



Published in final edited form as:

*J Psychopathol Behav Assess.* 2015 September ; 37(3): 493–503. doi:10.1007/s10862-015-9478-x.

## Psychometric Evaluation of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA) in a Predominately Hispanic, Low-Income Sample

Gabriela M. Hungerford, Dainelys Garcia, and Daniel M. Bagner

Department of Psychology, Florida International University, 11200 SW 8th St., AHC I Room 241, Miami, FL 33174, USA

Daniel M. Bagner: dbagner@fiu.edu

### Abstract

The goal of the present study was to evaluate the psychometric properties of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA) with 12- to 15-month-old infants from predominately Hispanic, low-income families. Mothers of 144 infants were screened at a pediatric clinic as part of a larger study examining a brief home-based intervention for infants at-risk for behavior problems. Reliability was good for the BITSEA problem scale in all analyses and acceptable for the BITSEA competence scale in most analyses. Discriminative validity was supported by scores on the BITSEA competence scale significantly predicting delayed status on all ASQ-3 subscales. BITSEA problem scale scores significantly predicted scores on the total problems scale of the Child Behavior Checklist, supporting predictive validity. Analyses revealed a main effect of group on BITSEA problem scale scores, providing preliminary support for sensitivity to change for the BITSEA problem scale. Results support the BITSEA as an effective screening tool for use with young infants, Hispanic and Spanish-speaking populations, and low-income families.

### Keywords

Infant behavior; Assessment; Psychometrics; Screening

---

Infants and toddlers engage in many behaviors that might be bothersome for parents or caregivers (e.g., tantrums) but are representative of typical development. However, for some young children, these behaviors may be early signs of more serious long-term problems (Campbell et al. 2000). Children who show high levels of disruptive behavior by age 2 years are at higher risk for developing more severe problems later in life (Shaw et al. 2003). Additionally, early behavioral and emotional problems have been demonstrated to be stable

---

Correspondence to: Daniel M. Bagner, dbagner@fiu.edu.

#### Conflict of Interest

Gabriela M. Hungerford, Dainelys Garcia, and Daniel M. Bagner declared that they have no conflict of interest.

#### Experiment Participants

All experimental protocols were approved by the appropriate institutional review boards and meet the guidelines of the responsible government agency. Informed consent was obtained from all adult participants.

over time (Briggs-Gowan et al. 2006; Mathiesen and Sanson 2000). In particular, proximal family and child risk factors (e.g., maternal depression, child fearlessness) occurring within the first 2 years of the child's life are associated with a trajectory of chronic conduct problems (Shaw et al. 2005). When these problems are detected early, interventions have been shown to successfully improve outcomes (Shaw et al. 2006). Therefore, it is critical to identify signs of behavioral and emotional problems as early in life as possible. Furthermore, young children from Hispanic (Gross et al. 1999) and low-income (Qi and Kaiser 2003) backgrounds have been found to have a higher incidence of behavior problems, suggesting it may be especially important to detect early signs of these problems in these at-risk populations.

Similar to early behavioral and emotional problems, developmental delay also places young children at higher risk for negative outcomes. For example, young children with developmental delay are more likely to experience later difficulties with self-regulation and activity level (Olson et al. 2002), peer relationships (Guralnick et al. 2007), behavioral functioning (Feldman et al. 2000), and academic achievement (Sonnander and Claesson 1999). Additionally, young children who live in poverty are almost twice as likely as other children to exhibit developmental delay (Simpson et al. 2003). Given that Hispanic children are more likely to live in poverty (Huston et al. 1994), it is critical to identify children with developmental delay as early as possible, particularly in children from Hispanic and low-income families (Sonnander and Claesson 1999).

In order to improve efforts to identify early signs of behavioral and emotional problems and developmental delay in young children, it is recommended that screening take place routinely in pediatric settings (American Academy of Pediatrics 2006). However, 83 % of pediatricians report time limitations as a barrier to conducting standardized behavioral, emotional, and developmental screenings for young children (Sand et al. 2005). Given the importance of early identification of behavioral and emotional problems and developmental delay, particularly among children from minority and disadvantaged backgrounds, and the time constraints reported by pediatricians, the purpose of this study was to examine the psychometric properties of a brief screening tool for behavioral and emotional problems and competence in infants between 12- and 15-months-old from predominately Hispanic and low-income families.

The Brief Infant-Toddler Social and Emotional Assessment (BITSEA; Briggs-Gowan and Carter 2002) is a 42-item parent-report questionnaire used to screen for social-emotional and behavioral problems and developmental delay in 12- to 36-month-olds. The brief 5 to 10 min completion time makes the BITSEA ideal for administration in pediatric primary care and other related settings with time constraints and can help identify children that may benefit from a more comprehensive evaluation and subsequent intervention. Items for the BITSEA were selected from the more comprehensive Infant-Toddler Social and Emotional Assessment (ITSEA; Carter and Briggs-Gowan 2006a), a 166-item, parent-report assessment measure used to identify social-emotional and behavioral problems and competencies. Criteria for BITSEA item selection included the 12 items with the highest loading on each of the ITSEA subscales, 1 item selected due to clinical significance, 1 item selected as most representative of the prosocial-peer parameter, and 28 items chosen by a

majority of an expert panel (Briggs-Gowan and Carter 2002). Specifically, relevant theories and literature were reviewed to ensure items were sensitive to early emerging problem behaviors, developmentally appropriate, and clinically sensitive. The BITSEA yields two scores: a problem total score and a competence total score. Cutpoints for each scale are based on age and sex and intended to identify 25 % of infants at highest risk for problems and 10–15 % of infants at highest risk for developmental delay on the problem and competence scores, respectively (Briggs-Gowan et al. 2004). The BITSEA also includes two separate parental “worry” items in which parents rate their concerns about their child’s behavior and language on a 5-point scale from 1 = *not at all worried* to 5 = *extremely worried* (Briggs-Gowan et al. 2004).

The psychometric properties of the BITSEA were examined in a representative healthy birth cohort of 1237 young children aged 12–36 months randomly selected from birth records (Briggs-Gowan et al. 2004). Participants in this study were mostly non-Hispanic white infants (66.3 %) from two-parent families (71.7 % married), and most parents reported an education beyond high school (73.7 %). Internal consistency was acceptable for the problem scale ( $\alpha=.79$ ) and marginal for the competence scale ( $\alpha=.65$ ). Test-retest reliability was excellent for both the problem ( $r=.87$ ) and competence ( $r=.85$ ) scales across 10 to 45 days, although one-year test-retest reliability estimates were lower for the problem ( $r=.65$ ) and competence ( $r=.53$ ) scales. Interrater reliability between parents was good for both the problem ( $r=.68$ ) and competence ( $r=.61$ ) scales. Interrater reliability between parents and child-care providers, however, was lower for both the problem and competence scales ( $r_s=.28$  and  $.59$ , respectively). Predictive validity for the BITSEA was examined using the Child Behavior Checklist for ages 1.5 to 5 years (CBCL/1.5–5; Achenbach and Rescorla 2001), a widely used measure of behavioral and emotional problems in young children, and the ITSEA. Findings suggested that BITSEA problem and competence scores significantly predicted CBCL/1.5–5 externalizing, internalizing, and total scores, as well as ITSEA scores in all four domains (externalizing, internalizing, dysregulation, competence) 1 year later (Briggs-Gowan et al. 2004).

