ijbd

International Journal of Behavioral Development
2019, Vol. 43(3) 271–277

DOI: 10.1177/0165025419830256
journals.sagepub.com/home/jbd
SAGE

Short-term test–retest reliability and continuity of emotional availability in parent–child dyads

Joyce J. Endendijk,[1] Marleen G. Groeneveld,[2] Maja Deković,[1] and Carlijn van den Boomen[3,4]

Abstract

The emotional availability scales (EAS), 4th edition, are widely used in research and clinical practice to assess the quality of parent–child interaction. This study examined the short-term reliability and continuity of the EAS (4th ed.) assessed in two similar observational contexts over a one-week interval. Sixty-two Dutch parents (85% mothers) and their 9- to 12-month-old infants ($M_{age} = 10.07$ months, $SD = 0.47$, 53% boys) were videotaped twice while they interacted with each other during several tasks (free play, structured play, book reading, toys taken away). The videotapes were coded with the EAS 4th edition by two reliable coders. Moderate to strong test–retest reliability was found for the three EA parent-dimensions: sensitivity, structuring, and nonintrusiveness. Child involvement was not reliable over a one-week period, and child responsiveness could only be reliably assessed in boys. Test–retest reliability of structuring was also higher for boys than for girls. Regarding continuity, mean levels of sensitivity, structuring, nonintrusiveness, and involvement did not change over a one-week interval, but responsiveness increased for girls only. Thus, the parenting dimensions of the 4th edition of the EAS reflect stable and consistent characteristics of the parent–child dyad on the short term, but the child measures do not.

Keywords

Emotional availability scales, 4th edition, test–retest reliability, continuity, parent–child interaction

Introduction

The importance of a healthy parent–child relationship for optimal development of children is unequivocal and evident in decades of research (Ainsworth & Bell, 1970; Bowlby, 1988; Belsky & Jaffee, 2006; Groh, Roisman, van IJzendoorn, Bakermans-Kranenburg, & Fearon, 2012). Therefore, many scholars have tried to conceptualize what constitutes a healthy parent–child relationship (e.g., Ainsworth, Blehar, Waters, & Wall, 1978; Emde, 1980; Biringen, 2008). An often-used conceptual framework is the emotional availability (EA) framework (Biringen, Derscheid, Vliegen, Closson, & Easterbrooks, 2014). EA is a dyadic construct that refers to the emotional connection between parent and child, the accessibility of parent and child to each other, and their ability to perceive and respond to each other's signals in an appropriate way (Biringen, 2008; Biringen, Robinson, & Emde, 1998). EA consists of four parent dimensions (i.e., sensitivity, structuring, nonintrusiveness, nonhostility) and two child dimensions (i.e., responsiveness, involvement) dynamically influencing each other. This means that any aspect of the parent's behavior needs to be viewed in light of the behavior of the child and vice versa. The child thus plays an important role in the formation of a positive parent–child relationship. There is ample evidence for the importance of EA for the quality of the attachment relationship between parent and child, and several aspects of infant and toddler development, such as emotion regulation, social behavior, and language development (for a review, see Biringen et al., 2014).

The Emotional Availability Scales (EAS) were designed to assess EA via ratings of observed parent–child interaction in a wide array of observation contexts, such as free play, (semi-)structured tasks, still-face situations, and separation–reunion situations (Biringen et al., 1998). Several improvements were made in the EAS over the years, resulting in a 4th edition in 2008. Over 50 studies explicitly mention using the 4th edition of the EAS. The scales are also used often in clinical contexts, particularly by parenting evaluators (e.g., child custody, social service) and sometimes by therapists to evaluate therapy process and potential progress (Saunders et al., 2017). The EAS is most often used in the infancy period (for a review, see Biringen et al., 2014), most likely because of the importance of EA for attachment formation between parent and child. However, little is known about the psychometric properties of the 4th edition of the EAS. Therefore, the goal of the current study was to examine the psychometric properties (i.e., test–retest reliability, continuity) of the 4th edition of the EAS in 10-month-old infants with their parents.

