal schooling became available. All of the original ASPI items and their unique contextual formats were administered intact for the national Head Start Impact Study (U.S. Department of Health and Human Services [DHHS], 2010a), spanning two years of preschool through kindergarten and first grade. Because the national sample reached horizontally well beyond Head Start into other public and private preschools and vertically through kindergarten and first grade, the instrument was renamed the Adjustment Scales for Early Transition in Schooling (ASETS). McDermott, Watkins, Rovine, and Rikoon (2013) resolved new phenotype dimensions for ASETS, calibrating them through item-response theory (IRT) and validating them for longitudinal assessment. As yet, contextually-based situtype dimensions have not been established or validated for the new national instrument.

This article reports the longitudinal factor analyses, IRT scaling and scoring, and concurrent and predictive validity of situtype dimensions of problem behavior assessed by ASETS. It further demonstrates via multilevel modeling the utility of the contextual dimensions for assessing relative risk of preschool children's later academic nonproficiency, for detecting change in adjustment levels over the long transition from early preschool into formal schooling, and for identifying the signature developmental trajectories that earmark eventually successful versus unsuccessful learners.

2. Method

2.1. Participants

The Head Start Impact Study (HSIS; DHHS, 2010b) was a nationwide randomized control trial designed to determine the relative effectiveness of Head Start and comparable prekindergarten programs. Participants were drawn randomly from 223 prekindergarten agencies across all geographic regions of the United States, provided that each child was eligible for Head Start entry (essentially a family income below or close to the federal poverty level). The youngest children were enrolled in prekindergarten in academic year 2002–2003 (AY0203) and followed through AY0506. ASETS was completed by each child's classroom teacher at the end of the first year of prekindergarten (PreK 1) and second prekindergarten year (PreK 2), the kindergarten year (K), and first-grade year (1st grade).

Because not all children randomly selected for prekindergarten entry actually entered school for PreK 1, and because others did not enter prekindergarten settings that would provide a teacher and/or classroom-type environs, the national sample size increased as children moved from PreK 1 to 1st grade (i.e., PreK 1 $N=1377$, PreK 2 $N=2764$, K $N=2873$, 1st grade $N=3077$). Considering the ASETS *full national* sample ($N=3077$), M age at entry to the study was 4.0 years ($SD=.5$), with 49.6% of children being females, 37.8% Hispanic, 29.5% African American, 32.7% White or other race/ethnicity, 25.7% primarily Spanish-speaking at entry, 12.8% identified with special needs, and 82.7% residing in urban areas. During PreK 1, children attended 540 preschool centers (867 classrooms) and during PreK 2 1032 centers (1815 classrooms), while during K, children attended 1469 schools (2280 classrooms) and during 1st grade 1617 schools (2576 classrooms). Through PreK years, as much as 80% of classrooms were not associated with conventional schools (approximately 60% being part-day environs such as day care or other non-school centers), with about 90% of post-PreK classrooms affiliated with public schools. Detailed sample characteristics are reported by McDermott et al. (2013) and DHHS (2010a, 2010b).

2.2. Instrumentation: sociobehavioral adjustment

Table 1 presents descriptive statistics for all instrumentation used in the current study by developmental level. ASETS contains 134 behavioral indicators embedded in 22 situational contexts. Each indicator may be checked or not by the responding teacher to describe the child's behavior over the past month. The 22 contexts cover relationships with the teacher and with other children, coping with classroom expectations, and demeanor during games and play. A typical context inquires, "How does the child react to correction?" Within that context, the teacher may describe child behavior by endorsing one of more of the following indicators: "Improves for the moment but it does not last long," "Accepts correction without fuss," "Takes correction badly (sulky, muttering, etc.)," "Answers back aggressively, makes threats, or creates a disturbance," and "Cries." In this manner, each context presents three to seven relatively negative problem behavior indicators that are theoretically and empirically reflective of a potential surface syndrome or phenotype (Aggression, Attention Seeking, Reticence/Withdrawal, or Low Energy; see McDermott et al., 2013) and most contexts provide one or two positive or healthy behavior choices. Altogether, ASETS features 112 problem and 22 healthy indicators where, given the intended purpose of the instrument, the number of problem indicators endorsed by the teacher for a given context comprises that context's score for problem behavior (score ranges varying from 0–3 to 0–7, depending on the number of problem behaviors embedded in different contexts). Moreover, all contextual and indicator

Table 1
Descriptive statistics for instrumentation by developmental level.

		Prekindergarten 1		Prekindergarten 2		Kindergarten		First Grade	
Instrument and contexts/subscales	# of indicators	*N*	*M* (*SD*)	*N*	*M* (*SD*)	*N*	*M* (*SD*)	*N*	*M* (*SD*)
Adjustment Scales for Early Transition in Schooling (contexts[a])		1377		2764		2873		3077	
Greeting the teacher	5		.5 (.6)		.4 (.6)		.3 (.5)		.3 (.5)
Helping the teacher with jobs	4		.3 (.5)		.2 (.5)		.2 (.5)		.2 (.4)
Answering questions	4		.5 (.6)		.5 (.6)		.5 (.6)		.5 (.6)
Seeking teacher's help	4		.2 (.4)		.2 (.4)		.2 (.5)		.3 (.5)
Talking with the teacher	5		.4 (.6)		.3 (.5)		.4 (.6)		.4 (.6)
Valuing the teacher's attention	4		.6 (.6)		.6 (.6)		.6 (.6)		.5 (.6)
General manner with the teacher	8		.6 (.8)		.6 (.8)		.5 (.8)		.5 (.8)
Behaving in the classroom	4		.5 (.7)		.4 (.7)		.5 (.7)		.5 (.7)
Telling the truth	3		.3 (.6)		.4 (.6)		.4 (.6)		.4 (.7)
Reacting to correction	4		.6 (.8)		.5 (.8)		.5 (.8)		.5 (.8)
Paying attention in the classroom	4		.6 (.7)		.4 (.6)		.5 (.7)		.6 (.8)
Coping with new learning tasks	3		.3 (.5)		.3 (.5)		.3 (.5)		.3 (.5)
Getting involved in class activity	4		.4 (.7)		.3 (.6)		.3 (.6)		.3 (.6)
Working with hands (artwork, etc.)	7		.5 (.8)		.4 (.8)		.5 (.9)		.5 (.9)
Sitting during directed activities	6		.7 (.9)		.6 (.9)		.6 (.0)		.6 (1.0)
Respecting others' belongings	4		.3 (.7)		.3 (.6)		.2 (0.6)		.2 (.7)
Taking part in games with others	7		.5 (.8)		.4 (.8)		.4 (.8)		.4 (.8)
Free play (self-choice activity)	8		.8 (1.0)		.7 (1.0)		.6 (.9)		.6 (.9)
Having companions	4		.4 (.6)		.3 (.5)		.3 (.5)		.2 (.5)
Behaving while standing in line	3		.5 (.6)		.4 (.6)		.4 (.6)		.4 (.6)
Getting along with agemates	6		.5 (.8)		.5 (.9)		.5 (.9)		.5 (1.0)
Handling conflicts with others	4		.5 (.8)		.4 (.7)		.4 (.7)		.4 (.7)
Pianta Child–Teacher Relationships Scale									
Closeness	7			2747	31.0 (4.0)			3058	29.7 (4.6)
Conflict	8			2743	13.7 (6.3)			3050	14.2 (6.9)
Positive Relationship	15							3059	63.3 (9.7)
Parent rating scale	14								
Total Behavior Problems				2626	5.5 (3.6)			3059	4.9 (3.9)
Peabody Picture Vocabulary Test, Third Edition				2699	297.0 (37.5)			2900	359.6 (30.5)
Woodcock–Johnson III Test of Achievement									
Letter Word Identification				2700	330.0 (27.9)				
Applied Problems				2683	399.3 (22.4)				
Pre-Academic Skills				2683	368.1 (21.0)				
Spelling				2701	374.8 (25.5)				
Basic Reading Skills								2873	449.8 (32.3)
Word Attack								2875	467.7 (31.0)
Quantitative Concepts								2877	461.3 (17.3)
Mathematics Reasoning								2879	457.9 (17.2)
Language and Literacy Ability	2							2188	.7 (.5)
Mathematics Ability	2							2182	.8 (.4)
Social Science Ability	2							2177	.8 (.4)

[a] Context descriptions are abbreviated for convenient presentation.

language is distinctly behavioral, avoiding clinical jargon or necessity for speculation about unobservable internal processes such as children's thoughts or feelings.