In addition to the psychometric properties in a predominately Caucasian, educated sample from the United States, psychometric properties of the BITSEA have been examined with a Turkish sample. Specifically, Karabekiroglu and colleagues (2009) recruited a community sample of 462 young children between the ages of 12 and 42 months whose parents had applied to health centers for immunization on their behalf in Turkey, and mothers and fathers completed the Turkish translation of the BITSEA. Internal consistency was excellent for the problem scale ( $\alpha=.82$ ) and good for the competence scale ( $\alpha=.72$ ). Test-retest reliability over 15 to 30 days was adequate for mothers (problem scale,  $r=.49$ ; competence scale,  $r=.41$ ) and fathers (problem scale,  $r=.83$ ; competence scale,  $r=.60$ ), and interrater reliability between mothers and fathers was good for both the problem ( $r=.68$ ) and competence ( $r=.71$ ) scales. Additionally, BITSEA problem scale scores were related to the CBCL/2–3 (Achenbach 1992) internalizing, externalizing, and total scores, supporting predictive validity.

Similarly, Kruizinga and colleagues (2012) examined the psychometric properties of the Dutch version of the BITSEA in a sample of 3,170 two-year-olds who were invited by

health care organizations to well-child visits in the Netherlands. Internal consistency and test-retest reliability (across 13–77 days) were adequate for the problem scale ( $\alpha=.76$ ;  $r=.75$ ) and marginal for the competence scale ( $\alpha=.63$ ;  $r=.61$ ). Interrater reliability between parents and child care providers was marginal for both scales (problem scale,  $r=.30$ ; competence scale,  $r=.17$ ). In support of convergent validity, BITSEA problem scale scores were highly correlated with CBCL/1.5–5 total problem scores ( $r=.75$ ). In support of discriminative validity, BITSEA problem scores were able to distinguish between parents who reported being worried about their children’s behavior, emotion, or relationships and parents who did not report being worried about their child. Taken together, the existing research supports the reliability and validity of the English, Turkish, and Dutch versions of the BITSEA.

Despite the demonstrated psychometric properties of the BITSEA in English, Turkish, and Dutch samples, no research study to our knowledge has examined the psychometric properties of the Spanish version of the BITSEA. This is particularly relevant in the United States, where the number of Spanish speakers continues to rise, with projections of 40 million Spanish speakers by the year 2020 (Ortman and Shin 2011). Additionally, previous research has examined psychometric properties of the BITSEA with participants from predominately middle socioeconomic status backgrounds (Briggs-Gowan et al. 2004). However, given the increased risk for behavioral, emotional, and developmental problems in children from low socioeconomic status and ethnic minority backgrounds (Gross et al. 1999; Qi and Kaiser 2003), it is important to examine the psychometric properties in this population. Additionally, psychometric properties of the BITSEA have been examined with participants from either the full age range on the BITSEA (i.e., 12- to 36-month-olds; Briggs-Gowan et al. 2004; Karabekiroglu et al. 2009) or among 2-year-olds (Kruizinga et al. 2012), but no studies have examined psychometric properties with infants within the youngest BITSEA age band (i.e., 12 to 18 months). As previously discussed, this very young age range is particularly relevant given the importance of early detection of problems. Finally, to our knowledge, no studies have examined whether the BITSEA can detect intervention effects (i.e., sensitivity to change).

## Current Study

The purpose of the current study was to examine the reliability and validity of the BITSEA with infants in the lower end (12–15 months) of the youngest BITSEA age band and from predominately Hispanic, low-income families. Additionally, to our knowledge, the current study was the first to examine the psychometric properties of the published Spanish version of the BITSEA. To examine reliability, we measured estimates of both internal consistency and test-retest reliability for both the BITSEA problem and competence scales. To examine validity, we measured estimates of both discriminative and predictive validity.

Discriminative validity refers to the ability of a measure to discriminate between groups which are presumed to differ on the construct addressed by the measure (Foster and Cone 1995). Due to the intended use of the BITSEA competence scales to identify children in need of a more comprehensive developmental assessment, the BITSEA’s discriminative validity is highly relevant. Additionally, predictive validity is particularly important for screening measures like the BITSEA, which are typically used to identify infants for preventive intervention programs (Bagner et al. 2012).

Finally, we are the first to measure the sensitivity to change of the BITSEA problem scale. Sensitivity to change has been defined as the degree to which a measure is likely to reflect changes that occur as a result of participation in an intervention and represents an important psychometric property (Lambert and Hawkins 2004). Thus, we examined the ability of the BITSEA to evaluate intervention outcome consistent with research suggesting a measure that is sensitive to change can be useful in evaluating effectiveness, tracking progress, and treatment planning (McClendon et al. 2011).

## Method

### Participants

Participants were 144 mothers and their 12- to 15-month-old infant who participated in a screening as part of a larger study examining the effect of a brief home-based preventive intervention for infants. Families were screened during infant well and sick visits at a large pediatric primary care clinic (housed in a children's hospital) serving a majority of families (85 %) without private insurance in a predominately Hispanic community. To participate in the screening, mothers were required to speak and understand either English or Spanish, and infants were required to be between 12- and 15-months-old. For bilingual families, the mother was given the option to complete the screening in English or Spanish.

During the recruitment period, we approached 315 families at the pediatric clinic and provided a brief description of the study in English or Spanish. One hundred and forty-six primary caregivers (46.3 %) agreed to participate and provided written consent before proceeding with the screening. Primary caregivers were the infant's mother in all cases. All procedures were approved by the Institutional Review Boards at both the affiliated university and the children's hospital. Two families did not complete the screening measures (due to time constraints) and were excluded from the current study. In total, 144 mother-infant dyads were included in the current study. Mothers were on average 29.68-years-old ( $SD=5.67$  years; range=17–42 years), and their infants (56 % male) were on average 13.26-months-old ( $SD=1.27$ ). A majority of the mothers (78.5 %) reported their ethnicity as Hispanic, and a majority of the mothers (80.6 %) reported their infant's ethnicity as Hispanic. Approximately half of the mothers ( $n=65$ ; 45.1 %) completed the screening and subsequent assessments (see below) in Spanish, and the other half of the mothers ( $n= 79$ ; 54.9 %) completed the screening and subsequent assessments in English. Income data were available for 74 families. Thirty-eight percent of these families reported yearly incomes below the poverty line, more than double the 2012 U.S. poverty rate of 15 % (DeNavas-Walt et al. 2013). See Table 1 for a summary of participant demographics.