[1] Child and Adolescent Studies, Utrecht University, The Netherlands
[2] Centre for Child and Family Studies, Leiden University, The Netherlands
[3] Department of Experimental Psychology, Utrecht University, The Netherlands
[4] Department of Developmental Psychology, Utrecht University, The Netherlands

Corresponding author:
Joyce J Endendijk, Child and Adolescent Studies, Utrecht University, Heidelberglaan1, PO Box 80140, 3508 TC, Utrecht, The Netherlands.
Email: J.J.Endendijk@uu.nl

For both clinical practice and research purposes it is important to know more about the short-term reliability and continuity of parent-relationship quality derived from the EAS. Reliability is defined as the consistency in rank order of EA scores over time. Reliability can be determined by calculating correlation coefficients between EA scores of a group of parent–child dyads on one timepoint with EA scores of the same group at a later timepoint. Significant and strong correlations indicate high stability (i.e., a dyad who scores low on EA relative to other dyads at one time point is likely to score low relative to other dyads at a second time point). For clinicians, information about the reliability of the EAS is crucial to determine whether they can base decisions about, for example, child custody on EA-measurement or whether they need more assessments to reliably determine quality of the parent–child relationship. For researchers, information about the reliability of the EAS is important as they are often interested in explaining antecedents and consequences of individual differences in EA.

Continuity is defined as consistency in mean EA level of a group across time. Continuity can be determined by comparing mean EA levels of a group of parent–child dyads on one time point with mean EA scores of the same group at a later time point. Nonsignificant mean-level EA differences indicate that a group behaves consistent over time. Short-term continuity of EA scores is needed to be able to interpret longer-term changes in EA as actual developmental change at a group level.

Several previous studies provide information on the reliability and continuity of the EAS across time. However, no studies yet examined the short-term test–retest reliability and continuity of the EAS, 4th edition. Most previous studies examined longer-term stability and continuity of the EAS over several months or years. An important problem with the investigation of longer-term reliability is that low to moderate test–retest correlations might not reflect low reliability per se, but instead indicate changes over time, which the measure is correctly picking up on. Examining short-term reliability can take into account this confounding effect of developmental change. Further, previous reliability research focused on older editions of the EAS (2nd or 3rd editions). The most important difference between older EAS editions and the newest edition is that the coding system has become more differentiated and specific (Biringen, 2008). Each EA dimension is now subdivided into several specific behaviors that are scored separately and added up to a total score representing global-relationship quality. This approach allows for more variation in scores compared to the older editions in which only one global score was given for each EA dimension on a 5- to 9-point rating scale (Biringen, 2008).

Regarding longer-term reliability of the EAS, Biringen, Matheny, Bretherton, Renouf, and Sherman (2000) assessed maternal sensitivity and structuring at 18, 24, and 39 months with the 2nd edition of the EAS. They found moderate correlations between the 18- and 24-month home-assessments, but no correlations with the 39-month lab-assessment. This could be because time and context were confounded in this study, as not all three assessments were conducted in the same context. A similar confound was present in a study by Bornstein et al. (2010; EAS 3rd ed.), who found low to moderate correlations for all EA dimensions assessed at 5 months in a naturalistic home setting and again at 20 months in a free-play setting at home. Further, Howes and Obregon (2009; EAS 2nd ed.) found marginal evidence for the reliability of EA dimensions assessed at 8, 14, 24, and 26 months during naturalistic home observations. Using the 4th edition of the EAS, Pillai Riddell et al. (2011) found moderate to strong correlations for a composite of all EA parental scales assessed at 2, 4, 6, and 12 months during hospital immunization. Finally, Hallers-Haalboom et al. (2017; EAS 4th ed.) found low to moderate correlations for parental sensitivity assessed at 12, 24, 36 months in a free-play setting at home.

Regarding longer-term continuity (or development) in EA scores, Bornstein et al. (2010) found decreases in sensitivity, structuring, and nonintrusiveness over time, while nonhostility, child responsiveness and involvement did not change. In contrast, Biringen et al. (1999; EAS 2nd ed.) showed increases in maternal sensitivity and child involvement assessed during naturalistic home observations from 9 to 14 months, while nonintrusiveness and child responsiveness did not change. Howes and Obregon (2009) found decreases in hostility, increases in structuring, responsiveness, and involvement, and no changes in sensitivity assessed from 8 to 26 months. Hallers-Haalboom et al. (2017) found that sensitivity increased from infancy to toddlerhood, and then decreased into early childhood.