2.3. Instrumentation: external validity measures

ASETS scores were validated against several teacher and parent ratings and direct assessments. Results are reported here for two developmental levels (PreK 2 and 1st grade), although available for all levels. Only a representative presentation of results was feasible, given the limitations of space. It was thought appropriate to present results from the culminating point of the PreK levels (PreK 2) and post-PreK levels (1st grade) because they allowed the maximum amount of participant data where psychometric properties for the external measures were acceptable. Certain measures administered for HSIS were eliminated from the current study because at a given developmental level, they failed to produce sufficient data to yield reasonable statistical power, or because the original instruments were altered for HSIS without report of requisite psychometric support (Smith, McCarthy, & Anderson, 2000), or because they failed to produce minimally adequate reliability (viz., ≥.70 as recommended by Fabrigar, Wegener, MacCallum, & Strahan, 1999 and Nunnally, 1978) for the HSIS population (DHHS, 2010b, pp. 3.32–3.43).

Teacher ratings. The Pianta Student–Teacher Relationships Scale (Pianta, 1996) features 15 items, such as "This child easily becomes angry at me," rated on a 5-point scale ranging from 1 = "Definitely does not apply" to 5 = "Definitely applies." Three subscales are available: Closeness (7 items), Conflict (8 items), and Total Positive Relationship (15 items). Substantial concurrent and predictive validity evidence is provided (Pianta, 2001; Pianta & Stuhlman, 2004) and internal consistency for the relevant HSIS developmental levels ranged between .73 and .82 for Closeness, .76 and .89 for Conflict, and .88 and .89 for Total Positive Relationship (DHHS, 2010b). Teacher report of Academic Ability is rated at the close of 1st grade for Language and Literacy, Mathematics, and Social Science, as based on observed attainment of multiple skills compared to the attainment of peers (DHHS, 2010b). Performance is rated as either 0 = "Below Average" (nonproficient) or 1 = "Proficient." Since each measure is a single index, internal consistency estimates are infeasible. Rather, the appropriate standard error of the *M* is reported here, where SE_M for Language and Literacy = 008, Mathematics = .008, and Social Science = .007. Given the discrete scaling, subsequent statistical analyses apply logit link functions and the Bernoulli response distribution.

Parent ratings. A parent rated each child's aggressive or defiant, hyperactive, and withdrawn or depressed behavior using the Total Behavior Problems scale. The scale contained 14 dichotomous items, such as "Is disobedient at home" and "Feels worthless or inferior." Development and validity evidence are provided for the FACES national study (DHHS, 2001, p. 2.27) and for HSIS in DHHS (2010b). Additional validity evidence was reported by Vaden-Kiernan et al. (2010) and Ziv, Alva, and Zill (2010). For the HSIS PreK 2 and 1st grade samples, internal consistency ranged .78–.79. For the Head Start and kindergarten samples as reported for the FACES national study, internal consistency ranged .76– .80.

Direct assessments. The Peabody Picture Vocabulary Test, Third Edition (PPVT; Dunn, Dunn, & Dunn, 1997) assesses receptive vocabulary. Criterion validity evidence is abundant (Dumont & Willis, 2006; Salvia, Ysseldyke, & Bolt, 2007) and reliability indexes ranged .70–.78 for the HSIS population. Also, various subscales of the Woodcock–Johnson III Tests of Achievement (WJ; Woodcock, McGrew, & Mather, 2002) were administered. The present study used scores from Letter-Word Identification (letter and word reading skills; HSIS internal consistency ranging .90–.91), Applied Problems (solving practical math problems by recognizing process and counting or calculating; internal consistency = 89–.90), Spelling (writing letters and words; internal consistency = .78–.81), Pre-Academic Skills cluster (composite of the three preceding subscales; internal consistency = .76–.78), Word Attack (applying phonic and structural analysis skills; internal consistency = .93–.94), Basic Reading Skills cluster (composite of the three Pre-Academic Skills cluster subscales and Word Attack; internal consistency = .90–.91), Quantitative Concepts (identifying number concepts and recognizing patterns and missing aspects; internal consistency = 86–.87), and Mathematics Reasoning cluster (composite of Applied Problems and Quantitative Concepts; internal consistency = 71–.78). Ample validity support has been reported for the WJ achievement subscales (Dumont & Willis, 2006; Salvia et al., 2007).

2.4. Procedure

Teachers responded to ASETS contexts in the spring semesters of AY0203–AY0506. The various criterion measures were administered during the same semesters, with PPVT and WJ scales applied by trained technicians and teacher reports on child academic ability provided at the close of 1st grade (DHHS, 2010a, 2010b). The average number of children assessed per classroom during AY0203 = 1.59, AY0304 = 1.52, AY0405 = 1.26, and AY0506 = 1.08. Factor analyses and vertical scale equating each required a mutually-exclusive cross-sectional sampling across PreK 1, PreK 2, K, and 1st grade. To this end, we consecutively drew at random one child from each developmental level (no child being drawn twice) until a sample of 1600 children was constructed, with 400 different children representing each level. The total size of this sample was dictated by the necessity to (1) generate for statistical power purposes the largest sample possible without redundant membership and (2) represent each developmental level equally. This sample was termed the *calibration sample*. This sample was important because it was used in subsequent longitudinal factor analyses to ensure that each developmental age group was represented equally, providing sufficient statistical power for subsequent confirmatory analyses investigating the longitudinal invariance of the factor structure (see below). This also precluded any sources of within-child variance in the calibration sample, as required for vertical equating and derivation of scoring parameters. Thereafter, the calibration sample of 1600 was randomly bifurcated to form an exploratory factor analysis (EFA) subsample of 800 and confirmatory factor analysis (CFA) subsample of 800.

Factor analyses. Inasmuch as the point scales for the 22 contexts were essentially ordinal, it was necessary to treat each score distribution as categorical. Numerous researchers have demonstrated the spurious factors and unstable dimensions that can arise when categorical data are treated as continuous data in EFA (Bernstein & Teng, 1989; McDonald & Ahlawat, 1974; Mislevy, 1986; Mooijaart, 1983; Muthén, 1987; Waller, 2001). Alternatively, per Waller's (2001) recommendation, we applied iterative common factoring with a smoothed polychoric matrix. Specifically, two-stage maximum-likelihood estimation (Olsson, 1979) was used to produce an initial polychoric correlation matrix among the 22 context scores for the EFA subsample, and the matrix was smoothed for positive semidefiniteness through least-squares approximation of the original matrix (Knol & Berger, 1991). The smoothed matrix was submitted for minimum average partialling (MAP; Garrido, Abad, & Ponsoda, 2011; Velicer, 1976) to suggest the number of factors for retention and thereafter submitted to iterative common factoring with varimax, equamax, and promax rotation. The ideal structure was that which satisfied multiple criteria; namely, the solution must (a) approximate simple structure as reflected in maximum hyperplane count (Yates, 1987) and coverage of contexts, (b) have at least four salient loadings per factor where loadings ≥.40 indicate salience, (c) produce internally consistent factors (i.e., $r \geq .70$), and (d) make theoretical sense in terms of parsimonious coverage of the data and compatibility with leading research in the area (Fabrigar et al., 1999).

The factor structure based on the salient markers from the ideal EFA solution were analyzed for the CFA subsample using maximum-likelihood estimation under the Satorra–Bentler scaled difference chi-square for nonnormal data (Satorra & Bentler, 2001), seeking acceptable fit where the Root Mean Squared Error of Approximation (RMSEA) ≤ .08 and Comparative Fit Index (CFI) ≥ .90 (Marsh, Liem, Martin, Morin, & Nagengast, 2011).

