Observed Quality and Psychometric Properties of the CLASS-T in the Early Head Start Family and Child Experiences Survey

MAY 2014
BRIEF 2014-34

OPRE Technical Brief

Eileen Bandel, Nikki Aikens, Cheri A. Vogel, Kimberly Boller, and Lauren Murphy

Background and Purpose

In this technical brief, we report on the use of the Toddler Classroom Assessment Scoring System (CLASS-T; Pianta et al. 2010; La Paro et al. 2012) in the Early Head Start Family and Child Experiences Survey (Baby FACES). We begin by providing a brief overview of the Baby FACES study, including its methodology and approach to data collection. Next, we provide a descriptive snapshot of process quality in center-based settings drawing on observations conducted in Early Head Start classrooms serving 2- and 3-year-old children in Baby FACES. Finally, we document evidence from Baby FACES of the instrument's psychometric properties, including results of factor analyses, internal consistency reliability, and concurrent and predictive associations to child development outcomes and other key indicators of quality.[1]

About Baby FACES

Baby FACES is an ongoing study of Early Head Start programs designed to inform policy and practice at both national and local levels. The study provides a descriptive snapshot of Early Head Start services, including their intensity and quality, the characteristics of the children and families served, and how children and families are faring in terms of key areas of development and well-being (Vogel et al. 2011). In 2007, the Office of Planning, Research and Evaluation in the Administration for Children and Families (ACF), U.S. Department of Health and Human Services, contracted with Mathematica Policy Research and its partners to implement this six-year longitudinal study in 89 Early Head Start programs around the country. Two cohorts of children were enrolled into the study in spring 2009: (1) a Newborn Cohort comprising those families in which the mother was pregnant or the child was less than 2 months old, and (2) a 1-year-old Cohort comprising children who were approximately age 1 at the time of the first data collection round. Baby FACES followed the children in both cohorts, collecting data in the spring of each year until they reached age 3 or exited the Early Head Start Program.

Data Collection and Measures

Baby FACES employed a comprehensive approach to data collection that gathers information on programs, staff, and families using multiple modes. Information on

[1] Additional information about the characteristics and quality of Early Head Start programs in Baby FACES can be found in the study's age 2 and age 3 reports (forthcoming).

MATHEMATICA Policy Research

OPRE

program operations and services was obtained from program directors, and frontline staff (teachers and home visitors) reported on their education, experience, and demographic characteristics. We learned about child and family characteristics through interviews with parents, direct child assessments when children were 2 and 3 years old, and staff members' reports on children's developmental progress. We gathered program service information through weekly staff reports on services offered by programs and received by families throughout their enrollment in the program, and we measured quality through observations of classrooms and home visits. Key data sources and measures used in this report are presented in Table 1.

Table 1. Data Sources and Measures Used in This Report

Staff Characteristics and Program Quality	
Teacher Demographic Characteristics	Teachers reported on their training and education experiences, including state-awarded and Child Development Associate (CDA) credentialing, experience working with infants and toddlers, and the likelihood of returning to their job in the coming year.
The Center for Epidemiologic Studies Depression Scale—Short Form (CESD-SF; Radloff 1977; Ross et al. 1983)	The CESD-SF is the short form of the full-version CESD, a self-administered screening tool used to identify symptoms of depression or psychological distress. The tool was used to measure depression symptoms in teachers.
Parent-Caregiver Relationship Scale (PCRS; Elicker et al. 1997)	The PCRS measures the perceived relationship between the parent and the child's teacher/caregiver. Items capture dimensions of the parent-caregiver relationship as reported by teachers, including trust and confidence, communication, respect/acceptance, caring, competence/knowledge, partnership/collaboration, and shared values.
Program Director Interview	Program directors reported on teacher turnover in the last 12 months, including reasons for staff departures (such as changes in career, higher compensation in the same field, personal reasons).
Child-Adult Ratio	Center-based classroom observations included assessments of group sizes and child-adult ratios.
Child Language Development	
Ages & Stages Questionnaires, Third Edition (ASQ-3; Squires et al. 2009)	The ASQ-3 is a parent-report tool for screening children ranging in age from 1 month to 5½ years for developmental delays in five developmental areas: (1) communication, (2) gross motor, (3) fine motor, (4) personal-social, and (5) problem solving. The scores reported here are based on parent reports in the area of communication obtained when children were 2 and 3 years old.
Preschool Language Scale—Fourth Edition (PLS-4; Zimmerman et al. 2002a, 2002b)	The PLS-4 is a direct child assessment used to evaluate the receptive and expressive language skills, and understanding and use of grammatical rules, of children from birth to 6 years of age. It is composed of two subscales: Auditory Comprehension (AC) and Expressive Communication (EC). Baby FACES used the AC subscale of the English and Spanish editions at the age 2 and age 3 assessments.
Peabody Picture Vocabulary Test-4th Edition (PPVT-4; Dunn and Dunn 2007)	The PPVT-4 is a measure of English receptive vocabulary suitable for a wide range of ages, spanning 2½ years through adulthood. For each item, children are presented with four pictures as response options, and are asked to point to the picture that best illustrates the target word spoken by the assessor. In Baby FACES, we administered the PPVT-4 to children at the age 3 assessment regardless of their primary language, yielding a measure of their ability to comprehend or understand vocabulary in English.

(continued)