### Procedures

As part of the screening, mothers completed the BITSEA (Briggs-Gowan et al. 2004) and the Ages and Stages Questionnaire–Third Edition (ASQ-3; Squires and Bricker 2009), a measure of child developmental functioning. Additionally, mothers completed either the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler 1999) if they spoke English or the Escala de Inteligencia Wechsler Para Adultos–Third Edition (EIWA-III; Pons et al. 2008) if they spoke Spanish. In order to screen into the intervention study, infants had to be

rated by their mothers above the clinically significant range (i.e., > 75th percentile) on the BITSEA problem scale. Additionally, mothers had to receive an estimated IQ score  $\geq 70$  on the two-subtest (vocabulary and matrix reasoning) version of the WASI (for English-speaking mothers) or an average standard score  $\geq 4$  on the vocabulary and matrix reasoning subtests of the EIWA-III (for Spanish-speaking mothers).

Sixty families screened into the intervention study based on these inclusion criteria and were randomized to an intervention ( $n=31$ ) or standard care control group ( $n=29$ ). Two families screened in and were randomized to condition, but did not respond to follow-up phone calls or complete the baseline assessment. The remaining 58 families completed a Time 1 (baseline) assessment an average of 11.35 days ( $SD=7.30$ ) after the initial screening. During this assessment, mothers completed the ITSEA (Carter and Briggs-Gowan 2006a) and other measures of child and parent functioning. The first 42 items on the ITSEA are identical to the 42 items on the BITSEA, which allows for test-retest reliability for the BITSEA to be examined using these two measures with this sample. We informed families of their group status at the conclusion of the Time 1 assessment after all measures were completed.

Families in the intervention group received a brief, home-based adaptation of Parent-child Interaction Therapy (PCIT) for at-risk infants with externalizing behavior problems. Specifically, the intervention consisted of an adaptation of the Child Directed Interaction (CDI) phase of PCIT designed to target behavior problems, in which parents learned to follow their infant's lead in play and use differential attention to strengthen the parent-child relationship. During the first session, families participated in a CDI teach session, which included an orientation to the intervention and a discussion of the specific rules the parents should use to follow their infant's lead in play. Following the teach session, each family participated in coach sessions in which the therapist actively coached the parent "live" in using the skills with the child. Families that completed the intervention participated in an average of 6.1 sessions (range=5-7 sessions).

Thirty-one families were randomly assigned to receive the intervention. Of these, one family did not complete the Time 1 assessment or any subsequent procedures. Two additional families completed the Time 1 assessment, but did not respond to follow-up phone calls after completing the Time 1 assessment to begin the intervention. Of the 28 families that began the intervention, 8 (29 %) dropped out of the intervention, which is comparable to dropout rates in studies examining standard PCIT (e.g., 36 %; Eyberg et al. 2014). Families that dropped out of the intervention completed an average of 3.1 sessions. Intervention sessions were videotaped and a research assistant coded 63 % of the sessions, selected at random, for fidelity (i.e., the percent with which the therapist adhered to key elements of the manual). Average fidelity across intervention sessions was 97 % (range=91-100 %). Forty percent of the tapes coded for fidelity were again randomly selected and coded for fidelity a second time, yielding an interobserver reliability estimate of 95 % (range=70-100 %).

In order to meet the needs of infants and their families, handouts were tailored for infants (e.g., replacing examples to be developmentally appropriate for infants), and therapists incorporated strategies relevant for the infant population (e.g., encouraging the use of physical touch to enhance verbal reinforcement). As families participated only in the CDI



phase of PCIT, they were encouraged, but not required, to meet mastery criteria (i.e., at least 10 behavioral descriptions, 10 reflections, and 10 labeled praises and no more than 3 commands, criticisms, or questions within 5 min). Open trial data for the intervention, including detailed information on specific adaptations for infants, have been presented elsewhere (blinded for review), and main outcome analyses for the current randomized trial are in progress.

Families in both the intervention and control groups participated in a Time 2 assessment ( $n=48$ ) about 6 weeks after the Time 1 assessment, which was the post-intervention assessment for families randomly assigned to the intervention. Families in both groups also participated in a Time 3 assessment ( $n=46$ ) about 3 months after Time 2, and in a Time 4 assessment ( $n=46$ ), which was the 6-month follow-up to the Time 2 assessment, about 3 months after Time 3. Families completed similar procedures at all assessments, with the exception of Time 4, which also included administration of the Child Behavior Checklist 1.5–5 (CBCL; Achenbach and Rescorla 2000). The CBCL was only administered at Time 4 when all children were at least 18 months-old, as it has not been validated with children younger than 18-months.

## Measures

**Brief Infant-Toddler Social and Emotional Assessment (BITSEA; Briggs-Gowan and Carter 2002)**—The BITSEA is a 42-item, nationally standardized screener designed to assess behavioral problems and competencies in 12- to 36-month-olds. The problem scale has excellent test-retest reliability and very good interrater reliability (Briggs-Gowan et al. 2004). Examples of items on the problem scale include “restless and can’t sit still,” “is destructive,” and “hits, bites, or kicks” and are rated on a scale from 0 (not true/rarely), 1 (somewhat true/sometimes), or 2 (very true/often). The competence scale has good test-retest reliability and acceptable interrater reliability. Examples of items on the competence scale include “follows rules,” “is affectionate with loved ones,” and “points to show you something far away” (Briggs-Gowan et al. 2004).

**Infant-Toddler Social and Emotional Assessment (ITSEA; Carter and Briggs-Gowan 2006b)**—The ITSEA, the larger version of the BITSEA, is a parent-report questionnaire consisting of 166 items designed to assess social-emotional and behavioral problems and competencies in 12- to 36-month-olds. As on the BITSEA, each item on the ITSEA is rated as “not true/rarely,” “somewhat true/sometimes,” or “very true/often.” The ITSEA yields scores in four broad domains: externalizing, internalizing, dysregulation, and competence. In a nationally-representative sample, internal consistency estimates were high, with Cronbach’s alpha coefficients ranging between 0.85 for the internalizing domain and 0.90 for the competence domain. Test–retest reliability after 6 days was also high, with coefficients ranging from 0.76 for the competence domain to 0.91 for the dysregulation domain. Inter-rater reliability between mother-father pairs was acceptable, with correlations ranging from 0.72 for the internalizing domain and 0.79 for the competence domain (Carter and Briggs-Gowan 2006b). Spanish versions of the ITSEA and BITSEA were developed by the original authors (Carter et al. 2004) but, as indicated above, the psychometric properties of the Spanish version have not yet been tested.