Only two studies examined short-term reliability and continuity of the EAS, 3rd edition. Bornstein, Gini, Suwalsky, Leach, and Haynes (2006b) found moderate to strong correlations for all EA dimensions assessed during two naturalistic home observations over a one-week interval. In a second study, short-term reliability of the EAS was similarly high across two different contexts, with moderate to strong correlations for all EA dimensions assessed during a home observation and a lab observation over a one-week interval (Bornstein et al., 2006a). Regarding continuity, both studies found no mean differences in all EA dimensions between visits. In sum, stronger evidence for the reliability and continuity of the EAS is found with short time intervals between assessments (i.e., 1 week) and assessments in similar contexts.

It is also important to examine whether the reliability and continuity of the EAS is similar in boys and girls. Several studies found gender differences in aspects of emotional availability, with parents being more sensitive with daughters than with sons (Lovas, 2005; Schoppe-Sullivan et al., 2006), using more optimal structure with daughters than with sons (Lovas, 2005), and being less intrusive with daughters than with sons (Hallers-Haalboom et al., 2014; Lovas, 2005; Tamis-Lemonda et al., 2004). Also, daughters have been found to show more involvement and responsiveness with parents (Lovas, 2005). These gender differences in EA can only be meaningfully interpreted when we have information about the reliability and continuity of the EA construct in both genders. For example, if EA can only be reliably assessed in parent–son dyads and not in parent–daughter dyads, we cannot be certain that mean-level differences found between these dyads reflect actual differences between the relationships parents have with their sons and daughters. There is some indication that long-term test–retest reliability (i.e., stability) of EA is higher for sons than for daughters (Lovas, 2005, EAS 3rd ed.). With regard to gender differences in continuity of the EA construct, mothers of 5-month-old girls decreased in structuring and nonintrusiveness over a 15-month period, whereas mothers of boys decreased on all EA-scales except involvement (Bornstein et al., 2010). However, there are also studies that do not find gender differences in reliability and continuity of the EAS (Biringen et al., 1999; Bornstein et al., 2006a, 2006b).

This study examines the short-term reliability and continuity of EA dimensions (4th ed.) assessed in two similar observational contexts over a one-week interval. Based on previous research with the 3rd edition of the EAS, we expected (a) moderate to strong correlations between the two assessments for all EA dimensions, and (b) no changes in mean levels of the EA dimensions over a one-week

Table 1. Demographic Characteristics of Sample.

	Mothers ($n = 51$)		Fathers ($n = 11$)	
	M (*SD*)	Range	*M* (*SD*)	Range
Parents' age (years)	34.02 (4.11)	25–44	31.40 (3.89)	24–36
% highly educated[1]	91		60	
Country of birth of parent, Netherlands[2] (%)	91		100	
Number of children in family (%)				
1	72		70	
2	14		30	
>2	14		0	
Child age (months)	10.07 (0.50)	9.11–12.36	10.04 (0.35)	9.37–10.59
% boys	47		82	

Note. [1]Higher vocational or university level.
[2]Other countries of birth of parents than the Netherlands included Suriname (2%), other European country (2%), other non-European country (5%).

interval. We also examined gender differences in reliability and continuity of the EAS in an explorative way, because of inconsistent previous findings.

Methods

Participants

This study was part of a larger project ("Consortium on Individual Development," n.d.), aimed at studying development and determinants of child behavioral control and social competence. Eighty infants between 9 and 12 months old and one of their parents were recruited for a two-visit study, through the local municipality of Utrecht, the Netherlands. Infants were excluded if they had a parent who did not speak Dutch, they were born preterm (i.e., before 37 weeks of pregnancy), they had known significant uncorrected hearing or vision impairment, or a significant developmental delay or condition that was likely to affect brain development or the ability to participate in the study. Eighteen parents and children were excluded, because of missing parent–child interaction data on one of the test days, resulting in a final sample of 62. See Table 1 for the demographic characteristics of the parents and children. This could be considered as a low-risk sample, because of generally high education of the parents and the absence of developmental delays or conditions that affected brain development of their children.