Scaling. Factor dimensions were scaled vertically, joining PreK 1 to PreK 2, PreK 2 to K, and K to 1st grade. For each dimension, a number of contexts were identified as linking contexts through multiple-group IRT analysis (Muraki & Bock, 2003) of Differential Item Functioning (DIF). A linking context is one that appears simultaneously at two adjacent developmental levels. DIF was assessed through χ^2 tests of the residuals (based on expected comparability of context difficulty parameters) for linking contexts across adjacent developmental levels (e.g., PreK 1 versus PreK 2). Contexts displaying statistically significant DIF were dismissed as potential linking contexts. One-third the number of contexts comprising a dimension for a given level were selected as linking contexts, being chosen so as to best distribute linking contexts across difficulty levels covering the dimension's distribution. Vertical equating was accomplished with the longitudinal calibration sample (N = 1600) using multiple-group IRT (Program PARSCALE; Muraki & Bock, 2003) testing both the Generalized Partial Credit Model (GPCM; Muraki, 1992) and the Graded Response Model (GRM; Samejima, 1996). Resultant item parameters were applied for the ASETS full national sample (N = 3077), with scores calculated via expected a posteriori (EAP) estimation (Thissen & Wainer, 2001), where the scaled score (*SS*) M = 50 and SD = 10 at Pre K 1, the reference level.

External validity. All validity analyses were performed using available data from the full national sample. Product–moment correlations were computed to show the direction and strength of relationships between ASETS dimension scores and scores for the external validity measures. Given the volume of data across developmental levels, reporting is limited to the most representative levels; i.e., PreK 2 (culminating the PreK period) and 1st grade (culminating the post-PreK period). Since the data were nested within teachers/classrooms, relationships also were assessed using hierarchical linear modeling (HLM), where each ASETS dimension served as the group-mean centered predictor in a two-level conditional HLM model, revealing the percentage of between-children within-teacher/classroom variance accounted for by ASETS variance (Waterman, McDermott, Fantuzzo, & Gadsden, 2012). Model specification was Criterion $\text{Measure}_{ij} = \gamma_{00} + \gamma_{10}\text{Context Problems}_i + \mu_{0j} + \mu_{1j} + r_{ij}$.

Predictive validity was examined for the relative risk of end-of-1st-Grade academic nonproficiency for teacher-reported Language and Literacy, Mathematics, and Social Science. Because of the binary nature of those reports and the nesting within teachers, two-level, generalized multilevel logistic models were constructed with teachers as the cluster variable, ASETS dimensions as explanatory variables, and teacher-reported binary outcomes as the response variables. Each model applied generalized multilevel linear modeling with the logit link function, Bernoulli response distribution, and adaptive quadratures to estimate the integral. These models were used to derive odds ratios indicating the increased risk for subsequent nonproficiency associated with each increment in ASETS *SS*s. Model specification was $\text{Academic Ability}_{\eta[\text{logit}]ij} = \gamma_{00} + \gamma_{10}\text{Context Problems}_i + \mu_{0j}$.

Change detection. A major element in any transition study is the sensitivity of the measurement to real change over time. Using the full national sample, each ASETS dimension was entered into a three-level, unconditional growth model, where level 1 estimated temporal variability within children over the four developmental levels, level 2 estimated variability between children, and the third level teacher/classroom variability. Models assessed random coefficients for intercepts and slopes, as well as linear, quadratic, and cubic trends for change. Model specification was $\text{Context Problems}_{ijk} = \gamma_{000} + \gamma_{100}\text{Time}_{ijk} + \gamma_{200}\text{Time}^2_{ijk} + \gamma_{300}\text{Time}^3_{ijk} + \gamma_{010}\text{Male}_j + \gamma_{020}\text{Black}_j + \gamma_{030}\text{Hispanic}_j + \gamma_{040}\text{Urban}_j + \gamma_{050}\text{LanguageStatus}_j + \gamma_{060}\text{SpecialNeeds}_j + (\mu_{00k} + \mu_{10k}\text{Time}_{ijk}) + (\mu_{0jk} + \mu_{1jk}\text{Time}_{ijk}) + r_{ijk}$.

3. Results

3.1. Dimensionality

MAP suggested that a minimum of two factors and scree suggested as many as three factors might be extracted from the smoothed polychoric matrix. Thus, 1- through 4-factor models were tested against the stated criteria. Having satisfied all criteria, the 3-factor, promaxian ($k = 2$) model was selected as optimal, where the Goodness of Fit Index = .996 and Root Mean Squared Residual = .036 (per Waller, 2001). The optimal solution retained 21 of the 22 contexts (compared to the 19 contexts retained for the former ASPI Head Start solution; Bulotsky-Shearer et al., 2008). Models extracting an additional factor produced underidentified and unreliable dimensions and those extracting fewer essentially excluded viable dimensions. Specifically, in contrast to the requisite criteria for an ideal factor solition, the 4-factor model produced three contexts with salient loadings on multiple factors, two factors that had no reasonable counterparts in established research (Bulotsky-Shearer et al., 2008; McDermott et al., 2006), one factor with only three salient loadings, and one factor with internal consistency < .70. Alternatively, the 2-factor model collapsed two reliable and traditionally recognized factors from the ideal 3-factor model to form a single and markedly less interpretable dimension, thus violating the recommendations outlined by Fabrigar et al. (1999) and Wood, Tataryn, and Gorsuch (1996) preferring extraction of the maximum number of meaningful latent dimensions.

Table 2 presents rotated pattern loadings, final communalities, and context/total scale *r*s. Based on contextual content and the patterns of descending loadings, the scales were named Peer Context Problems (10 contexts), Learning Context Problems (seven contexts), and Teacher Context Problems (5 contexts). Over the four developmental levels, Peer Context Problems and Learning

Table 2
Dimensional structure for context problems of the adjustment scales for early transition in schooling.

Situational context[a]	Scale pattern loadings[b] 1	2	3	Communality	Context/scale *r*[c]
Scale 1: Peer Context Problems					
Reacting to correction	*.75*	−.03	.18	.69	.71
Behaving in the classroom	*.66*	.13	.16	.72	.70
Respecting others' belongings	*.85*	.05	−.08	.70	.66
Sitting during teacher-directed activities	*.65*	.22	.09	.73	.75
Getting along with agemates	*.83*	−.01	.07	.75	.75
Telling the truth	*.63*	.11	−.09	.45	.54
Behaving while standing in line	*.82*	.01	.07	.74	.73
Handling conflicts with other children	*.61*	.08	.15	.57	.65
Free play (self-choice activity)	*.60*	.05	.26	.65	.68
Coping with new learning tasks	*.42*	.49	−.00	.66	.58
Scale 2: Learning Context Problems					
Getting involved in classroom activities	−.11	*.76*	.14	.62	.61
Working with hands (artwork, etc.)	.33	*.68*	−.14	.70	.65
Paying attention in the classroom	.17	*.67*	.07	.67	.64
Seeking teacher's help	−.18	*.59*	.31	.51	.45
Coping with new learning tasks	*.42*	*.49*	−.00	.66	.58
Having companions	.03	*.43*	.18	.33	.43
Taking part in games with other children	.31	*.41*	.20	.62	.61
Scale 3: Teacher Context Problems					
Talking with the teacher	−.14	.24	*.74*	.70	.54
General manner with the teacher	−.02	.19	*.66*	.59	.57
Helping the teacher with jobs	.30	.08	*.47*	.55	.48
Greeting the teacher	.37	−.20	*.45*	.36	.38
Answering questions	.16	.22	*.43*	.49	.51

Note: $N = 800$ comprising the random exploratory analysis sample.
[a] Context descriptioms are abbreviated for convenient presentation.
[b] Values are promaxian pattern loadings at $k = 2$, where hyperplane count is maximized. 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.

Context Problems correlate .738, Peer Context Problems and Teacher Context Problems correlate .581, and Learning Context Problems and Teacher Context Problems correlate .619.