**Ages & Stages Questionnaires: A Parent-Completed, Child-Monitoring System, Third Edition (ASQ-3; Squires and Bricker 2009)**—The ASQ-3 is a screening tool designed to identify children ages 4 to 60 months in need of further developmental assessment. Each item (on versions matched to the child’s age) is rated as “yes,” “sometimes,” or “not yet.” Example items include “Does your baby shake his head when he means ‘no’ or ‘yes’?” and “Does your baby play with a doll or stuffed animal by hugging it?” (Squires and Bricker 2009). Questionnaires yield scores in five subscales: communication, gross motor, fine motor, problem solving, and personal-social. Further developmental assessment is recommended if a child’s score in a given domain is 2 SD below the mean in the normative sample. The ASQ-3 has demonstrated good test-retest reliability over 2 weeks ( $r=.75$  to  $.82$ ) and acceptable interobserver agreement ( $r=.43$  to  $.69$ ). In addition, ASQ-3 subscales have demonstrated good to acceptable ( $r=.51$  to  $.87$ ) estimates of internal consistency (Squires et al. 2009). The five scales of the ASQ-3 were included in the current study to examine the discriminative validity of the BITSEA competence scale. In the current sample, internal consistency estimates for the ASQ-3 subscales ranged from 0.61 to 0.92.

**Child Behavior Checklist for Ages 1½–5 (CBCL; Achenbach and Rescorla 2000)**—The CBCL is a widely-used 99-item parent-rating scale designed to measure the frequency of young children’s behavior and emotional problems now or within the past 2 months. Parents rate each item as “not true (as far as you know),” “somewhat or sometimes true,” or “very true or often true.” The CBCL yields Syndrome scales, DSM-oriented scales, and internalizing, externalizing, and total problem scales. The CBCL total problems scale yields an interrater (mother-father) reliability coefficient of 0.65 and a 1-week test-retest reliability coefficient of 0.90 (Achenbach and Rescorla 2000). This scale was included in the current study to examine the predictive validity of the BITSEA problem scale. In the current sample, the internal consistency estimate for the CBCL total problems scale was 0.93.

### Statistical Analyses

Prior to analysis, the data were evaluated for multivariate outliers by examining leverage indices for each individual and defining an outlier as a leverage score four times greater than the mean leverage. No outliers were detected using this method. Missing data were minimal for most variables but non-trivial for some. For example, for the ASQ-3, 13 % of the cases had missing data. Values for missing data were imputed using the Expectation-Maximization method in SPSS 20 (Allison 2003). To examine internal consistency in the full sample, Cronbach’s alpha coefficients were computed for the BITSEA problem and competence scales at the screening. In addition, internal consistency was examined separately for families who completed the screening in English and in Spanish and separately for boys and girls. These comparisons were only performed in the analyses using the full sample, as the reduced sample size in the subsample was not large enough to make these comparisons. Test-retest reliability for the BITSEA problem and competence scales was evaluated with the subset of families that screened into the larger study and completed the ITSEA at the subsequent Time 1 assessment ( $n= 58$ ). Screening scores were correlated to scores at the Time 1 assessment (from the identical first 42 items on the larger ITSEA).



In order to examine the discriminative validity of the BITSEA competence scale, we used a binary logistic regression to determine whether BITSEA competence scores at the screening predicted infant delayed status (i.e., delayed or not delayed) on each of the five ASQ-3 subscales. The BITSEA competence scale consists of 11 items and is intended to be a brief screener to identify infants in need of a comprehensive developmental assessment. Therefore, an important index of the utility of the BITSEA competence scale would be its ability to discriminate between infants who are classified as “delayed” (i.e., more than 2 SD below the mean on communication, gross motor, fine motor, problem solving, and personal-social development) and infants who are classified as “not delayed” on the ASQ-3, a longer and more thorough screening measure of developmental functioning. Discriminative validity also was examined separately for families completing screenings in English and Spanish, as well as for boys and girls. Predictive validity was measured with a subset of the sample completing the Time 4 assessment ( $n=46$ ) by regressing the BITSEA problem scale score at the screening onto the CBCL total problems scale score at the Time 4 assessment and controlling for group.

Finally, in order to examine sensitivity to change for the BITSEA problem scale in a subset of the sample completing the Time 2 assessment ( $n=48$ ), we ran a regression analysis including BITSEA Problem scores at Time 1, group (i.e., intervention vs. standard care), and the group by Time 1 scores interaction as predictors and BITSEA Problem scores at Time 2 as the dependent variable. Sensitivity to change was only examined using the problem scale, as behavior problems were the intended intervention target.

## Results

In the current sample, 78 (54 %) children were above the BITSEA problem scale cut score, corresponding to the children in the normative sample with scores at the 75th percentile and higher. Of the children above the cut score, 45 (58 %) were boys. Twenty-six children (18 %) were below the BITSEA competence scale cut score, corresponding to children in the normative sample with scores at the 25th percentile and lower. Of these children, 14 (54 %) were boys. Overall, 26 (18 %) mothers reported being “worried” or “very worried” about their child’s behavior, emotions, or relationships. Similarly, 35 (24 %) mothers reported being “worried” or “very worried” about their child’s language development. Mothers’ worry about their child’s behavior, emotions, or relationships was moderately correlated to BITSEA problem scores,  $r(144)=.58, p<.001$ . Mother’s worry about their child’s language development was moderately correlated to BITSEA competence scores,  $r(144)=-.43, p<.001$ .

## Reliability

Cronbach’s alpha coefficients obtained in the current study using the full sample were good for the BITSEA problem scale (0.85) and low for the BITSEA competence scale (0.57). When including only mothers who completed the BITSEA in English, Cronbach’s alpha coefficients were 0.87 for the problem scale and 0.63 for the competence scale. When including only mothers who completed the BITSEA in Spanish, Cronbach’s alpha coefficients were 0.83 for the problem scale and 0.47 for the competence scale. For boys,

Cronbach's alpha coefficients were 0.86 for the problem scale and 0.66 for the competence scale. For girls, Cronbach's alpha coefficients were 0.84 for the problem scale and 0.45 for the competence scale. To examine test-retest reliability, Pearson product moment correlations were calculated for a subset of the sample ( $n=58$ ) completing the Time 1 assessment between BITSEA problem and competence scale scores at the screening and at the Time 1 assessment, an average of 11.35 days apart ( $SD=7.30$ ). Correlations were high for both the problem,  $r(58)=.81, p<.001$ , and competence scales,  $r(58)=.80, p<.001$ .