Table 1. Data Sources and Measures Used in This Report *(continued)*	
MacArthur-Bates Communicative Development Inventories—Infant Short Form (CDI-III; Fenson et al. 2000; Vagh, Mançilla-Martinez, and Pan, unpublished manuscript)	The CDI-III is designed to assess children's early receptive and expressive language and communication skills. Teachers/caregivers reported on children's vocabulary comprehension and vocabulary production using the English checklist; staff who themselves spoke Spanish reported on Spanish-speaking children's vocabulary comprehension and production using the English and Spanish CDI. The scores reported here are based on reports obtained when children were 2 and 3 years old.
Child Social-Emotional Development	
The Brief Infant Toddler Social Emotional Assessment (BITSEA; Briggs-Gowan and Carter 2006)	The BITSEA is the screener version of the longer Infant Toddler Social Emotional Assessment (ITSEA), designed to detect delays in the acquisition of social-emotional competencies and behavior problems in children 12 to 36 months of age. The scores reported here are based on teacher reports obtained when children were 2 and 3 years old.
Parent-Child Interaction Rating Scales for the Two-Bag Assessment (Mathematica Policy Research 2010)	The Parent-Child Interaction (PCI) Rating Scales for the Two-Bag Assessment consist of twelve scales that assess a range of child and parent behaviors. Assessments are based on a semi-structured, parent/child play task that was administered as part of the age 2 and age 3 assessments (and video recorded for later coding). Four scales address the child's behavior during the 8-minute interaction: (1) engagement of parent (extent to which the child initiates and/or maintains interaction with the parent); (2) sustained attention with objects (degree of involvement with and focused exploration of the play materials); (3) enthusiasm (degree of vigor and confidence during the task); and (4) negativity toward parent (displays of anger, hostility, or disdain). Each behavior was rated from 1 to 7, ranging from a very low incidence to a very high incidence of the behavior.
Bayley Behavioral Rating Scale (BRS; Bayley 1993)	The BRS is one of the three component scales of the Bayley Scales of Infant Development—Second Edition (Bayley 1993), and was used by assessors to rate children's behavior during the direct assessment. Two subscales of the BRS were used in Baby FACES: (1) Orientation/Engagement (cooperation with the assessor during the assessment, positive affect, and interest in the test materials); and (2) Emotional Regulation (ability to transition between tasks and test materials, negative affect, and frustration with tasks during the assessment). Assessors completed the BRS as part of the age 2 and age 3 assessments.
Behavior Problems Index (BPI; Peterson and Zill 1986)	At age 3, parents and Early Head Start teachers rated children's behavior problems using the BPI. The BPI measures child externalizing behavior problems (such as aggression and hyperactivity) and internalizing behavior problems (such as anxiety and depression) exhibited by children in the past 3 months.

Measuring Classroom Quality Using the CLASS-T

The CLASS-T is an adaptation of the Pre-K CLASS (Pianta et al. 2008) that focuses on effective teacher-child interaction in classrooms serving toddlers. The CLASS-T is intended for use with children ages 15 to 36 months and measures process quality along eight dimensions (see Table 2). Dimension scores are defined by observable indicators along a seven-point scale, with ratings reflecting scores in the low (1-2), mid (3-5), and high (6-7) ranges. These dimensions comprise two domains: (1) Emotional and Behavioral Support and (2) Engaged Support for Learning.

The Baby FACES team used the CLASS-T to assess process quality in Early Head Start classrooms when sample children were 2 and 3 years old. Here, we present findings based on observations conducted in 220 center-based classrooms serving the 1-year-old Cohort in their age 2 year (spring 2010), and in 235 classrooms serving the 1-year-old

Table 2. Dimensions of Process Quality Measured by the CLASS T	
Dimension	**Description**
Emotional and Behavioral Support	
Positive Climate	Degree of warmth, respect, and mutual enjoyment communicated between the teacher and children, either verbally or non-verbally
Negative Climate	Frequency and intensity of teacher and child expressions of negativity
Teacher Sensitivity	Responsiveness to and awareness of children's individual needs and emotional functioning, encompassing the extent to which the teacher is available to provide reassurance and encouragement
Regard for Child Perspectives	Degree to which teacher-child interactions reflect children's interests and motivations, and encourage children's responsibility and independence
Behavior Guidance	The teacher's ability to promote children's self-regulation by using proactive approaches, supporting positive behaviors, and guiding and curtailing problem behavior
Engaged Support for Learning	
Facilitation of Learning and Development	The manner by which the teacher facilitates activities that support children's learning and developmental opportunities
Quality of Feedback	Degree to which the teacher provides feedback that promotes learning and understanding, and extends children's participation
Language Modeling	The quality and quantity of the teacher's use of language to support and encourage children's language development

Source: La Paro et al. (2012).

and Newborn Cohorts in their age 3 year (spring 2011 and 2012, respectively). At the age 2 quality assessment, Baby FACES children were between 22 and 27 months of age; and at the age 3 assessment, they were between 33 and 39 months of age.[2] Classroom observations included counts of infants and toddlers and the adults caring for them, which we used to assess features of quality care including child-to-staff ratios and group sizes.

The CLASS-T manual was under development during the Baby FACES study data collection planning (2009–2010). However, a pilot version of the instrument was available (Pianta et al. 2010). As the developers continued to refine the instrument during the course of the project, we used the pilot version of the instrument for the duration of the study.[3]

We trained field observers in the use of the CLASS-T in a rigorous and comprehensive way. Two Mathematica survey leaders on the project traveled to the Frank Porter Graham Child Development Institute to meet with the CLASS-T developers and receive a preliminary training. These project staff led planning for the classroom observation field staff training and oversaw field observations throughout the data collection. One of the measure authors and another Mathematica staff member conducted the field training over a two-day period. Field staff observed video clips, practiced coding, and discussed their ratings. They took a certification test using three videotaped observations. Only staff who averaged at least 80 percent agreement (exact or within one point) with the master codes across the three clips were allowed to conduct observations in the field. Interrater reliability

[2] Early Head Start is available to low-income pregnant women and families with children up to age 3. Although specific program policies varied, most programs continued to serve children who turned 3 until the end of the program year. Thus, the analyses presented in this brief include observations in Early Head Start Classrooms serving some children over 36 months of age (at the age 3 data collection).

[3] Noteworthy changes to the CLASS-T include the provision of additional details, examples, and guidance to observers for the indicators within dimension ratings.

checks throughout the field period met standards established by the measure developers (80 percent agreement within one point); across all observers, agreement with a paired gold standard observer averaged 95 percent.

Quality of Early Head Start Classrooms in Baby FACES

Based on observations of Early Head Start classrooms serving toddlers, Baby FACES provides a descriptive snapshot of process quality.

Classrooms Offer Group Sizes and Ratios Within the Performance Standards and Professional Recommendations

Two- and three-year-old children are in classrooms with observed group sizes of approximately 6 children and child-teacher ratios of 2.7 children per teacher, on average (Table 3). These numbers fall within the Head Start Program Performance Standards (four children per adult and a maximum group size of eight; ACF 2009). Nearly all Baby FACES children (99 and 98 percent at ages 2 and 3, respectively) are in classrooms with observed group sizes of eight or fewer children; at each age, 98 percent of children are in classrooms with ratios of 4 to 1 or better.