Fit of this 3-factor model for the confirmatory subsample was adequate, with the Satorra–Bentler $\chi^2(185) = 597.59$, CFI = .991, RMSEA = .053 (90% CI = .048/.058). To test longitudinal replication of the structure, analysis was repeated for all confirmatory subsample participants in PreK 1 plus PreK 2 and K plus 1st grade, respectively (each analysis involving 400 children, as requisite for sufficient statistical power) (Meade & Bauer, 2007; Meade, Johnson, & Braddy, 2008). Again, adequate fit was supported, where for the PreK children, $\chi^2(185) = 429.96$, CFI = .988, RMSEA = .057 (90% CL = .050/.064) and for the post-PreK sample $\chi^2(185) = 318.65$, CFI = .994, RMSEA = .043 (90% CL = .035/.051). The disparities between CFI indices $\leq$.01 and RMSEA indices $\leq$.015 across developmental levels indicate that the two levels do not practically differ (Chen, 2007). Further, multiple-group statistical tests of the pattern of loadings (configured invariance) and the magnitude of loadings (metric invariance) revealed statistically nonsignificant differences between the developmental levels. Consequently, the total confirmatory subsample structure is deemed properly representative of the structure across levels.

3.2. Scaling and reliability

Multiple-group DIF analyses identified approximately one-third of the contexts comprising each dimension at each adjacent developmental level to serve as linking contexts. Thus, for example, the 10 Peer Context Problems at PreK 1 were joined by four linking contexts to PreK 2 and another four nonDIF linking contexts from PreK 2 were joined in common to K, etc., where the linking contexts were partly selected so as to represent (via distributions of mean thresholds) the points across the dimensional continuum. For each ASETS dimension, multiple-group vertical equating was applied, contrasting the GPCM and GRM solutions for convergence effectiveness, relative fit, total test information (information is the inverse of measurement error or $1/SE^2$), and reliability. The GPCM uniformly performed better and was adopted for all solutions. Longitudinally, the equating process yielded as follows: Peer Context Problems = 28 contexts (with 12 linking contexts), M slopes ranging 1.18–1.31, M information = .98, and maximum information = 9.78 at $\theta = 2.07$; Learning Context Problems = 19 contexts (nine linking contexts), M slopes ranging .99–1.17, M information .67, and maximum information 6.25 at θ 1·24; and Teacher Context Problems = 14 contexts (six linking contexts), M slopes ranging .66–.74, M information .35, and maximum information 2.10 at θ 2.07. Each ASETS scale was scored for the full national sample via Bayesian EAP with $M = 50$ and $SD = 10$ for the PreK 1 reference level.

Fig. 1a–c illustrates the overplots of total scale information and the standard error for each scale. The information curve essentially displays the degree of precision or reliability of a scale across the continuum of a given trait. The standard error curve conversely represents the degree of imprecision and, in general, decisions based on IRT scores should only be made within the range of trait levels where a test's information exceeds its standard error. It is clear that for every scale, scores ≥ 40 SS points provide sufficient accuracy to support distinctions between adjustment levels. This is particularly important since, as ordinarily applied, users would want to be able to discriminate between adjustment and maladjustment and between varied levels of maladjustment (e.g., SS 60–69 are considered at risk or subclinical, $SS \geq 70$ maladjusted). Coefficient α for Peer Context Problems = .91, Learning Context Problems = .83, and Teacher Context Problems = .73.

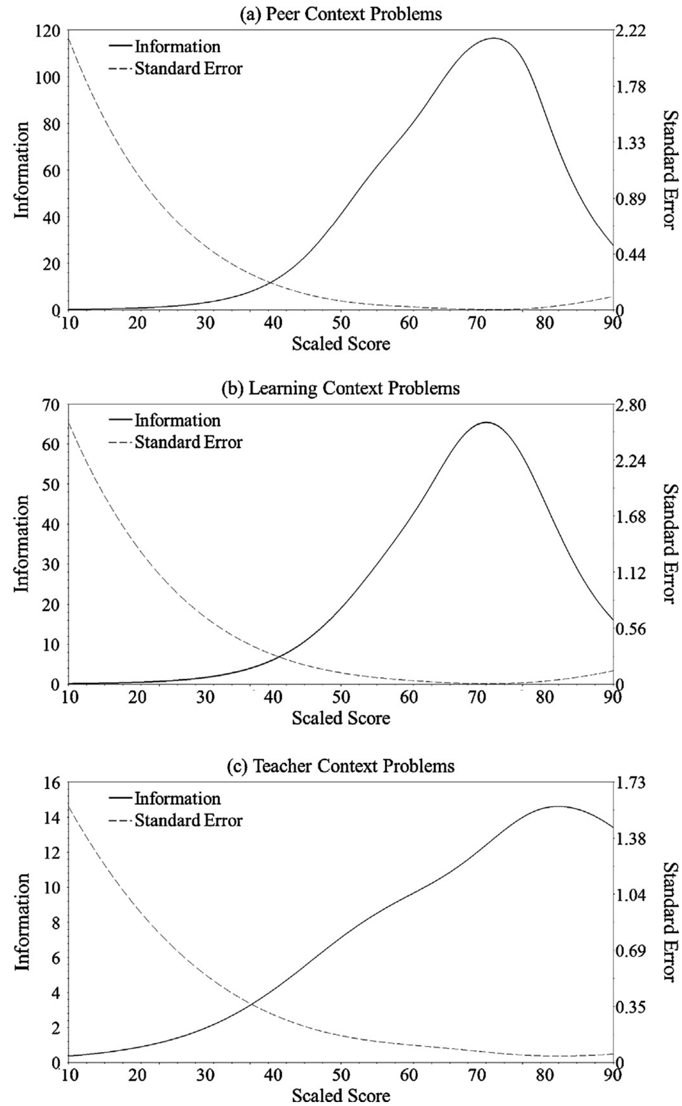

Fig. 1. Distributions of estimated information functions and standard errors for ASETS scales.

3.3. Criterion validity

Table 3 shows concurrent relationships between spring, PreK 2 and spring 1st grade ASETS scores and other relevant measures. Whereas all statistically significant correlations are as directionally expected, ASETS scores evince low moderate to strong relationships with other teacher measures and weak relationship with parent measures and direct assessments of achievement. Given the nested nature of the data, the fourth and eighth columns of the table list the percentage of criterion measure variance that actually pertains to children's individual differences, whereas the parenthetical values reveal how much of that variance is accounted for by a given ASETS scale. Thus, for instance, while Table 3's last column entry for the Pianta teacher Conflict scale indicates that 82.2% of score variance evolves from children's individual differences (rather than teacher or classroom characteristics), it is found that 59.6% of that variance is predictable from children's ASETS Peer Context Problems scores, 30.8% from Learning Context Problems scores and 25.5% from Teacher Context Problems scores. ASETS Learning Context Problems scores are characteristically at least twice as effective as either Peer or Teacher Context Problems scores in accounting for children's individual differences in all areas of WJ academic achievement.

Table 3
Relationships between ASETS context problem scores and concurrent criterion measures.