## Validity

Using the full sample, binary logistic regressions predicting delayed status on each of the five ASQ-3 subscales from BITSEA competence scale scores were all significant at the  $p<.001$  level (Table 2). Specifically, the BITSEA competence scale score predicted delayed status on the ASQ-3 subscales accurately for 79.2 % (Fine Motor Subscale) to 91.7 % (Communication Subscale) of cases, with odds ratios ranging from 0.61 (Communication Subscale) to 0.79 (Gross Motor Subscale). Thus, the odds of delayed status on the ASQ-3 subscales decreased between 0.61 and 0.79 for each one-point increase in the BITSEA competence scale. For mothers completing the BITSEA in English, competence scores significantly predicts delayed status on all ASQ-3 subscales ( $p<.05$ ), with odds ratios ranging between 0.59 and 0.82. Similarly, for mothers who completed the BITSEA in Spanish, competence scores significantly predicts delayed status on all ASQ-3 subscales ( $p<.05$ ), with odds ratios ranging from 0.67 to 0.76.

To examine predictive validity in the subset of the sample completing the Time 4 assessment ( $n=46$ ), BITSEA problem scale scores at screening were regressed onto the CBCL total problems scale score at Time 4, controlling for group. BITSEA problem scale scores significantly predicted scores on the CBCL total problems scale,  $\beta=1.28, t(43)=3.86, p<.001$ . BITSEA problem scale scores also explained a significant proportion of the variance in CBCL total problem scores,  $R^2=.26, F(2, 43)=7.45, p=.002$ . Values for the families who dropped out of the study prior to Time 4 ( $n=14$ ) were imputed using the Expectation-Maximization method in SPSS 20 (Allison 2003) and yielded comparable results.

## Sensitivity to Change

In the subset of the sample ( $n=48$ ) completing both Time 1 and Time 2 assessments, we conducted a regression analysis including BITSEA Problem scores at Time 1, group (i.e., intervention vs. standard care), and the group by Time 1 scores interaction as predictors and BITSEA Problem scores at Time 2 as the dependent variable. The interaction term was not significant ( $p=.192$ ), suggesting the slopes for the intervention and standard care groups were not significantly different from each other.

Given the non-significant interaction term, we reran the regression without the interaction term, using only BITSEA Problem scores at Time 1 and group as predictors and BITSEA problem scores at Time 2 as the dependent variable. Results indicated the two predictors explained a significant proportion of the variance in BITSEA problem scores at Time 2,  $R^2=.56, F(2, 45)=28.38, p<.001$ . Specifically, results revealed a main effect of group on

Time 2 BITSEA Problem scores when holding Time 1 BITSEA scores constant,  $B=-3.66$ ,  $t(45)=-2.20$ ,  $p=.033$ . Furthermore, BITSEA Problem scores at Time 1 significantly predicted scores at Time 2 when holding group status constant, suggesting a significant main effect of time (i.e., slope),  $B=0.755$ ,  $t(45)=7.05$ ,  $p<.001$  (see Fig. 1).

Values for the families who dropped out of the study prior to Time 2 ( $n=12$ ) were imputed using the Expectation-Maximization method in SPSS 20 (Allison 2003) and both regression analyses were rerun. When including imputed scores, the interaction term in the first regression analysis (i.e., including BITSEA Problem scores at Time 1, group, and the group by Time 1 scores interaction as predictors and BITSEA Problem scores at Time 2 as the dependent variable) was non-significant ( $p=.229$ ). When including imputed scores, the second regression analysis (i.e., including only BITSEA Problem scores at Time 1 and group as predictors and BITSEA problem scores at Time 2 as the dependent variable) revealed significant main effects for both group ( $p=.034$ ) and BITSEA problem scores at Time 1 ( $p<.001$ ) on BITSEA problem scores at Time 2. Thus, results did not differ when we ran the analyses with and without missing data. All infants who were randomized to intervention or standard care groups were above the BITSEA problem cutoff at screening. Forty percent of infants in the intervention group were below the BITSEA problem cutoff at Time 2, compared with only 21 % of infants in the standard care group.

## Discussion

The purpose of the current study was to examine the psychometric properties of the BITSEA in 12- to 15-month-old infants from predominately Hispanic, low-income families. Results provided evidence for the reliability of the BITSEA problem and competence scales. Specifically, internal consistency was good for the BITSEA problem scale and test-retest reliability was good for both the problem and competence scales. Results also provided evidence for the validity of the BITSEA scales. Specifically, scores on the BITSEA competence scale significantly predicted delayed status on each of the ASQ-3 subscales, supporting discriminative validity. Additionally, scores on the BITSEA problem scale at Time 1 significantly predicted scores on the CBCL at Time 4 when controlling for group, supporting predictive validity. Finally, group status significantly predicted BITSEA problem scores at Time 2 when controlling for Time 1 scores, providing preliminary support for intervention sensitivity.

Consistent with previous psychometric studies of the BITSEA in Turkish and Dutch samples (Karabekiroglu et al. 2009; Kruizinga et al. 2012), internal consistency was good for the BITSEA problem scale and acceptable for the competence scale when including only mothers who completed the BITSEA in English and when including only boys. However, when including only mothers who completed the BITSEA in Spanish, internal consistency was good for the problem scale but poor for the competence scale. Though the Spanish version of the BITSEA (translated by the developers) was administered in Spanish by a bilingual assessor, it is possible that it may be linguistically, but not culturally, equivalent to the original English version (Geisinger 1994; Preciado and Henry 1997). To our knowledge, this was the first study to examine the psychometric properties of the Spanish version of the BITSEA. Thus, the lower internal consistency estimates for the Spanish BITSEA in the

current study highlight the need for additional psychometric evaluations to establish its reliability and validity, as well as the extent to which it can be compared to the English version. Additionally, future research with larger samples should examine measurement invariance between the English and Spanish versions.

Similarly, when including only girls, internal consistency was good for the problem scale but poor for the competence scale. These findings of lower Cronbach's alphas for parent ratings of girls' behavior are consistent with previous research on other measures (Vaz et al. 2013). Additionally, research has found that, despite identical levels of motor performance, mothers of infants underestimate their daughters' motor abilities and overestimate their sons' abilities (Mondschein et al. 2000). Thus, it is possible similar gender biases contributed to poor internal consistency for the competence scale for girls. Correspondence across an average of 11 days was high for both the problem and competence scales, supporting test-retest reliability in a subset of the sample. Taken together, these results extend the evidence for the reliability of the BITSEA.