Procedure

Parent–child interactions were videotaped in two sessions that were apart for an average of 7.5 days ($SD = 2.91$). Both sessions were conducted with the same parent (mother or father). To avoid confounding time with context, both sessions were conducted in the same context: in the lab (95%) or in the home. At the start of each session, parents provided written informed consent. During each session, parent and child were filmed for a total of 15 minutes, subdivided in 3-minute episodes of successively (a) free play, (b) structured play, (c) book reading, (d) free-play, (e) toys taken away. The different settings were chosen to be better able to detect variations in emotional availability (McElwain & Booth-LaForce, 2006). Parents were asked to play with their child as they would normally do. Parents received the same instructions and the same toys in both sessions. In the first free-play episode, parent and child could play with a set of four standardized age-appropriate toys (doll, car, pop-up toy, rattle). In the structured-play context parents were instructed to play with a shape sorter. In the book-reading setting parent and child received a feel-and-learn book to play with. In the second free-play setting parent and child could play freely with all the previously introduced toys (doll, car, pop-up toy, rattle, shape sorter, book). In the toys taken away episode, all toys were put in a basket by the experimenter, followed by free-play without toys. Parents and children respectively received a financial compensation (€10 per session) and a small present for their participation. The study protocol was approved by the Ethical Committee of the University Medical Center Utrecht.

Emotional Availability Assessment

The Emotional Availability Scales, 4th edition (EAS; Biringen, 2008) were used to assess three dimensions of parenting (i.e., sensitivity, structuring, nonintrusiveness) and two dimensions of child behavior (i.e., responsiveness, involvement). The EAS includes a 4th parenting dimension, nonhostility, that was not included in the current study. In low-risk samples and play contexts, there is generally too little variation in nonhostility to be able to examine this dimension in an adequate way (Biringen et al., 2014; Bornstein et al., 2006b).

Sensitivity refers to the parent's ability to be warm and appropriately responsive to the child. Structuring refers to the parent's ability to support the child's learning and exploration without overwhelming the child's autonomy and in a way to which the child is receptive. Nonintrusiveness refers to the parent's ability to give the child space to explore and to refrain from intrusions on the child's activities. Responsiveness refers to the child's ability to explore on his/her own in a way that is appropriate considering the context and the child's developmental level, while still being responsive and affectively available to the parent. Involvement refers to the child's ability to involve the parent in interaction by initiating eye contact or physical contact, asking questions, or showing and demonstrating materials to the parent.

Each dimension is divided into seven subscales; the first two subscales are coded on a 7-point Likert scale and the other subscales are coded using a 3-point Likert scale (potential score range: 7 = nonoptimal, to 29 = optimal). Following Hallers-Haalboom et al. (2014), subscale 7 of the nonintrusiveness dimension (the adult is made to "feel" or "seem" intrusive) was excluded because it refers to child behavior rather than parental behavior (leading to score range: 6 = nonoptimal, to 26 = optimal). Subscale 2 of the involvement dimension (elaborative initiatives such as creating a storyline) was also omitted, because of the young age of the infants (leading to score range: 6 = nonoptimal, to 22 = optimal; Biringen, 2008). More detailed information about the EAS dimensions and subscales can be found in Biringen (2008) and Biringen, Derscheid, Vliegen, Closson, and Easterbrooks (2014).

The two observation sessions were coded independently by two trained and reliable coders (first and second authors). The coders have experience coding hundreds of parent–child observations with the EAS for previous studies (see, Hallers-Haalboom et al., 2014, 2017). Parent and child dimensions were coded separately in order to avoid a "halo-effect" whereby the coder codes parent and child dimensions similarly high (or low) (Biringen et al., 2014). Each

Table 2. Descriptives, Correlations, and Comparisons of Emotional Availability Scales at Two Measurement Waves.

	Measurement 1			Measurement 2			$r_{test-retest}$			$r_{test-retest}$*		
EA-scales	All *M* (*SD*)	Boys *M* (*SD*)	Girls *M* (*SD*)	All *M* (*SD*)	Boys *M* (*SD*)	Girls *M* (*SD*)	All	Boys	Girls	All	Boys	Girls
Parent scales												
Sensitivity	22.48 (4.02)	22.48 (4.21)	22.48 (3.88)	22.58 (3.97)	22.58 (4.23)	22.59 (3.74)	.55**	.68**	.38*	.50**	.66**	.38
Structuring	21.53 (3.77)	21.42 (4.12)	21.66 (3.40)	21.89 (3.80)	21.58 (4.12)	22.24 (3.44)	.54**	.69**	.29	.55**	.73**	.24
Nonintrusiveness	19.26 (3.87)	19.94 (3.14)	18.48 (4.50)	19.10 (3.81)	19.36 (3.62)	18.79 (4.07)	.33**	.40**	.27	.32*	.38*	.31
Child scales												
Responsiveness	19.60 (3.55)	19.73 (3.83)	19.45 (3.27)	20.37 (4.13)	18.79 (4.41)	22.17 (2.93)	.19	.37*	−.07	.19	.32*	−.05
Involvement	13.26 (2.85)	13.03 (3.12)	13.52 (2.54)	13.79 (3.01)	13.30 (3.10)	14.34 (2.87)	.14	.09	.20	.17	.12	.22