Toddlers were in Early Head Start classrooms that scored in the mid-range of quality. Classrooms were strongest in the area of Emotional/ Behavioral Support; lower scores were observed in the area of Engaged Support for Learning.

Table 3. Most Children Are in Classrooms in the Mid Range of Quality

	Weighted Mean (Standard Error)	
	Age 2	Age 3
Group Size	5.9 (0.14)	6.1 (0.14)
Child-Adult Ratio	2.7 (0.06)	2.7 (0.09)
CLASS-T Emotional and Behavioral Support	5.3 (0.07)	5.3 (0.09)
Positive Climate	5.6 (0.12)	5.5 (0.13)
Negative Climate	1.3 (0.04)	1.4 (0.07)
Teacher Sensitivity	4.8 (0.10)	4.8 (0.10)
Regard for Child Perspectives	4.7 (0.09)	4.7 (0.10)
Behavior Guidance	4.8 (0.10)	4.7 (0.11)
CLASS-T Engaged Support for Learning	3.6 (0.15)	3.3 (0.13)
Facilitation of Learning and Development	3.9 (0.12)	3.7 (0.13)
Quality of Feedback	3.5 (0.18)	3.1 (0.12)
Language Modeling	3.4 (0.16)	2.9 (0.15)
Sample Size	**295–302**	**297–304**

Source: Spring 2010, 2011, and 2012 Baby FACES Classroom Observations. Reported observations are for children in the 1-year- old Cohort at age 2 and for children in the 1-year-old and Newborn Cohorts at age 3. Estimates are at the child level.
CLASS-T = Classroom Assessment Scoring System-Toddler.

Most Children Are in Classrooms of Mid-Range Quality

Using the CLASS-T, we found that 2- and 3-year-old children were in Early Head Start classrooms that scored in the mid-range of quality (scores of 3 to 5) in the domains of Emotional and Behavioral Support and Engaged Support for Learning (Table 3). Classrooms were strongest in the Emotional and Behavioral Support domain, including the dimensions of Positive Climate, Teacher Sensitivity, Regard for Child Perspectives, and Behavior Guidance; classrooms were rated in the low range on Negative Climate, indicating that interactions characterized by negativity were infrequently observed. Classrooms scored lower in the area of Engaged Support for Learning, including the dimensions of Facilitation of Learning and Development, Quality of Feedback, and Language Modeling. At age 3, mean scores in the area of Language Modeling were in the low range.

These findings are similar to those found in other work on the quality of interactions between teachers and children in center-based programs. For example, in the pilot study of the adapted toddler measure (Thomason and La Paro 2009), Language Modeling was likewise rated the lowest of all dimensions (mean = 1.9) in observations of classrooms serving children between 15 and 36 months of age. Additionally, Baby FACES findings indicating lower instructional quality compared to emotional aspects of the classroom are consistent with those in studies of older preschool children (for example, the Head Start Family and Child Experiences Survey, Aikens et al. 2010; and the Chicago Program Evaluation Project, Ross et al. 2008) and suggest that providers face challenges in their attempts to offer high quality instructional support to children, including facilitating activities that support children's learning and development, providing individualized feedback to promote children's understanding of concepts, providing opportunities for children to use language, and engaging in conversations that extend children's language skills. Notably, across all the CLASS-T dimensions, these Early Head Start classroom scores compare favorably with those reported by the CLASS-T authors in a number of reports (personal communication with Robert Pianta and Karen La Paro, November 2, 2011).

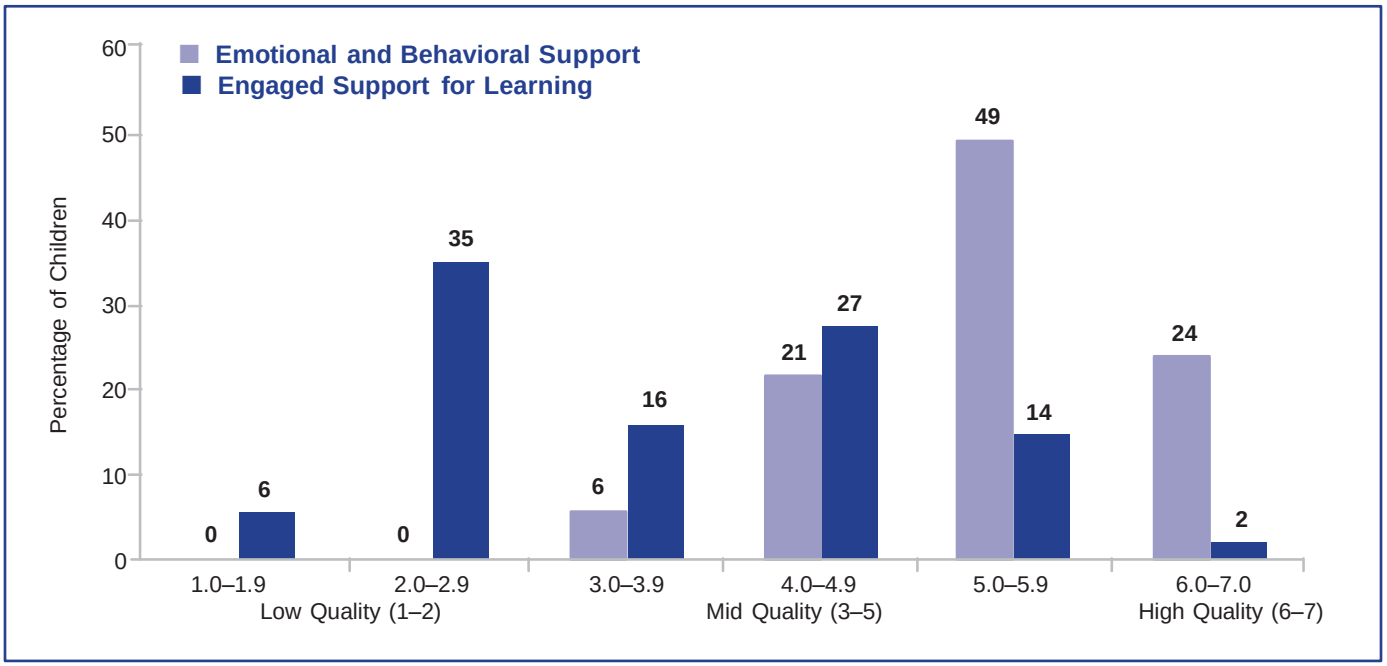
Findings suggest providers are strongest in interactions that provide emotional and behavioral support but face more challenges in their attempts to offer high quality support for learning.