Criterion measure	Prekindergarten 2 Spring Pearson *r* (% of variance explained between children within classrooms)				First Grade Spring Pearson *r* (% of variance explained between children within classrooms)			
	Peer Context Problems	Learning Context Problems	Teacher Context Problems	% of Explainable variance[a]	Peer Context Problems	Learning Context Problems	Teacher Context Problems	% of Explainable variance[a]
Pianta Child–Teacher Relationships Scale (teacher rating)								
Closeness[b]	−.25 (10.4)	−.35 (24.5)	−.34 (24.7)	74.9	−.28 (8.2)	−.37 (20.9)	−.41 (26.8)	76.7
Conflict[c]	.62 (52.4)	.45 (27.7)	.35 (34.1)	83.1	.72 (59.6)	.56 (30.8)	.52 (25.5)	82.2
Positive Relationship [d]					−.64 (45.2)	−.57 (37.1)	−.56 (33.7)	77.9
Parent rating								
Total Behavior Problems[e]	.20 (4.0)	.19 (7.7)	.14 (8.0)	90.9	.28 (8.5)	.28 (8.4)	.19 (5.8)	88.4
Peabody Picture Vocabulary Test, Third Edition (direct assessment)								
Receptive Vocabulary[f]	−.04† (1.6)†	−.12 (5.6)	−.07 (6.5)	61.3	−.03^{+} (4.8)	.09 (6.3)	−.10 (1.6)†	56.8
Woodcock–Johnson III Test of Achievement (direct assessment)								
Letter Word Identification[g]	−.12 (3.9)	−.21 (8.8)	−.09 (2.4)	75.6				
Applied Problems[h]	−.14 (3.5)	−.21 (7.3)	−.12 (2.7)	78.4				
Pre-Academic Skills[i]	−.18 (6.1)	−.27 (13.8)	−.13 (4.8)	78.0				
Spelling[j]	−.19 (4.6)	−.25 (9.9)	−.12 (4.0)	85.0				
Basic Reading Skills[k]					−.19 (9.2)	−.28 (21.6)	−.18 (7.7)	49.2
Word Attack[l]					−.18 (4.9)	−.27 (12.9)	−.17 (5.8)	56.3
Quantitative Concepts[m]					−.16 (3.1)	−.26 (11.6)	−.16 (2.9)	72.9

Note: Parenthetical values equal the reduction in the residual coefficient (100) as estimated via hierarchical linear modeling. Each two-level random coefficients model entered a given ASETS scale as the covariate. All correlations and fixed effects associated with ASETS scales are significant statistically at $p < .001$ unless indicated † (nonsignificant). ASETS = Adjustment Scales for Early Transition in Schooling.

[a] Total % of potentially explainable variance between children within classrooms. Values equal 1 − intraclass correlation (100), where the intraclass correlation was estimated via hierarchical linear modeling. Each two-level, unconditional means model applied random intercepts for classrooms, where the random effect was significant at $p < .001$.
[b] Prekindergarten 2 $n = 2747$. First Grade Spring $n = 3058$.
[c] Prekindergarten 2 $n = 2743$. First Grade Spring $n = 3050$.
[d] First Grade Spring $n = 3059$.
[e] Prekindergarten 2 $n = 2626$. First Grade Spring $n = 3059$.
[f] Prekindergarten 2 $n = 2699$. First Grade Spring $n = 2900$.
[g] Prekindergarten 2 $n = 2700$.
[h] Prekindergarten 2 $n = 2683$.
[i] Prekindergarten 2 $n = 2683$.
[j] Prekindergarten 2 $n = 2701$.
[k] First Grade Spring $n = 2873$.
[l] First Grade Spring $n = 2875$.
[m] First Grade Spring $n = 2877$.

Table 4
Increased risk of first-grade teacher-reported academic nonproficiency associated with prekindergarten 2 ASETS context problem scores.

Prekindergarten ASETS scale	Odds ratio[a]	95% Confidence limits (lower/upper)	% Risk increment[b]
First-Grade Language and Literacy Ability ($n = 2188$, estimated variance between children = 82.4%)[c]			
Peer Context Problems	1.04	1.03/1.06	4.3
Learning Context Problems	1.07	1.05/1.09	7.2
Teacher Context Problems	1.02	1.01/1.04	2.4
First-Grade Mathematics Ability ($n = 2182$, estimated variance between children = 82.4%)[c]			
Peer Context Problems	1.05	1.03/1.07	4.9
Learning Context Problems	1.08	1.06/1.10	7.5
Teacher Context Problems	1.03	1.02/1.05	3.4
First-Grade Social Science Ability ($n = 2177$, Estimated variance between children = 76.0%)[c]			
Peer Context Problems	1.06	1.05/1.08	5.6
Learning Context Problems	1.08	1.06/1.09	8.3
Teacher Context Problems	1.03	1.02/1.05	3.2

Note: Entries are based on generalized multilevel logistic regression modeling using adaptive quadratures to estimate the multilevel generalized linear model. A separate model was constructed for each academic performance area. ASETS = Adjustment Scales for Early Transition in Schooling.

[a] All values are statistically significant at $p < .001$.
[b] Values = (odds ratio − 1) × 100 and express the percentage increase in risk of future academic nonproficiency per each 1 scaled score increase in the respective ASETS scale.
[c] Based on unconditional models, values = the intraclass correlation × (100), where the intraclass correlation = estimated coefficient for random intercepts/(estimated coefficient for random effects + estimated coefficient for residuals).

Table 5
Linear and higher-order growth parameters for context problem scores of the Adjustment Scales for Early Transition in Schooling (ASETS) over four years.

ASETS scale	Parameter estimate for change (Standard error)		
	Linear	Quadratic	Cubic
Peer Context Problems	.1749** (.2055)	.2548* (.0714)	
Learning Context Problems	−.5768 (.4940)	−1.2347 (.4425)	−.4016** (.0987)
Teacher Context Problems	−1.8185 (.5479)	−1.8720 (.4921)	−.4406** (.1101)

Note: Values are estimated through multilevel individual growth-curve modeling. Models for the 3 ASESTS scales incorporated statistically significant coefficients for random intercepts. Random linear and higher-order slopes were uniformly nonsignificant and thus excluded. Only statistically significant fixed effects parameters are reported unless nonsignificant linear and quadratic estimates appear as requisite for subsequent sequential F tests associated with higher-order estimates. Specification for the full model was, although terms associated with nonsignificant fixed effects for a given model were dropped. Parameters reflect change in ASETS scaled scores per year through 4 years spanning Prekindergarten 1 to First Grade. $N = 3077$.

* $p < .001$.

** $p < .0001$.

Table 4 shows the ability of ASETS scores to forecast increased risk of teacher-assessed academic nonproficiency two years later, at the close of 1st grade. Again, because outcomes are nested within teachers/classrooms, odds ratios are estimated through multilevel modeling. For each outcome (Language and Literacy, Mathematics, Social Science), all ASETS scales are able to indicate significant risk for subsequent nonproficiency. To interpret results, refer to the last column where, for example, the 4.3 entry for Peer Context Problems indicates that, for every 1 *SS* point increase in ASETS Peer Context Problems during spring PreK 2, there is a 4.3% increment in the risk of Language and Literacy nonproficiency at the conclusion of 1st grade.

3.4. Change detection

Multilevel individual growth-curve modeling was applied to test sensitivity to change and to reveal the direction and trends for change for each scale.

Table 5 shows the statistically significant change parameters. Each value indicates the estimated magnitude and direction of change in ASETS *SS* points per developmental level (year). To illustrate, the entry for Peer Context Problems shows that change is linear and curvilinear (quadratic) over time. A quadratic change implies the simplest type of curvature in a change trajectory over time, where in the current study a linear increase of .17 *SS* points and a quadratic increase of .25 *SS* points per year were observed. In contrast, both the Learning Context Problems and Teacher Context Problems scales present long-term decrements in scores as children move through the transitions. These changes are characteristically more complex than those evident for Peer Context Problems, with distinct cubic decrements found for Learning Context Problems (−.40 *SS* points per year) and Teacher Context Problems (−.44 *SS* points per year). A cubic coefficient indicates a more complicated curvature, with multiple shifts in the direction of a change trajectory over time.