Note. *r** represents correlations corrected for interrater variation (Nunnally, 1978), parent gender, time difference between observation sessions, observation setting, difference between the sessions in comparability of parents' behavior to their normal behavior.

coder coded half of the first sessions and half of the second sessions. Each coder never coded the same parent or child twice. Coder reliability was determined on a random selection of 20 videos (13% of the total number of videos). Half of the reliability videos were coded at the start of coding and the other half was coded at the end of coding to determine whether coder drift was an issue. Intercoder reliability coefficients (intraclass correlation coefficient, absolute agreement) were adequate for each EAS dimension on respectively the first and second half of the reliability videos: Sensitivity, .74–.79; Structuring, .73–.87; Nonintrusiveness, .78–.73; Responsiveness, .71–.91; Involvement, .75–.92.

Results

Table 2 displays descriptive statistics for all study variables. All variables approached a normal distribution and no outliers were identified.

Validity of the Observations

As a validity check of the naturalistic observations, at the end of both observation sessions the parent evaluated how comparable both their own and their child's behavior was to how they normally behaved on a 3-point scale (1 = *not at all comparable*, 2 = *a bit comparable*, 3 = *completely comparable*). Children's behavior was more comparable to their normal behavior than parents' behavior was in both the first (child: $M = 2.76$, $SD = .47$, parent: $M = 2.10$, $SD = .65$, $t(61) = -6.99$, $p < .01$, $d = 1.16$) and the second observation session (child: $M = 2.76$, $SD = .50$, parent: $M = 2.31$, $SD = .53$, $t(61) = -5.75$, $p < .01$, $d = 0.87$). Comparability of children's behavior to their normal behavior was not different between the observation sessions ($t(61) = 0.00$, $p = 1.0$). Parents evaluated their own behavior as more natural in the second session compared to the first session ($t(61) = -2.08$, $p < .05$, $d = 0.26$). Therefore, a difference score was computed for the difference between the two sessions in comparability of parents' behavior to how they normally behaved, which was included as a covariate in reliability and continuity analyses.

Reliability

In Table 2, the "*r*" column shows Pearson's correlation coefficients between EAS scores in the two observation sessions. The "*r**" column shows correlations that have been corrected for relevant covariates (parent gender, coder, time difference between observation sessions, observation setting, parent behavior difference). Test–retest correlation coefficients were moderate to strong for all EA parent-scales, with or without controlling for relevant covariates. However, for the child scales responsiveness and involvement the test–retest correlations were low and non-significant. We checked whether low test–retest correlations for the child scales could be due to differences between the two sessions with regard to the time of day the children were assessed and the time since the infants had last slept or eaten. However, controlling for differences between the sessions in time of day, and time since the infants had slept or eaten did not change the test–retest correlations (responsiveness: $r = 10$, $p = .46$; involvement: $r = .08$, $p = .55$).

Regarding gender differences, *z*-tests for differences between correlations from independent samples revealed that test–retest correlations were not significantly different between boys and girls for sensitivity ($z = 1.60$, $p = .06$), nonintrusiveness ($z = 0.55$, $p = .29$), and involvement ($z = 0.42$, $p = .34$). However, the test–retest correlations for structuring ($z = 2.05$, $p < .05$) and responsiveness ($z = 1.71$, $p < .05$) were significantly stronger for boys than for girls. The same gender differences were found for test–retest correlations that were controlled for covariates.

Continuity

Table 2 also displays means and standard deviations for EAS scores for the two observation sessions in the whole sample and separately for boys and girls. To test continuity in scores, repeated-measures analyses of variance were conducted with EAS scores in the two observation sessions as repeated measures, and child gender and relevant covariates (parent gender, coder, time difference between observation sessions, observation setting, parent behavior difference) as between-subjects variables. There was no significant change in EAS scores for sensitivity ($F(1,55) = 0.05$, $p = .83$), structuring ($F(1,55) = 0.00$, $p = .97$), nonintrusiveness ($F(1,55) = 0.03$, $p = .86$), responsiveness ($F(1,55) = 0.59$, $p = .45$), and involvement ($F(1,55) = 1.04$, $p = .31$). No main effects of child gender on the EAS scores were found (sensitivity: $F(1,55) = 0.31$, $p = .58$; structuring: $F(1,55) = 0.02$, $p = .89$; nonintrusiveness: $F(1,55) = 1.56$, $p = .22$; responsiveness: $F(1,55) = 3.52$, $p = .07$; involvement: $F(1,55) = 1.84$, $p = .18$). There were also no gender differences in continuity for sensitivity ($F(1,55) = 0.01$, $p = .92$), structuring ($F(1,55) = 0.15$, $p = .70$), nonintrusiveness ($F(1,55) = 0.23$, $p = .64$), and involvement ($F(1,55) = 0.69$, $p = .41$).