Using the developer-provided definitions of the CLASS-T dimension scores, all 2-year-old children in Baby FACES were in classrooms rated in the mid-to-high range in the domain of Emotional and Behavioral Support (Figure 1). One-quarter (24 percent) were in classrooms rated as 6 or higher. In contrast, only 59 percent of children were in classrooms receiving scores in the mid-to-high range for Engaged Support for Learning. Nearly half (41 percent) of all children were in classrooms scoring in the low range, with far fewer children (2 percent) in classrooms rated as 6 or higher. The distribution of scores was similar for observations conducted in classrooms when Baby FACES children were 3 years old (see Figure A.1 in Appendix A).

Psychometric Properties of the CLASS-T in Baby FACES

In Baby FACES, we addressed a number of conceptual and analytic questions related to the psychometric properties of the CLASS-T. While there is some evidence of the validity of the measure based on a pilot study of 30 classrooms (Thomason and La Paro 2009), Baby FACES represents the first large-scale effort sufficiently powered to examine the underlying factor structure of the CLASS-T and assess its reliability and validity. In this section, we report findings of factor analyses using the CLASS-T data and present an overview of the relationships among observed quality, teacher and program characteristics, and children's language and social-emotional competence.

Figure 1. Distribution of CLASS-T Domain Scores in Classrooms at Age 2



Source: Spring 2010 Baby FACES Classroom Observations. Reported observations are for children in the 1-year-old Cohort at age 2. Estimates are at the child level.
Note: The overall mean score was 5.3 for Emotional and Behavioral Support and 3.6 for Engaged Support for Learning.
Sample size = 323 children in 220 classrooms.
CLASS-T = Classroom Assessment Scoring System-Toddler.

Baby FACES CLASS-T Factor Analyses

The CLASS-T pilot manual used in Baby FACES (Pianta et al. 2010) defines classroom interactions along eight dimensions grouped into three domains: (1) Emotional Support (Positive Climate, Negative Climate, Teacher Sensitivity, and Regard for Child Perspectives), (2) Classroom Organization (Behavior Guidance and Facilitation of Learning and Development), and (3) Instructional Support (Quality of Feedback and Language Modeling). In 2012, La Paro and colleagues released a revised iteration of the instrument. Among the most significant changes to the instrument is the reconceptualization of the dimensions into only two domains. Specifically, the authors proposed a revised organizational structure in which the dimension of Facilitation of Learning and Development—previously part of the Classroom Organization domain—became a component of the Instructional Support domain. The dimension of Behavior Guidance moved to the Emotional Support domain. Thus, the final instrument assesses teacher-child interactions along the same eight quality dimensions, but it offers two rather than three overarching domains of toddlers' classroom experience: (1) Emotional and Behavioral Support and (2) Engaged Support for Learning (La Paro et al. 2012). The authors provide support for the validation of the original organizational structure using data from more than 3,000 classrooms ranging from preschool to 5th grade (Hamre and Pianta 2007). However, there is less available evidence to support the validity of the instrument's organizational structure with classrooms serving toddlers. We thus conducted analyses to cross-validate the underlying factor structure of the CLASS-T in our sample and guide our approach to scoring at the domain level.

Using data from 217 Early Head Start classrooms serving 2-year-old children, we conducted a principal components factor analysis with Varimax rotation using the eight component dimensions of the CLASS-T (Table 4). We

The CLASS-T measures two aspects of teacher–child interaction in classrooms serving toddlers: Emotional/ Behavioral Support and Engaged Support for Learning

Table 4. CLASS T Dimension Scores Load Into a Two Factor Solution, Age 2

	Factor	
Dimension	**Factor 1: Emotional and Behavioral Support**	**Factor 2: Engaged Support for Learning**
Positive Climate	0.74	0.30
Negative Climate	0.71	-0.07
Teacher Sensitivity	0.79	0.47
Regard for Child Perspectives	0.79	0.45
Behavior Guidance	0.82	0.37
Facilitation of Learning and Development	0.41	0.83
Quality of Feedback	0.18	0.94
Language Modeling	0.18	0.93
Mean (Standard Deviation)	5.30 (0.80)	3.54 (1.17)
Standardized Alpha	0.89	0.94
% of Total Variance Explained	63.15	15.95
Sample Size	**217**	**217**

Source: Spring 2010 Baby FACES Classroom Observations. Includes observations of classrooms of the 1-year-old Cohort at age 2.
Note: Three of the 220 classrooms in the study were missing data for at least one dimension and were excluded from the factor analysis. As a result, the sample size was 217 classrooms.
[a] Standardized alpha calculated among items with loadings of 0.55 or higher.
CLASS-T = Classroom Assessment Scoring System-Toddler.

a priori retained two factors for rotation.[4] As shown in Table 4, item loadings ranged from 0.71 to 0.82 for factor 1 and from 0.83 to 0.94 for factor 2. These factors align closely with the domains identified by the CLASS-T developers in the updated manual (La Paro et al. 2012), presented in Table 2 above. We repeated this analysis with 235 classroom observations when children were 3 years old collected during spring 2011 and 2012. The pattern of findings closely mirrored those obtained in analysis of classrooms serving 2-year-old children (see Table A.1 in Appendix A).

Correlations among the dimension scores at age 2 are presented in Table 5. For Factor 1 (Emotional and Behavioral Support), the Positive Climate, Teacher Sensitivity, Regard for Child Perspectives and Behavior Guidance dimensions were strongly and significantly correlated (r = 0.63 to 0.86); associations between Negative Climate and the other dimensions comprising this factor were moderate in magnitude (r = -0.30 to -0.45). Associations among the three component dimensions corresponding to Factor 2 (Engaged Support for Learning) were strong and statistically significant (r = 0.80 to 0.91). Associations at age 3 are presented in Table A.2 in Appendix A.