BITSEA competence scale scores significantly predicted delayed status on each of the five ASQ-3 subscales in the full sample, both when including only mothers completing the BITSEA in English and when including only mothers completing the BITSEA in Spanish. Specifically, higher competence scale scores were associated with decreased odds of delayed status on the ASQ-3. These results provide support for the discriminative validity for the BITSEA competence scale. Discriminative validity is particularly important for screening measures, as they are used to make categorical decisions regarding the need for further evaluation (Bagner et al. 2006). Although the ASQ-3 is a longer and more thorough measure of infant development than the BITSEA competence scale, it is also a parent-report screening measure intended to identify children in need of a comprehensive developmental assessment. Thus, future research should examine the discriminative validity of the BITSEA competence scale with a more comprehensive developmental assessment tool, such as the Bayley Scales of Infant and Toddler Development (Bayley 2006). In the subset of the sample completing the Time 4 assessment, BITSEA problem scale scores significantly predicted scores on the CBCL total problems scale, supporting predictive validity for the BITSEA problem scale. Predictive validity is especially important for instruments, such as the BITSEA, used to identify children for preventive intervention programs.

In a subset of the sample that completed both the Time 1 and Time 2 assessments, a regression analysis including BITSEA problem scores at Time 1, group, and the group by Time 1 scores interaction as predictors and BITSEA problem scores at Time 2 as the dependent variable resulted in a non-significant interaction term, suggesting the slopes for the intervention and standard care groups were not significantly different from each other. It is possible this was due to our small sample size. Additionally, the intervention was intended to target externalizing problems specifically, while the BITSEA problem scale measures a broader range of problems, which could have affected these results.

When we regressed BITSEA problem scores at Time 1 and group onto BITSEA problem scores at Time 2, results revealed a main effect of group on Time 2 scores when holding Time 1 scores constant. Thus, the BITSEA detected an intervention effect, such that the

intervention led to lower BITSEA scores at Time 2 (holding constant Time 1 scores) for the intervention group compared to the standard care group. Furthermore, BITSEA problem scores at Time 1 significantly predicted scores at Time 2 when holding group status constant, suggesting a significant main effect of time (i.e., slope). Therefore, BITSEA Time 2 scores for the intervention group were significantly lower than BITSEA Time 2 scores for the standard care group at every level of BITSEA Time 1 scores.

The proportion of infants with elevated problems in the current sample was another interesting finding. In a normative sample, we would expect 25 % of children to score above the cutoff for the BITSEA problem scale. However, more than half of the infants in the current study scored above the cutoff, highlighting the high-risk nature of the sample. Interestingly, only 18 % of mothers reported being “worried” or “very worried” about their infant’s behaviors, emotions, or relationships despite more than half of the mothers rating their infants above the BITSEA problem scale cutoff. The discrepancy hypothesis has been proposed to explain similar findings with another parent-report measure of child behavior. Briefly, the discrepancy hypothesis proposes that parents who report high frequencies of disruptive behavior but report few problems with these behaviors may be permissive or overly tolerant of disruptive behaviors (Butler et al. 2008). Additionally, as the disruptive behaviors assessed are considered early risk factors for more severe behavior problems later in life, it is possible that mothers in the current study were more tolerant of their infants’ disruptive behavior, perhaps due to their young age. Future research should compare mother ratings on the BITSEA to behavior ratings by an objective observer in order to evaluate the accuracy of mothers’ report of their infants’ behavior.

There were some limitations in the current study that should be considered when interpreting the results. First, the sample size was relatively small compared to other studies examining the psychometric properties for the BITSEA. However, the sample size was sufficient given the number of items on the BITSEA scales (Peterson 1994). Nevertheless, the current results should be considered preliminary evidence for psychometric properties in this sample of predominately low-income, Hispanic infants and with the Spanish version of the BITSEA. Future research should examine psychometric properties in larger samples with similar demographic compositions. Additionally, the sample size was not sufficient to test the BITSEA’s latent factor structure (MacCallum et al. 1999), which has not yet been examined and should be addressed in future research on the BITSEA.

Another limitation is the short timeline of the current research. Future research should examine psychometric properties which are time sensitive, such as test-retest reliability and predictive validity, over longer periods of time. Additionally, the constraints presented by the main study limit the generalizability of the current psychometric findings. For example, participants included in the test-retest analyses were drawn from those scoring above the 75th percentile on the BITSEA Problem scale (due to study inclusion criteria). Given the limited range of problem scores (raw score range=13–48, out of a possible 62), test-retest reliability estimates for this scale should be interpreted with caution. However, the BITSEA Problem scale is intended to be used as a screener to identify infants displaying early signs of behavior problems (i.e., infants scoring above the 75th percentile). Therefore, it is important to demonstrate that scores for these infants, in particular, are reliable over time.

Furthermore, as the BITSEA Competence scale score was not used as a screening criterion, the test-retest analysis for this scale was not compressed (raw score range=3–21, out of a possible 22).

Another limitation is that only mothers were included in the current study, which did not allow examination of interrater reliability. Fathers are often excluded from research studies in child psychopathology (Phares et al. 2005) even though they can offer unique and important information about the child (Treutler and Epkins 2003). Thus, future research should include fathers and other care-givers in order to both measure interrater reliability and obtain additional information about the infant's behavior. Finally, while the examination of psychometric properties in the youngest age range was a strength of the current study, it did not allow for examination of the psychometric properties of the Spanish version with the full age range of 12 to 36 months. Therefore, future research should examine these psychometric properties for the Spanish version of the BITSEA using larger samples with the full age range.

Despite these limitations, the current study was the first to examine the psychometric properties of the BITSEA in the youngest applicable age range (i.e., 12 to 15 months) in a predominately Hispanic, low-income sample. The current study was also the first to examine intervention sensitivity to change and psychometric properties of the Spanish version of the BITSEA. Additionally, the current research extended evidence for internal consistency, test-retest reliability, discriminative validity, and predictive validity in the BITSEA. Taken together, this research extends support for the use of the BITSEA as a screening tool in pediatric primary care settings and suggests that the BITSEA is effective for use with the youngest patients from predominately Hispanic and low-income families. Results also suggest a need for continued research on the psychometric properties of the Spanish version of the BITSEA and with girls. Finally, this research study supports the utility of the BITSEA to identify children for early intervention and as a measure of intervention outcome.

## Acknowledgments

This study was funded by a career development award from the National Institute of Mental Health to the third author (K23 MH085659).

## References

- Achenbach, TM. Manual for the child behavior checklist/2–3 and 1992 profile. Burlington: University of Vermont, Department of Psychiatry; 1992.
- Achenbach, TM.; Rescorla, LA. Manual for the ASEBA preschool forms & profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families; 2000.
- Allison PD. Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology*. 2003; 112(4):545. [PubMed: 14674868]
- American Academy of Pediatrics, Council on Children With Disabilities, Section on Developmental Behavioral Pediatrics, Bright Futures Steering Committee, Medical Home Initiatives for Children With Special Needs Project Advisory Committee. Identifying infants and young children with developmental disorders in the medical home: An algorithm for developmental surveillance and screening. *Pediatrics*. 2006; 118(1):405. [PubMed: 16818591]
- Bagner D, Harwood M, Eyberg S. Psychometric considerations in child behavioral assessment. *Handbook of child behavioral assessment*. 2006:63–79.