3.5. Differential change detection

The ability to detect change invites many questions for practice and research. Here we explore one interesting avenue of inquiry that demonstrates how ASETS scores can be used to discover the nature of long-term sociobehavioral changes that distinguish children who reach successful versus unsuccessful outcomes at the end of 1st grade. For this exploration, we extended each of the multilevel models discussed above to isolate the average change trajectories for different future outcomes. Results are illustrated in Fig. 2a–f. The change trajectories are controlled for the effects of children's gender, ethnicity, primary language and special-needs status, and urban residence (age not found significant for any model). Model specification was:

$$\begin{aligned}\text{Context Problems}_{ijk} &= \gamma_{000} + \gamma_{100}\text{Time}_{ijk} + \gamma_{200}\text{Time}^2_{ijk} + \gamma_{300}\text{Time}^3_{ijk} + \gamma_{400}\text{Time}^4_{ijk}\\ &+ \gamma_{010}\text{Male}_j + \gamma_{020}\text{Black}_j + \gamma_{030}\text{Hispanic}_j\\ &+ \gamma_{050}\text{LanguageStatus}_j + \gamma_{060}\text{SpecialNeeds}_j + \gamma_{040}\text{Urban}_j\\ &+ \gamma_{070}\text{Nonproficient}_j(\gamma_{170}\text{Time}_{ijk} * \text{Nonproficient}_j)\\ &+ (\gamma_{270}\text{Time}^2_{ijk} * \text{Nonproficient}_j) + (\gamma_{370}\text{Time}^3_{ijk} * \text{Nonproficient}_j)\\ &+ (\gamma_{470}\text{Time}^4_{ijk} * \text{Nonproficient}_j) + (\mu_{00k} + \mu_{10k}\text{Time}_{ijk})\\ &+ (\mu_{0jk}) + r_{ijk}.\end{aligned}$$

Fig. 2a shows what occurs when the change trajectories for Peer Context Problems are distinguished by whether children perform nonproficiently (the lowest quintile) versus proficiently in spring, 1st grade, WJ Mathematics Reasoning. Specifically, the upper growth trajectory shows the average levels and changes in Peer Context Problems for children who ultimately failed in mathematics. The lower growth trajectory shows the level and changes in Peer Context Problems for children who were adequately successful in mathematics. The eventual nonproficient children show more Peer Context Problems all along (effect size [*ES*] = .49 by the end of 1st grade), with a common decrease over time for all children. To clarify, effect sizes in this study were calculated using population parameter estimates (least-squares means) corrected for group imbalance. In this first example, the effect size is calculated as 50.49 (least-squares mean for nonproficient children at status) – 47.38 (least-squares mean for proficient children at status)/6.19 (estimated maximum-likelihood standard deviation of predicted scores of all children at status). See Fantuzzo, Gadsden, and McDermott (2011, p. 789) for further details on effect size computation.

Fig. 2b also tracks changes in Peer Context Problems, but this time for children who end up manifesting high levels of problem behavior (upper quintile) or not as observed by parents. The separation of the trajectories is most noticeable at the close of 1st grade (*ES* = .66). Fig. 2c and d both pertains to changing Learning Context Problems, the first where children are distinguished by future WJ Basic Reading Skills level and the second by future child conflict levels with teachers. Both illustrations reveal complex change patterns where all children tend to decrease in Learning Context Problems as they transition through PreK 1 and 2. Thereafter the trajectories for children who eventually encounter reading or conflict problems show marked increments in Learning Context Problems, where by the end of 1st grade, *ES* = .72 for the reading example and *ES* = 1.48 for the teacher–child conflict example. In contrast, Fig. 2e displays distinct cubic change patterns for Teacher Context Problems,

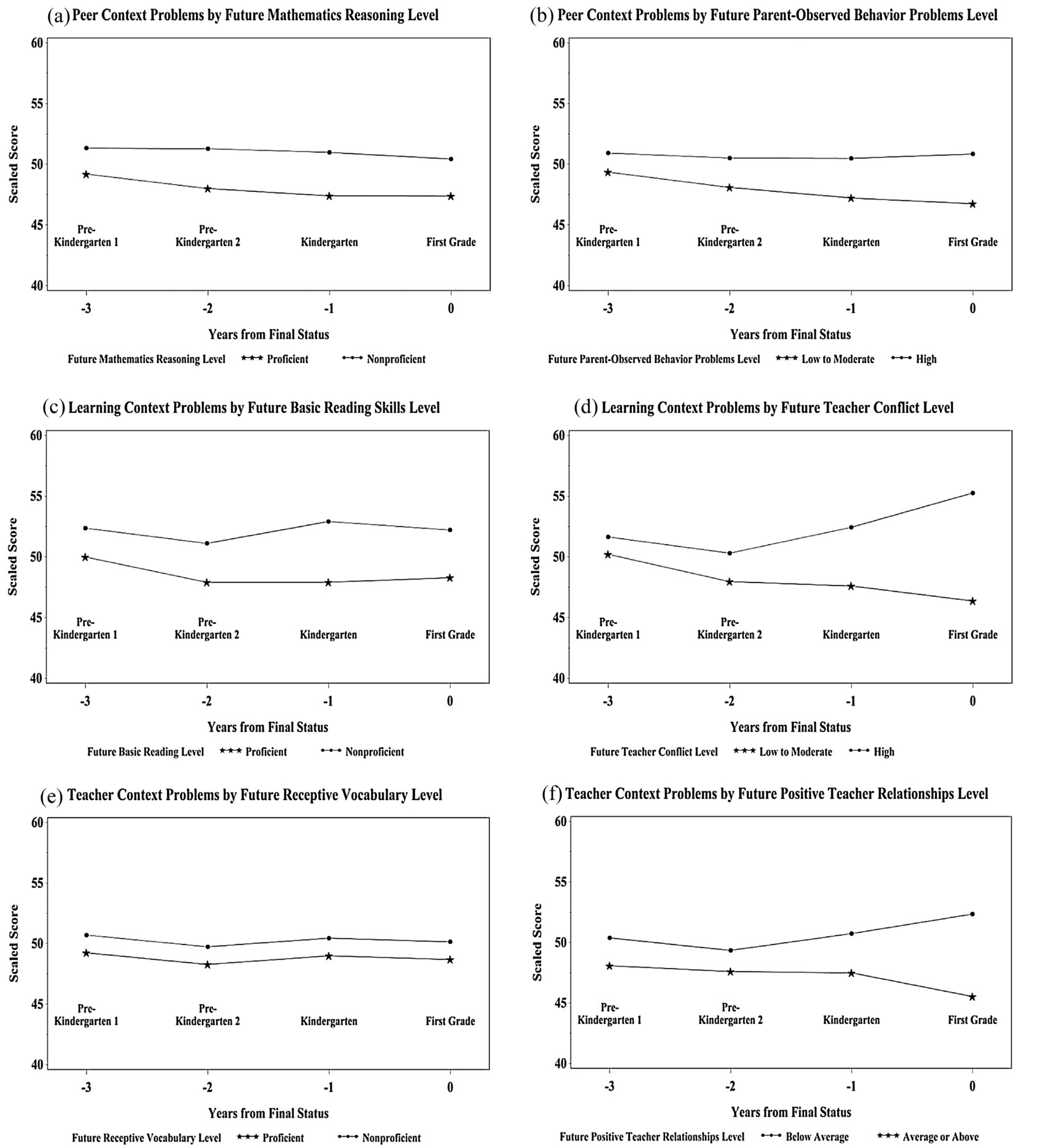

Fig. 2. Estimated average growth trajectories for ASETS scales. Estimated average growth trajectories for ASETS scales.

where children who eventually show lack of success in future receptive vocabulary consistently have more Teacher Context Problems (*ES*s > .40 across the years). Lastly, Fig. 2f shows changes in levels of ASETS Teacher Context Problems based on whether or not children manifest positive relationships with teachers in 1st grade. The departure of the trajectories is striking (*ES* = 1.49 at close of 1st grade), beginning as children leave PreK levels. A notable feature of every set of ASETS trajectories presented is that the trajectories for the eventually successful and unsuccessful children are significantly different statistically even at the initial ASETS assessments in PreK 1 (*ES*s ranging .38–.50).

4. Discussion

The current study of the ASETS represents a significant contribution to the literature given the current research and policy priority on developing a better understanding of socio-emotional development in young children. The current study has described the design and validity evidence for contextually-specific measures of early childhood social and behavioral adjustment within school using the ASETS. Primary analyses of representative nationwide data from the Head Start Impact Study informed developmental-transitional stability and change in adjustment across numerous school contexts. Support was found for the concurrent validity of ASETS contextual scales and their ability to assess future risk of academic and behavioral problems. ASETS scales were also shown to reveal differential, contextually-based, change trajectories across four years of early school transition.