At each age, the two-factor solution demonstrated high internal consistency and explained a substantial portion of common variance. Thus, we created two composite scores, Emotional and Behavioral Support and Engaged Support for Learning, by

[4] In accord with the organizational structure of the domains in the CLASS-T pilot manual, we also examined a three-factor solution. Behavior Guidance, however, did not emerge as a separate factor when a third factor was retained for rotation. Instead, Negative Climate loaded singly onto the additional factor, likely due to the low variability observed on this dimension.

Table 5. Correlations Among CLASS T Dimension Scores in Classrooms at Age 2

	1	2	3	4	5	6	7
Factor 1: Emotional and Behavioral Support							
1. Positive Climate							
2. Negative Climate	0.30***						
3. Teacher Sensitivity	0.71***	-0.43***					
4. Regard for Child Perspectives	0.63***	-0.44***	0.86***				
5. Behavior Guidance	0.70***	-0.45***	0.80***	0.78***			
Factor 2: Engaged Support for Learning							
6. Facilitation of Learning and Development							
7. Quality of Feedback						0.82***	
8. Language Modeling						0.80***	0.91***
Sample Size	**218–220**	**217–218**	**219–220**	**219**		**220**	**220**

Source: Spring 2010 Baby FACES Classroom Observations. Includes observations of classrooms of the 1-year-old Cohort at age 2.
**p < 0.01; ***p < 0.001.

averaging the component items corresponding to each of the derived factors.[5] These estimates are consistent with those reported for the domain of Emotional Climate (Cronbach's alpha = 0.88) in a pilot study of the adapted measure conducted in 30 toddler classrooms (Thomason and La Paro 2009).[6]

Concurrent Associations with Measures of Quality

In Baby FACES, we found some significant concurrent associations between CLASS-T domain scores and other indicators of classroom quality. These associations were generally small to modest in size and all were in the expected direction (Table 6).[7] The pattern of associations varied by age, with few significant associations emerging at age 3 for Emotional and Behavioral Support and no associations for Engaged Support for Learning. At age 2, teachers' experience working with infants and toddlers and having a Child Development Associate (CDA) credential were positively related to Emotional and Behavioral Support scores (r = 0.16 to 0.21). At age 3, state awarded credentialing was related to higher Emotional and Behavioral Support scores (r = 0.22). Emotional and Behavioral Support scores at ages 2 and 3 were related to teachers' self-reported depressive symptoms, with higher ratings of depressive symptomatology related to lower classroom quality scores (r = -0.21 to -0.23). Teachers who reported better relationships with parents had classrooms characterized by higher scores on Emotional and Behavioral Support (r = 0.19) and Engaged Support for Learning (r = 0.22) at age 2. Finally, teacher turnover and job satisfaction was associated with classroom quality. At age 2, domain scores were higher when teachers reported a greater likelihood of returning to their jobs in the coming year (r = 0.17 to 0.20). In addition, high teacher turnover rates were associated with lower Emotional and Behavior Support domain scores (r = -0.17). At age 3, quality scores were lower when program directors reported staff left the program for higher compensation or improved benefits (r = -0.21).

[5] Based on the findings of this analysis and at the suggestion of the CLASS-T developers, we refer to the two derived composite domain scores throughout this report using these modified descriptors.

[6] The measure used in the Thomason and La Paro (2009) pilot study included only six of the eight component dimensions represented in the version used in Baby FACES. Consequently, the study authors did not report findings for the dimensions of Facilitation of Learning and Development and Quality of Feedback, or the resulting composite domain score (Instructional Support) derived from these dimensions. The Emotional Climate domain score reported by the authors is similar to the composite measure of Emotional and Behavioral Support derived in Baby FACES; a key dissimilarity is the inclusion of Behavior Guidance in the Emotional and Behavioral Support composite score.

[7] Analyses examined associations between the two classroom quality factors and a wide range of program and staff characteristics. These include: whether the program has unfilled positions; program turnover of home visitors, teachers, and management staff; teachers' years of experience working with young children; educational level (including categorical educational level and dummy codes for whether the teacher has a high school degree plus some college, an A.A., or a B.A.), training (whether the teacher has a degree in early childhood education and whether he or she is currently participating in child-care-related training), and credentials (whether teacher has a CDA or state-awarded credential); depressive symptoms (whether the teacher has moderate to severe levels of depressive symptoms); job satisfaction; reported relationship with the parent; and parent involvement in the program. Only relationships that were statistically significant are presented in Table 6.

Table 6. Concurrent Correlations of CLASS T Factor Scores and Measures of Quality

	Age 2		Age 3	
	Emotional/ Behavioral Support	Engaged Support for Learning	Emotional/ Behavioral Support	Engaged Support for Learning
Years Teaching Infants/Toddlers	0.16**	0.03	0.04	-0.01
Teacher Has State Certificate	0.08	-0.03	0.22***	0.06
Teacher Has CDA Credential	0.21**	0.10	0.05	-0.04
Teacher Depression Score	-0.21**	-0.03	-0.23***	-0.08
Relationship with Parent	0.19**	0.22**	0.08	0.03
Teacher Is Very Likely to Return to Job Next Year	0.20**	0.17**	0.11	0.05
Number of Teachers Who Left Program in Past Year	-0.17*	-0.04	-0.05	0.05
Staff Left for Higher Compensation	-0.05	-0.14	-0.21**	0.00
Sample Size	**175-220**	**175-220**	**197-241**	**197-241**

Source: Spring 2010, 2011, and 2012 Baby FACES Classroom Observations and Teacher Interviews; 2010 and 2011 Program Director Interviews and Program Director Self-Administered Questionnaires.
Note: Reported observations are for the 1-year-old Cohort at age 2 and for the 1-year-old and Newborn Cohorts at age 3. Only characteristics yielding statistically significant correlations are presented.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Although small in magnitude, CLASS-T domain scores were significantly associated with features of quality at age 2 and with a few measures of child language and social-emotional competence.