- Bagner DM, Rodríguez GM, Blake CA, Linares D, Carter AS. Assessment of behavioral and emotional problems in infancy: A systematic Review. *Clinical Child and Family Psychology Review*. 2012; 15(2):113–128. [PubMed: 22262040]
- Bayley, N. Bayley scales of infant and toddler development: Bayley-III. Harcourt Assessment, Psych. Corporation; 2006.
- Briggs-Gowan, MJ.; Carter, AS. Brief Infant-Toddler Social and Emotional Assessment (BITSEA) manual, version 2.0. New Haven, CT: Yale University; 2002.
- Briggs-Gowan M, Carter AS, Irwin JR, Wachtel K, Cicchetti DV. The brief infant-toddler social and emotional assessment: screening for social-emotional problems and delays in competence. *Journal of Pediatric Psychology*. 2004; 29(2):143–155. [PubMed: 15096535]
- Briggs-Gowan MJ, Carter AS, Bosson-Heenan J, Guyer AE, Horwitz SM. Are infant-toddler social-emotional and behavioral problems transient? *Journal of the American Academy of Child and Adolescent Psychiatry*. 2006; 45(7):849–858. [PubMed: 16832322]
- Butler AM, Brestan EV, Eyberg SM. Examination of the Eyberg Child Behavior Inventory discrepancy hypothesis. *Child and Family Behavior Therapy*. 2008; 30(3):257–262.
- Campbell SB, Shaw DS, Gilliom M. Early externalizing behavior problems: toddlers and preschoolers at risk for later maladjustment. *Development and Psychopathology*. 2000; 12(03):467–488. [PubMed: 11014748]
- Carter, AS.; Briggs-Gowan, MJ. Infant-toddler social and emotional assessment. San Antonio, TX: Harcourt; 2006a.
- Carter, AS.; Briggs-Gowan, MJ. ITSEA/BITSEA: Infant toddler and brief infant toddler social and emotional assessment Examiner's manual. San Antonio, TX: Harcourt Assessment; 2006b.
- Carter AS, Briggs-Gowan MJ, Davis NO. Assessment of young children's social-emotional development and psychopathology: recent advances and recommendations for practice. *Journal of Child Psychology and Psychiatry*. 2004; 45(1):109–134. [PubMed: 14959805]
- DeNavas-Walt, C.; Proctor, BD.; Smith, JC. Current population reports, P60–245, income, poverty, and health insurance coverage in the United States: 2012. Washington, DC: U.S. Census Bureau; 2013.
- Eyberg S, Boggs S, Jaccard J. Does maintenance treatment matter? *Journal of Abnormal Child Psychology*. 2014; 42(3):355–366. [PubMed: 24413969]
- Feldman MA, Hancock CL, Rielly N, Minnes P, Cairns C. Behavior problems in young children with or at risk for developmental delay. *Journal of Child and Family Studies*. 2000; 9(2):247–261.
- Foster SL, Cone JD. Validity issues in clinical assessment. *Psychological Assessment*. 1995; 7(3):248.
- Geisinger KF. Cross-cultural normative assessment: translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment*. 1994; 6(4):304.
- Gross D, Sambrook A, Fogg L. Behavior problems among young children in low-income urban day care centers. *Research in Nursing & Health*. 1999; 22(1):15–25.10.1002/(sici)1098-240x(199902)22:1<15::aid-nur3>3.0.co;2-2 [PubMed: 9928960]
- Guralnick MJ, Neville B, Hammond MA, Connor RT. The friendships of young children with developmental delays: a longitudinal analysis. *Journal of Applied Developmental Psychology*. 2007; 28(1):64–79. [PubMed: 17558442]
- Huston AC, McLoyd VC, Coll CG. Children and poverty: issues in contemporary research. *Child Development*. 1994; 65(2):275–282. [PubMed: 7516847]
- Karabekiroglu K, Rodopman-Arman A, Ay P, Ozkesen M, Akbas S, Tasdemir GN, Boke O, Peksen Y. The reliability and validity of the Turkish version of the brief infant-toddler social emotional assessment (BITSEA). *Infant Behavior and Development*. 2009; 32(3):291–297.10.1016/j.infbeh.2009.03.003 [PubMed: 19411111]
- Kruizinga I, Jansen W, De Haan CL, Van der Ende J, Carter AS, Raat H. Reliability and validity of the dutch version of the brief infant-toddler social and emotional assessment (BITSEA). *PLoS One*. 2012; 7(6):e38762. [PubMed: 22715411]
- Lambert MJ, Hawkins EJ. Measuring outcome in professional practice: considerations in selecting and using brief outcome instruments. *Professional Psychology: Research and Practice*. 2004; 35(5):492.