It should be no surprise that disparities might arise between the dimensional structure of the earlier Head Start ASPI (Bulotsky-Shearer et al., 2008) and the new ASETS. ASPI was based exclusively on Head Start children (73.5% being African American) from one Northeastern school district and covering the full enrollments of 233 classrooms (an average of 16.4 children per teacher). Alternatively, ASETS was drawn on a nationally representative sample across many demographic strata, more than 6700 classrooms (typically assessing <2 children per classroom), and sampling both non-Head Start and Head Start PreK and longitudinally through K and 1st grade. Also, the ASETS factor analytic strategy took advantage of techniques now available for ordinal response scales and longitudinal factor extraction. Nonetheless, the resulting ASETS structure is remarkably similar to the earlier ASPI structure, featuring Peer, Learning, and Teacher Context Problems, although the membership of component contexts is shifted somewhat for ASETS and two additional contexts (viz., Handling conflicts with other children, Coping with new learning tasks) were successfully incorporated into the ASETS dimensional structure. This general continuity in structure, not only to PreK children outside of Head Start but temporally to formal schooling, is indicative of the overall generalizability of ASETS's situtype dimensions.

Recent research with the national ASETS (McDermott et al., 2013) has concentrated on individual problem behaviors (rather than contexts) and defined reliable and valid measures of phenotypic maladjustment. As noted, the phenotype dimensions indicate collections of behaviors that reflect similarity in appearance and function across multiple contexts. These phenotype dimensions include Aggression and Attention Seeking (and their higher-order composite, Overactivity) and Reticence/Withdrawal and Low Energy (and their composite, Underactivity). The current study alternatively concentrates on the dimensional structure of the contexts within which problems emerge, irrespective of phenotypic similarity. This construction was designed to inform *what* types of adjustment problems emerge and transition across time (the phenotype dimensions) and *when* and *where* those problems emerge (the situtype dimensions). For example, a practitioner may now readily detect, based on ASETS national norms, that a particular child is manifesting relatively high levels ($SS \geq 60$) of Attention Seeking behavior and at once discover through similar score elevation in Teacher Context Problems that the behavior is isolated to situations involving the teacher, rather than situations involving classmates or structured learning per se. In turn, this information provides important clues to motivation and potential intervention, especially as related to the fit between a child's behavioral disposition and the evocative and reactive situational contexts within the classroom (Chess & Thomas, 1991).

As illustrated, the ASETS scales have the capacity to detect differential socio-behavioral change patterns across the transitions, even when controlling for important alterative factors (child sex, ethnicity, special needs, language status, etc.). Thus, it is now possible to assess the same children on multiple occasions over the developmental period connecting early prekindergarten and subsequent formal schooling and to have reasonable confidence that the phenomena under study have construct continuity and psychometric integrity. This is ideal for program evaluation purposes and for long-term longitudinal research. Moreover, because the earlier ASPI (Bulotsky-Shearer et al., 2008; Noone-Lutz et al., 2002) and ASETS are identical in format and content, it means that the new ASETS dimensions and IRT scoring systems can be applied to rescore any ASPI data so that they will acquire national nomothesis and capacity for use in follow-up investigations reaching into kindergarten and first grade.

ASETS opens several other lines of research. First, the multidimensional and transitional features of ASETS phenotypic and situtype scales invite research on the typological (latent class transition) change of early childhood socio-behavioral adjustment, thereby informing children's membership in distinct subpopulation and temporal movements into other subpopulations. Second, the longitudinal ASETS scales make possible the identification and study of distinct types of long-term change trajectories (growth mixture modeling) as they relate to familial, neighborhood, and school factors.

ASETS was designed and validated in partial response to the recent National Research Council (2008) call for production of technically advanced and purposeful assessment tools for use in early childhood education. In this enterprise, we began with a novel approach that could dually assess the typical nature and contextual circumstances of young children's social and behavioral adjustment. The instrument was grounded in the nexus of modern contextual and transitional theory, and implemented through model-based measurement. Perhaps most importantly, ASETS contributes a nationally standardized and multifaceted perspective on childhood adjustment as it develops in time and place.

References

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms & profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Benner, A. D., & Crosnoe, R. (2011). The racial/ethnic composition of elementary schools and young children's academic and socioemotional functioning. *American Educational Research Journal*, *48*, 621–646.

Bernstein, I. H., & Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence for multidimensionality due to item categorization. *Psychological Bulletin*, *105*, 467–477.

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In R. M. Lemer (Ed.), *Handbook of child psychology (Vol. 1, 8th ed.): Theoretical models of human development* (pp. 793–828). Hoboken, NJ: Wiley.

Bulotsky-Shearer, R., Fantuzzo, J. F., & 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*, 139–154.

Buss, K. A. (2011). Which fearful toddlers should we worry about? Context, fear regulation, and anxiety risk. *Developmental Psychology*, *47*, 804–819.

Campbell, S. B. (2001). *Behavior problems in preschool children*. New York, NY: Guilford.

Campbell, J. M., & James, C. L. (2007). Assessment of social and emotional development in preschool children. In B. A. Bracken, & J. Nagle (Eds.), *Psychoeducational assessment of preschool children* (4th ed., pp. 111–135). Mahwah, NJ: Erlbaum.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464–504.

Chess, S., & Thomas, A. (1991). Temperament and the concept of goodness of fit. In J. Strelau, & A. Angleitner (Eds.), *Explorations in temperament: International perspectives on theory and measurement* (pp. 15–28). New York, NY: Plenum.

Dumont, R., & Willis, J. O. (2006). Test descriptions and reviews. In E. Fletcher-Janzen, & C. R. Reynolds (Eds.), *The special education almanac* (pp. 39–146). Hoboken, NJ: Wiley.

Dunn, L. M., Dunn, L. L., & Dunn, D. M. (1997). *Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance.

Egger, H. L., & Angold, A. (2006). Common emotional and behavioral disorders in preschool children: Presentation, nosology, and epidemiology. *Journal of Child Psychology and Psychiatry*, *47*, 313–337.

Eivers, A. R., Brendgen, M., & Borge, A. I. H. (2010). Stability and change in prosocial and antisocial behavior across the transition to school: Teacher and peer perspectives. *Early Education and Development*, *21*, 843–864.

Entwisle, D. R., & Alexander, K. L. (1993). Entry into school: The beginning school transition and educational stratification in the United States. *Annual Review of Sociology*, *19*, 401–423.

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (2005). First grade and educational attainment by age 22: A new story. *American Journal of Sociology*, *110*, 1458–1502.

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*, 272–299.

Fantuzzo, J. W., Gadsden, V. L., & McDermott, P. A. (2011). An integrated curriculum to improve mathematics, language, and literacy for Head Start children. *American Educational Research Journal*, *48*, 763–793.

Feeney-Kettler, K. A., Kratochwill, T. R., & Kettler, R. J. (2011). Identification of preschool children at risk for emotional and behavioral disorders: Development and validation of a universal screening system. *Journal of School Psychology*, *49*, 197–216.

Fulmer, S. M., & Frijters, J. C. (2009). A review of self-report and alternative approaches in the measurement of student motivation. *Educational Psychology Review*, *21*, 219–246.

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2011). Performance of Velicer's minimum average partial factor retention method with categorical variables. *Educational and Psychological Measurement*, *71*(3), 551–570.

Goldsmith, H. G., & Davidson, R. J. (2004). Disambiguating the components of emotion regulation. *Child Development*, *75*, 361–365.

Hemmeter, M. L., & Ostrosky, M. (2006). Social and emotional foundations for early learning: A conceptual model for intervention. *School Psychology Review*, *33*, 583–601.

Horn, W. F., Wagner, A. E., & Ialongo, N. (1989). Sex differences in school-aged children with pervasive attention deficit hyperactivity disorder. *Journal of Abnormal Child Psychology*, *17*, 109–125.