Concurrent Associations with Child Developmental Outcomes

A few significant associations emerged between the CLASS-T and concurrent child language and social-emotional outcomes. These were generally small in magnitude and most were in the expected direction (Table 7).[8] Notably, the pattern of significant associations was not consistent across the CLASS-T domains or across the observed ages. Emotional and Behavioral Support domain scores were significantly and positively associated with children's language outcomes, including PLS-4 English standard scores at age 2 ($r = 0.22$) and teacher-reported Spanish CDI-III vocabulary comprehension scores at age 3 ($r = 0.30$). However, a negative association between Emotional and Behavioral Support scores and teacher-reported English CDI-III vocabulary comprehension scores emerged at age 3, with higher quality associated with lower English comprehension scores ($r = -0.12$), although the association was small in magnitude. Scores on this domain of quality were also significantly associated with children's social-emotional competencies, including teacher-reported BITSEA Competence scores at age 2 ($r = 0.15$), BITSEA Problem scores at age 3 ($r = -0.20$), BRS Emotional Regulation scores at age 3 ($r = 0.14$), and teacher-reported BPI Externalizing Behavior scores at age 3 ($r = -0.13$).

[8] Analyses examined associations between the two classroom quality factors and a wide range of child language and social-emotional outcomes, obtained using a combination of direct child assessments, observer ratings, and parent and teacher reports. Language measures include the ASQ-3, PLS-4, PPVT-4, Early Communication Indicator (ECI), and English and Spanish CDI-III. Assessments of children's social-emotional competence include the BITSEA, PCI Child Rating Scales for the Two-Bag Assessment, BRS, and BPI. Only relationships that were statistically significant are presented in Table 7.

Table 7. Concurrent Correlations of Classroom Quality with Child Outcomes

	Age 2		Age 3	
	Emotional/ Behavioral Support	Engaged Support for Learning	Emotional/ Behavioral Support	Engaged Support for Learning
Language Outcomes				
ASQ-3 Communication IRT Score	0.03	0.14**	-0.04	-0.03
PLS-4 English Standard Score	0.22***	0.23***	0.06	0.09
CDI-III Vocabulary Comprehension Raw Score	-0.01	0.09	-0.12*	0.08
CDI-III Vocabulary Production Raw Score	-0.02	0.12*	-0.10	0.07
Spanish CDI-III Vocabulary Comprehension Raw Score	0.10	-0.10	0.30*	-0.30*
Spanish CDI-III Vocabulary Production Raw Score	0.16	0.09	0.22	-0.35*
Social-Emotional Outcomes				
BITSEA Problem Domain Raw Score	-0.09	-0.06	-0.20***	-0.07
BITSEA Competence Domain Raw Score	0.15**	0.16**	0.06	0.10
PCI Rating Scales—Sustained Attention with Objects	0.05	0.15**	-0.04	0.00
BRS Orientation/ Engagement Raw Score	0.05	0.08	0.10	0.18**
BRS Emotional Regulation Raw Score	0.01	0.05	0.14*	0.21***
BPI Externalizing Behaviors	n.a	n.a	-0.13*	0.13
Sample Size	**234–309**	**234–309**	**205–298**	**205–298**

Source: Spring 2010, 2011, and 2012 Baby FACES Classroom Observations, Teacher Interviews, Parent Interviews, Two-Bag Task, and Direct Child Assessments. Estimates are at the child level.

Note: Reported observations are for children in the 1-year-old Cohort at age 2 and for children in the 1-year-old and Newborn Cohorts at age 3. Only outcomes yielding statistically significant correlations are presented. The CDI, BITSEA, and BPI scores reported here are based on teacher reports; ASQ-3 Communication scores are parent-reported.

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

n.a. = not applicable.

Scores on the Engaged Support for Learning domain were positively associated with children's language outcomes at age 2, including ASQ-3 communication scores ($r = 0.14$), PLS-4 English standard scores ($r = 0.23$), and teacher-reported English CDI-III vocabulary production scores ($r = 0.12$). At age 3, the associations between Engaged Support for Learning and teacher-reported Spanish CDI-III scores were negative, with higher quality associated with lower comprehension and production scores ($r = -0.30$ and -0.35, respectively). In terms of children's social-emotional competencies, scores on this domain of quality were associated with teacher-reported BITSEA Competence scores ($r = 0.16$) and ratings of Sustained Attention from the PCI Rating Scales for the Two-Bag Assessment ($r = 0.15$) at age 2, and BRS Orientation/Engagement and Emotional Regulation scores ($r = 0.18$ and 0.21, respectively) at age 3.

Predictive Associations with Child Developmental Outcomes

We also examined associations between CLASS-T scores age 2 and children's later language and social-emotional competencies (at age 3). Few consistent predictive relationships emerged. In terms of children's language skills, Emotional and Behavioral Support and Engaged Support for Learning scores at age 2 were significantly associated with children's language outcomes at age 3, including PLS-4 English standard scores (r = 0.19 for both quality domains) and PPVT-4 standard scores (r = 0.17 and 0.29, respectively). However, there were also relationships in the opposite direction, indicating higher quality was associated with poorer outcomes. Emotional and Behavioral Support domain scores were significantly (but negatively) associated with teacher-reported English CDI-III vocabulary comprehension scores (r = -0.14). Scores on Engaged Support for Learning were associated with higher (more) parent-reported BPI Internalizing and Externalizing Behavior scores (r = 0.16 and 0.19, respectively).

Summary

Baby FACES represents the first large-scale research effort that provides an opportunity to examine the reliability and validity of the CLASS-T observational tool. Observers were certified in accord with criteria established by the instrument's developers and maintained or exceeded this level of reliability throughout the field period. Analyses revealed two domains of effective teacher-child interaction in classrooms serving 2- and 3-year-old children: (1) Emotional and Behavioral Support and (2) Engaged Support for Learning. Most children in center-based programs were in classrooms of mid-range quality. Providers were strongest in the area of Emotional and Behavioral Support; lower scores were observed in the area of Engaged Support for Learning, suggesting that providers face more challenges in their attempts to offer high quality instructional support to children. Although generally small in magnitude, CLASS-T domain scores were significantly associated with features of quality at age 2 and with a few measures of child language and social-emotional competence. The pattern of associations, however, was not consistent across the CLASS-T domains or across the observed ages. Nonetheless, the findings in Baby FACES are generally consistent with findings from classroom observations with older children, including some small to modest associations among quality measures and between CLASS observations and children's outcomes (Moiduddin et al. 2012). Furthermore, there is emerging theory that particular quality supports are more likely to be associated with outcomes in the same domain (e.g., social and behavioral supports are more likely to be associated with social-emotional outcomes compared to language outcomes; Zaslow et al. 2010), which may account for some of the inconsistencies in findings across CLASS-T domains.