- MacCallum RC, Widaman KF, Zhang S, Hong S. Sample size in factor analysis. *Psychological Methods*. 1999; 4(1):84.
- Mathiesen KS, Sanson A. Dimensions of early childhood behavior problems: stability and predictors of change from 18 to 30 months. *Journal of Abnormal Child Psychology*. 2000; 28(1):15–31. [PubMed: 10772347]
- McClendon DT, Warren JS, Green KM, Burlingame GM, Eggett DL, McClendon RJ. Sensitivity to change of youth treatment outcome measures: a comparison of the CBCL, BASC-2, and Y-OQ. *Journal of Clinical Psychology*. 2011; 67(1):111–125. [PubMed: 21046644]
- Mondschein ER, Adolph KE, Tamis-LeMonda CS. Gender bias in mothers' expectations about infant crawling. *Journal of Experimental Child Psychology*. 2000; 77(4):304–316. [PubMed: 11063631]
- Olson SL, Bates JE, Sandy JM, Schilling EM. Early developmental precursors of impulsive and inattentive behavior: from infancy to middle childhood. *Journal of Child Psychology and Psychiatry*. 2002; 43(4):435–447. [PubMed: 12030590]
- Ortman, JM.; Shin, HB. Language projections: 2010 to 2020. Paper presented at the annual meetings of the American Sociological Association; Las Vegas, NV. August; 2011.
- Peterson RA. A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research*. 1994; 21(2):381–391.10.2307/2489828
- Phares V, Fields S, Kamboukos D, Lopez E. Still looking for Poppa. *American Psychologist*. 2005; 60(7):735–736.10.1037/0003-066x.60.7.735 [PubMed: 16221013]
- Pons JI, Matías-Carrelo L, Rodríguez M, Rodríguez JM, Herrans LL, Jiménez ME, Jiménez K. Estudios de validez de la Escala de Inteligencia Wechsler para Adultos Versión III, Puerto Rico (EIWA-III). *Revista Puertorriqueña de Psicología*. 2008; 19:75–111.
- Preciado J, Henry M. Linguistic barriers in health education and services. 1997
- Qi CH, Kaiser AP. Behavior problems of preschool children from low-income families review of the literature. *Topics in Early Childhood Special Education*. 2003; 23(4):188–216.
- Sand N, Silverstein M, Glascoe FP, Gupta VB, Tonniges TP, O'Connor KG. Pediatricians' reported practices regarding developmental screening: do guidelines work? Do they help? *Pediatrics*. 2005; 116(1):174–179. [PubMed: 15995049]
- Shaw DS, Gilliom M, Ingoldsby EM, Nagin DS. Trajectories leading to school-age conduct problems. *Developmental Psychology*. 2003; 39(2):189–200. [PubMed: 12661881]
- Shaw DS, Lacourse E, Nagin DS. Developmental trajectories of conduct problems and hyperactivity from ages 2 to 10. *Journal of Child Psychology and Psychiatry*. 2005; 46(9):931–942. [PubMed: 16108996]
- Shaw DS, Dishion TJ, Supplee L, Gardner F, Arnds K. Randomized trial of a family-centered approach to the prevention of early conduct problems: 2-year effects of the family check-up in early childhood. *Journal of Consulting and Clinical Psychology*. 2006; 74(1):1. [PubMed: 16551138]
- Simpson GA, Colpe L, Greenspan S. Measuring functional developmental delay in infants and young children: prevalence rates from the NHIS-D. *Paediatric and Perinatal Epidemiology*. 2003; 17(1): 68–80. [PubMed: 12562474]
- Sonnander K, Claesson M. Predictors of developmental delay at 18 months and later school achievement problems. *Developmental Medicine and Child Neurology*. 1999; 41(3):195–202. [PubMed: 10210252]
- Squires, J.; Bricker, D. Ages & stages questionnaires, Third Edition (ASQ-3): A parent-completed child monitoring system. 3. Baltimore, MD: Brookes Publishing; 2009.
- Squires, J.; Twombly, E.; Bricker, D.; Potter, L. ASQ-3 user's guide. Baltimore, MD: Brookes Publishing; 2009.
- Treutler CM, Epkins CC. Are discrepancies among child, mother, and father reports on children's behavior related to parents' psychological symptoms and aspects of parent-child relationships? *Journal of Abnormal Child Psychology*. 2003; 31(1):13–27. [PubMed: 12597696]
- Vaz S, Parsons R, Passmore AE, Andreou P, Falkmer T. Internal consistency, test-retest reliability and measurement error of the self-report version of the Social Skills Rating System in a sample of Australian adolescents. *PLoS One*. 2013; 8(9):e73924. [PubMed: 24040116]

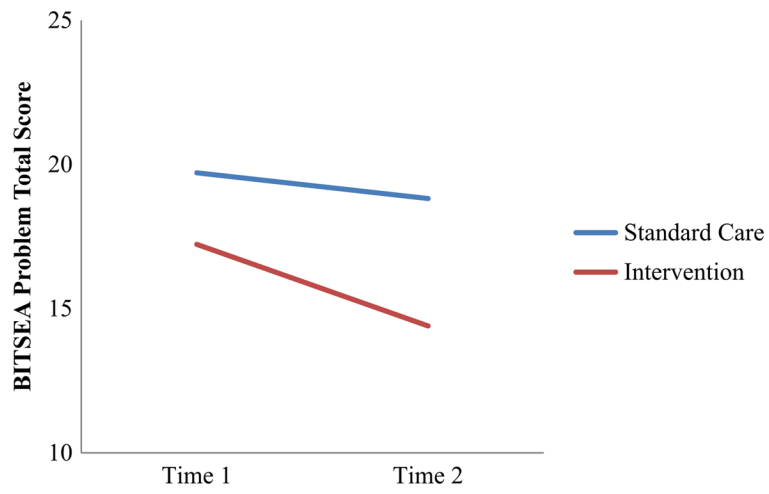
Wechsler, D. Wechsler abbreviated scale of intelligence. San Antonio, TX: Psychological Corporation; 1999.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Fig. 1.** BITSEA problem scale scores at Time 1 and Time 2 by Group. BITSEA = Brief Infant-Toddler Social and Emotional Assessment

**Table 1**

## Sample demographic characteristics

<b>Characteristic</b>	<b><i>M</i> (<i>SD</i>)</b>	<b><i>N</i> (%)</b>
Child age (months)	13.26 (1.27)	
Child sex (% male)		80 (55.6)
Child ethnicity (% Hispanic)		116 (80.6)
Child race (% White)		105 (72.9)
Mother age (years)	29.68 (5.67)	
Mother ethnicity (% Hispanic)		113 (78.5)
Mother race		102 (70.8)
Mother marital status (% married)		75 (52.1)
Mother education (% high school or less)		45 (31.3)
Family Income (% below poverty line) <sup>a</sup>		28 (37.8)

<sup>a</sup>Out of 74 families with available income data

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2**  
Binary logistic regression predicting ASQ-3 delayed status from BITSEA competence scale

Sample	ASQ-3 Subscale	$\beta$ (SE)	Wald	95% CI for odds ratio		
				Lower	Upper	
Full sample (n=144)	Communication	-0.50** (0.11)	19.94	0.49	0.61	0.76
	Gross Motor	-0.23** (0.07)	12.39	0.70	0.79	0.90
	Fine Motor	-0.26** (0.06)	17.32	0.68	0.77	0.87
	Problem Solving	-0.32** (0.07)	21.14	0.63	0.73	0.83
English Version (n=79)	Personal-Social	-0.44** (0.09)	25.82	0.54	0.64	0.76
	Communication	-0.53** (0.14)	13.21	0.45	0.59	0.79
	Gross Motor	-0.20* (0.08)	7.05	0.70	0.82	0.95
	Fine Motor	-0.26** (0.08)	10.96	0.66	0.77	0.90
Spanish Version (n=65)	Problem Solving	-0.35** (0.09)	14.34	0.59	0.71	0.85
	Personal Social	-0.41** (0.10)	15.35	0.54	0.67	0.82
	Communication	-0.43* (0.18)	5.55	0.45	0.65	0.93
	Gross Motor	-0.30* (0.13)	5.48	0.58	0.74	0.95
	Fine Motor	-0.27* (0.11)	6.64	0.62	0.76	0.94
	Problem Solving	-0.28* (0.11)	6.17	0.61	0.76	0.94
	Personal Social	-0.50** (0.15)	10.85	0.45	0.61	0.82

df = 1

ASQ-3 Ages and Stages Questionnaires—Third Edition

\*  $p < .05$ .

\*\*  $p < .001$