Joosten, A. V., Bundy, A. C., & Einfeld, S. L. (2012). Context influences the motivation for stereotypic and repetitive behaviour in children diagnosed with intellectual disability with and without autism. *Journal of Applied Research in Intellectual Disabilities*, *25*, 262–270.

Justice, L. M., Petscher, Y., Schatschneider, C., & Mashburn, A. (2011). Peer effects in preschool classrooms: Is children's language growth associated with their classmates' skills? *Child Development*, *82*, 1768–1777.

Kagan, J. (2003). Biology, context, and developmental inquiry. *Annual Review of Psychology*, *34*, 1–23.

Kataoka, S. H., Zhang, L., & Wells, K. B. (2002). Unmet need for mental health care among U.S. children. Variation by ethnicity and insurance status. *American Journal of Psychiatry*, *159*, 1548–1555.

Kern, L., & Clemens, N. H. (2007). Antecedent strategies to promote appropriate classroom behavior. *Psychology in the Schools*, *44*, 65–75.

Knol, D. L., & Berger, M. P. (1991). Empirical comparison between factor analysis and multidimensional item response models. *Multivariate Behavioral Research*, *26*, 457–477.

LeBuffe, P. A., & Naglieri, J. A. (1999). *Devereaux Early Childhood Assessment-Technical manual*. Lewisville, NC: Kaplan Press.

Marsh, H. W., Lien, G. A. D., Martin, A. J., Morin, A. J. S., & Nagengast, B. (2011). Methodological measurement fruitfulness of exploratory structural equation modeling (ESEM): New approaches to key substantive issues in motivation and engagement. *Journal of Psychoeducational Assessment*, *29*, 322–346.

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

McDermott, P. A., Steinberg, C. M., & Angelo, L. E. (2006). Situational specificity makes the difference in assessment of youth behavior disorders. *Psychology in the Schools*, *42*, 121–136.

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*, 97–115.

McDonald, R. P., & Ahlawat, K. S. (1974). Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology*, *27*, 82–99.

Meade, A. W., & Bauer, D. J. (2007). Power and precision in confirmatory factor analytic tests of measurement invariance. *Structural Equation Modeling*, *14*, 611–635.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, *93*, 568–592.

Merrell, K. W. (2003). *Preschool and Kindergarten Behavior Scales-Second Edition (PKBS-2)*. Austin, TX: PRO-ED.

Mian, N. D., Wainwright, L., Briggs-Gowan, M. J., & Carter, A. S. (2011). An ecological risk model for early childhood anxiety: The importance of early child symptoms and temperament. *Journal of Abnormal Child Psychology*, *39*, 501–512.

Mischel, W. (2004). Toward an integrative science of the person. *Annual Review of Psychology*, *55*, 1–22.

Mischel, W., Shoda, Y., & Mendoza-Denton, R. (2002). Situation-behavior profiles as a locus of consistency in personality. *Psychological Science*, *11*, 50–54.

Mislevy, R. J. (1986). Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics*, *11*, 3–31.

Moll, H., & Tomasello, M. (2012). Three-year-olds understand appearance and reality—Just not about the same object at the same time. *Developmental Psychology*, *48*, 1124–1132.

Mooijaart, A. (1983). Two kinds of factor analysis for ordered categorical variables. *Multivariate Behavioral Research*, *18*, 423–441.

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

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

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

National Research Council. (2008). *Early childhood assessment: Why, what, and how*. Washington, DC: National Academies Press.

Noone-Lutz, M., Fantuzzo, J. F., & McDermott, P. A. (2002). Multidimensional assessment of emotional and behavioral adjustment of low-income preschool children: Development and initial validation. *Early Childhood Research Quarterly*, *17*(338–355), 596–601.

Norwood, S. (2007). Validity of self-reports of psychopathology from children of 4–11 years of age. *Vulnerable Children and Youth Studies*, *2*, 89–99.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw Hill.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Pychometrika*, *44*, 443–460.

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., Cox, M. J., & Snow, K. L. (Eds.). (2007). *School readiness and the transition to kindergarten in the era of accountability*. Baltimore, MD: Brookes.

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.

Powell, D. R., Burchinal, M. R., File, N., & Kontos, S. (2008). An eco-behavioral analysis of children's engagement in urban public school preschool classrooms. *Early Childhood Research Quarterly*, *23*, 108–123.

Rescorla, L. A., Achenbach, T. M., Ginzburg, S., Ivanova, M., Dumenci, L., Almqvist, F., et al. (2007). Consistency of teacher-reported problems for students in 21 countries. *School Psychology Review*, *36*, 91–110.

Rescorla, L. A., Achenbach, T. M., Ivanova, M. Y., Harder, V. S., Otten, L., Bilenberg, N., et al. (2011). International comparisons of behavioral and emotional problems in preschool children: Parents' reports from 24 societies. *Journal of Clinical Child & Adolescent Psychology*, *40*, 456–467.

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior Assessment System for Children-Second Edition (BASC-2)*. Circle Pines, MN: American Guidance Service.

Sabol, T. J., & Pianta, R. C. (2012). Recent trends in research on teacher–child relationships. *Attachment and Human Development*, *14*, 213–231.

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

Samejima, F. (1996). The graded response model. In W. J. van der Linden, & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 85–100). New York, NY: Springer.

Sameroff, A. J. (2010). A unified theory of development: A dialectic integration of nature and nurture. *Child Development*, *81*, 6–22.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*, 507–514.

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, *12*, 162–171.

Stott, D. H. (1966). *The social adjustment of children*. London, UK: University of London Press.

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

Thorsen, K. L., Goldberg, W. A., Osann, K., & Spence, M. A. (2008). Birthday and non-birthday videotapes: The importance of context for the behavior of young children with autism. *Journal of Autism Developmental Disorders*, *38*, 1047–1058.

U.S. Department of Health and Human Services [DHHS]. (2001). *Head Start FACES: Longitudinal findings on program performance. Third progress report*. Washington, DC: Administration for Children and Families.

U.S. Department of Health and Human Services [DHHS]. (2010a). *Head Start Impact Study final report*. Washington, DC: Administration for Children and Families.

U.S. Department of Health and Human Services [DHHS]. (2010b). *Head Start Impact Study technical report*. Washington, DC: Administration for Children and Families.

U.S. Department of Health and Human Services [DHHS] (President's New Freedom Commission on Mental Health). (2003). *Achieving the priorities: Transforming mental health care in America*. Rockville, MD: U.S. Department of Health and Human Services.

Vaden-Kiernan, M., D'Elio, M. A., O'Brien, R. W., Tarullo, L. B., Zill, N., & Hubbell-McKey, R. (2010). Neighborhood as a developmental context: A multilevel analysis of neighborhood effects on Head Start families and children. *American Journal of Community Psychology*, *45*, 49–67.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*, 321–327.

Visser, M., Kunnen, S. E., & van Geert, P. L. C. (2010). The impact of context on the development of aggressive behavior in special elementary school children. *Mind, Brain and Education*, *4*, 34–43.

von Baeyer, C. L., Forsyth, S. J., Stanford, E. A., Watson, M., & Chambers, C. T. (2009). Response biases in preschool children's ratings of pain in hypothetical situations. *European Journal of Pain*, *13*, 209–213.

von Suchodoletz, A., Trommsdorff, G., Heikamp, T., Wieber, F., & Gollwitzer, P. M. (2009). Transition to school: The role of kindergarten children's behavior regulation. *Learning and Individual Differences*, *19*, 561–566.

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

Wannarka, R., & Ruhl, K. (2008). Seating arrangements that promote positive academic and behavioral outcomes: A review of empirical research. *Support for Learning*, *23*, 89–93.

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*, 46–54.
Wood, J. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychological Methods*, *1*, 354–365.
Woodcock, R. W., McGrew, K. S., & Mather, N. (2002). *Woodcock–Johnson III Tests of Achievement*. Itasca, IL: Riverside.
Yates, A. (1987). *Multivariate exploratory data analysis: A prospective on exploratory factor analysis*. Albany, NY: State University of New York Press.
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*, 396–408.