References

Administration for Children and Families (ACF). "Head Start Program Performance Standards." Cited in *Federal Register*, October 1, 2009. Retrieved from http://eclkc.ohs.acf.hhs.gov/hslc/standards/Head%20Start%20Requirements/45%20CFR%20Chapter%20XIII on February 7, 2012.

Aikens, N., L. Tarullo, L. Hulsey, C. Ross, J. West, and Y. Xue. "A Year in Head Start: Children, Families and Programs." Report submitted to Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Washington, DC: Mathematica Policy Research, October 2010.

Bayley, N. *Bayley Scales of Infant Development–Second Edition*. San Antonio, TX: The Psychological Corporation, 1993.

Briggs-Gowan, M, and A. Carter. *The Brief Infant–Toddler Social and Emotional Assessment (BITSEA)*. San Antonio, TX: Harcourt Assessment, 2006.

Dunn, L.M., and D.M. Dunn. *Peabody Picture Vocabulary Test–Fourth Edition*. Bloomington, MN: NCS Pearson, Inc., 2007.

Elicker, J., I. Noppe, L. Noppe, and C. Fortner-Wood. "The Parent-Caregiver Relationship Scale: Rounding Out the Relationship System in Infant Care." *Early Education and Development*, vol. 8, no. 1, 1997, pp. 83–100.

Fenson, L., S. Pethick, C. Renda, J.L. Cox, P.S. Dale, and J.S. Reznick. "Short-Form Versions of the MacArthur Communicative Development Inventories." *Applied Psycholinguistics,* vol. 21, 2000, pp. 95–115.

Hamre, B.K., and R.C. Pianta. "Learning Opportunities in Preschool and Early Elementary Classrooms." In *School Readiness and the Transition to Kindergarten in the Era of Accountability*, edited by R.C. Pianta, M.J. Cox, and K.L. Snow. Baltimore, MD: Paul H. Brookes Publishing Co., 2007.

La Paro, K.M., B.K. Hamre, and R.C. Pianta. *Classroom Assessment Scoring System Manual, Toddler (CLASS-T).* Baltimore, MD: Paul H. Brookes Publishing, Co., 2012.

Mathematica Policy Research. "Baby FACES and Building Strong Families: Video Coding Manual for Two Bags Task." Unpublished scales adapted from Brady-Smith et al. "36-Month Child-Parent Interaction Rating Scales for the Three-Bag Assessment. Early Head Start Research and Evaluation Project," 2000; and Cox. "Qualitative Ratings for Parent-Child Interaction at 24–36 Months of Age," 1997, 2010.

Moiduddin, E., N. Aikens, L. Tarullo, J. West, Y. Xue. (2012). Child Outcomes and Classroom Quality in FACES 2009. OPRE Report 2012-37a. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Peterson, J. L., and N. Zill. "Marital Disruption, Parent-Child Relationships, and Behavior Problems in Children." *Journal of Marriage and Family*, vol. 48, 1986, pp. 295–307.

Pianta, R.C., K.M. La Paro, and B.K. Hamre. *Classroom Assessment Scoring System (CLASS)*. *Pre-K version*. Baltimore, MD: Paul H. Brookes Publishing Co., 2008.

Pianta, R.C., K.M. La Paro, and B.K. Hamre. *Classroom Assessment Scoring System Pilot Toddler Manual (CLASS-T)*. Charlottesville, VA: Teachstone, Inc., 2010.

Radloff, L.S. "The CES-D Scale: A Self-Report Depression Scale for Research in the General Population." *Applied Psychological Measurement*, vol. 1, 1977, pp. 385–401.

Ross, C.E., J. Mirowsky, and J. Huber. "Dividing Work, Sharing Work, and in Between: Marriage Patterns and Depression." *American Sociological Review*, vol. 48, 1983, pp. 809–823.

Ross, C. M., E. Moiduddin, C. Meagher, and B. L. Carlson. "The Chicago Program Evaluation Project: A Picture of Early Childhood Programs, Teachers, and Preschool-Age Children in Chicago." Final external report submitted to The Erickson Institute, The Chicago Department of Children and Youth Services, and the Chicago Public Schools. Princeton, NJ: Mathematica Policy Research, December 2008.

Squires, J., E. Twombly, D. Bricker, and L. Potter. *ASQ-3 User's Guide*. Baltimore, MD: Paul H. Brookes Publishing Co., 2009.

Thomason, A.C., and K.M. La Paro. "Measuring the Quality of Teacher-Child Interactions in Toddler Child Care." *Early Education and Development*, vol. 20, no. 2, 2009, pp. 285–304.

Vagh, S.B., J. Mançilla-Martinez, and B.A. Pan. "An Upward Extension Based on the Inventarios del Desarrollo de Habilidades Comunicativas Short Form for Use with Children of Ages 30-37 Months from Low-Income Bilingual Families." Unpublished manuscript.

Vogel, C.A., K. Boller, Y. Xue, R. Blair, N. Aikens, A. Burwick, Y. Shrago, B.L. Carlson, L. Kalb, L. Mendenko, J. Cannon, S. Harrington, and J. Stein. "Learning As We Go: A First Snapshot of Early Head Start Programs, Staff, Families, and Children." OPRE Report 2011-7. Washington, DC: U.S. Department of Health and Human Services, Office of Planning, Research and Evaluation, Administration for Children and Families, 2011.

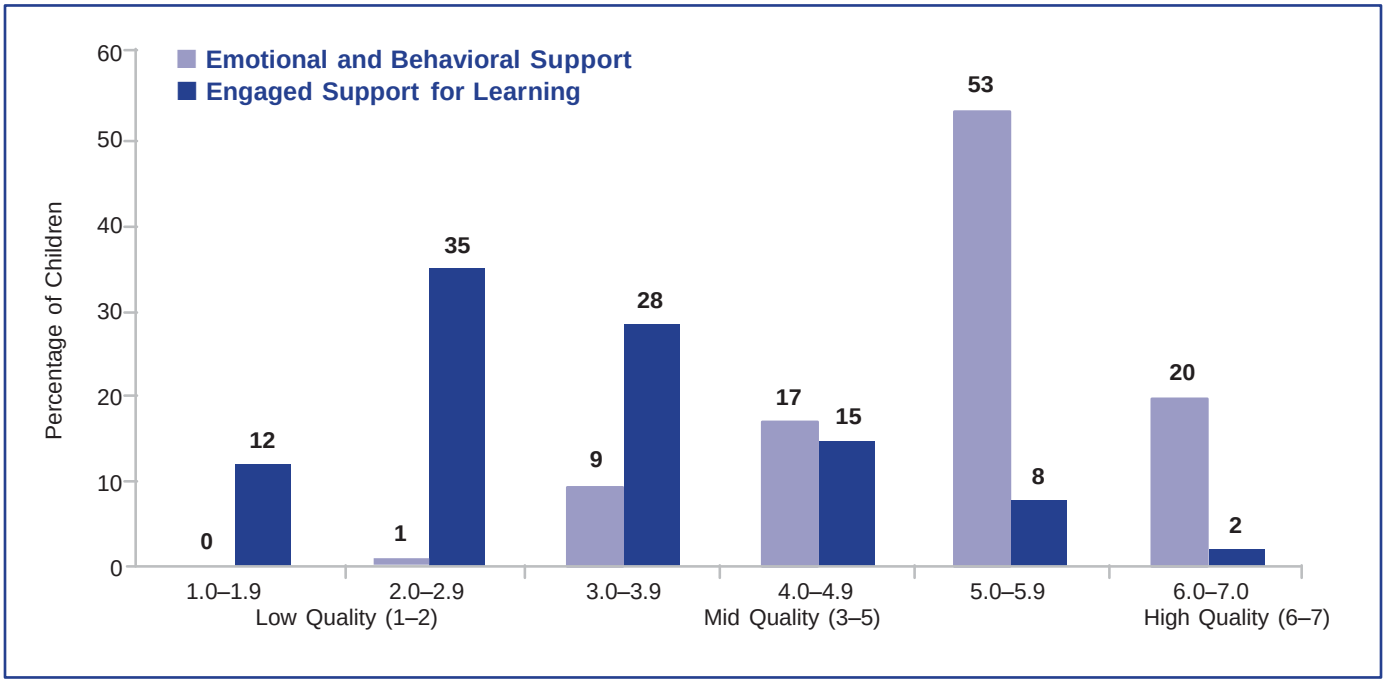
Zaslow, M., R. Anderson, Z. Redd, J. Wessel, L. Tarullo, and M. Burchinal. "Quality Dosage, Thresholds, and Features in Early Childhood Settings: A Review of the Literature." Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation, August 2010.

Zimmerman, I.L., V.G. Steiner, and R.E. Pond. *Preschool Language Scale, Fourth Edition (PLS-4)*. San Antonio, TX: Pearson, 2002a.

Zimmerman, I.L., V.G. Steiner, and R.E. Pond. *Preschool Language Scale, Fourth Edition (PLS-4) Spanish Edition*. San Antonio, TX: Pearson, 2002b.

Appendix A

Figure A.1. Distribution of CLASS-T Domain Scores in Classrooms at Age 3



Source: Spring 2011 and 2012 Baby FACES Classroom Observations. Reported observations are for children in the 1-year-old and Newborn Cohorts at age 3. Estimates are at the child level.
Note: The overall mean score was 5.3 for Emotional and Behavioral Support and 3.3 for Engaged Support for Learning.
Sample size = 312 children in 235 classrooms.
CLASS-T = Classroom Assessment Scoring System-Toddler.

Table A.1. CLASS T Dimensions Load into a Two Factor Solution, Age 3

Dimension	Factor	
	Factor 1: Emotional and Behavioral Support	Factor 2: Engaged Support For Learning
Positive Climate	0.85	0.30
Negative Climate	0.75	-0.08
Teacher Sensitivity	0.81	0.46
Regard for Child Perspectives	0.82	0.40
Behavior Guidance	0.78	0.40
Facilitation of Learning and Development	0.37	0.88
Quality of Feedback	0.18	0.94
Language Modeling	0.18	0.94
Mean (Standard Deviation)	5.24 (0.88)	3.19 (1.18)
Standardized Alpha	0.91	0.96
% of Total Variance Explained	64.48	18.31
Sample Size	**235**	**235**

Source: Spring 2011 and 2012 Baby FACES Classroom Observations. Includes observations of classrooms of the 1-year-old and Newborn Cohorts at age 3.

[a] Standardized alpha calculated among items with loadings of 0.55 or higher.

CLASS-T = Classroom Assessment Scoring System-Toddler.

Table A.2. Correlations Among CLASS T Dimension Scores in Classrooms at Age 3

	1	2	3	4	5	6	7
Factor 1: Emotional and Behavioral Support							
1. Positive Climate							
2. Negative Climate	-0.48***						
3. Teacher Sensitivity	0.85***	-0.43***					
4. Regard for Child Perspectives	0.78***	-0.46***	0.86***				
5. Behavior Guidance	0.75***	-0.45***	0.77***	0.77***			
Factor 2: Engaged Support for Learning							
6. Facilitation of Learning and Development							
7. Quality of Feedback						0.87***	
8. Language Modeling						0.87***	0.90***
Sample Size	**231–333**	**231–233**	**231–233**	**231**		**233**	**233**

Source: Spring 2011 and 2012 Baby FACES Classroom Observations. Includes observations of classrooms of the 1-year-old and Newborn Cohorts at age 3.
**$p < 0.01$; ***$p < 0.001$.

To Learn More

Amy Madigan, *Project Officer*

Office of Planning, Research and Evaluation; Administration for Children and Families; Department of Health and Human Services

acf.hhs.gov/programs/opre

Contract Number: HHSP23320072914YC

Cheri A. Vogel, *Project Director*

Mathematica Policy Research

600 Alexander Park

Princeton, NJ 08540

Suggested Citation

E. Bandel, N. Aikens, C.A. Vogel, K. Boller, and L. Murphy (2014). Observed Quality and Psychometric Properties of the CLASS-T in the Early Head Start Family and Child Experiences Survey. OPRE Technical Brief 2014-34. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Disclaimer

The views expressed in this publication do not necessarily reflect the views or policies of the Office of Planning, Research and Evaluation, the Administration for Children and Families, or the U.S. Department of Health and Human Services.