The gender difference in continuity of responsiveness was significant ($F(1,55) = 9.13$, $p < .01$, *partial* $\eta^2 = .14$), indicating that girls showed an increase in responsiveness ($t(28) = -3.24$, $p < .01$), whereas boys' responsiveness did not change between the two observation sessions ($t(32) = 1.16$, $p = .26$).[1]

Discussion

This study was set up to examine the short-term reliability and continuity of the 4th edition of the EAS in 10-month-old infants with their parents. First, moderate to strong test–retest reliability was found for the three EA parent dimensions: sensitivity, structuring, and nonintrusiveness. However, regarding the EA child dimensions, involvement was not reliable over a one-week period, and the responsiveness dimension was only reliable for boys, not for girls. Test–retest reliability of structuring was also higher for boys. Regarding continuity, mean levels of sensitivity, structuring, nonintrusiveness, and involvement did not change over a one-week interval, but responsiveness increased for girls only. Together, these findings suggest that the parenting dimensions of the 4th edition of the EAS appear to reflect stable and consistent characteristics of the parent–child dyad on the short-term, but the child measures do not.

The moderate to strong test–retest reliability of the EA parent-scales was expected based on reliability studies of the 3rd edition of the EAS, that also found evidence of moderate to strong short-term reliability over a one-week interval (Bornstein et al., 2006a, 2006b). Similar to our findings these studies also demonstrated lower reliability of parental nonintrusiveness, compared to the other EA parent-scales. Bornstein et al. (2006a, 2006b) attributed this lower reliability to a restriction of the range, which could not have been the case in our study as the range was similar in all scales. Another reason for the lower reliability of nonintrusiveness might be that parental behavior that is passive/withdrawn is considered highly nonintrusive, in the 3rd and 4th edition of the EAS. It might be better to represent individual differences in nonintrusiveness on a curvilinear scale (e.g., optimal scores at midpoint of scale, passive/withdrawn at one extreme of the scale, and poorly timed/negative intrusive at the other extreme of the scale), as was done in the second edition of the EAS.

Previous studies (Bornstein et al., 2006a, 2006b) also demonstrated short-term reliability of the EA child dimensions – responsiveness and involvement – whereas we did not. This indicates that a child who scores low on EA relative to other children at one time point is not likely to score low relative to other children at a second time point. This could be due to the unpredictable nature of infants' behavior (Kenrick, Griskevicius, Neuberg, & Schaller, 2010). The low test–retest correlations for the infant scales could not be attributed to differences between the observation sessions in time of day, or infant tiredness and hunger, because after controlling for these variables the test–retest correlations remained low. These findings suggest that the child-scales of the 4th edition of the EAS do not yield reliable predictions about behavior of *individual* infants as young as 10 months old.

The lower test–retest reliability of the EA child-scales in this study compared to previous studies, might be primarily due to the fact that we examined the 4th edition of the EAS. In the 4th edition subscales are each differentiated in seven specific behaviors, whereas in previous editions subscales were coded more holistically by assigning one score. It is possible that infant social behavior in particular might not be differentiated as easily in so many sub-behaviors, because it is not yet as multifaceted as the social behavior of older children and adults (Rosenblum, Dayton, & Muzik, 2009). The developers of the 4th edition of the EAS acknowledged this partially by omitting the subscale "elaborate involvement" from the involvement dimension. Yet, it might be better to code infant responsiveness and involvement on a more global level, or to increase the observation length to provide more substantial information about the infant's behavior (Bornstein et al., 2006b).

With regard to continuity of the EAS, mean-level EA scores did not change over a one-week period, except for girls' responsiveness. This finding is largely consistent with previous studies demonstrating the short-term continuity of EA assessed with the 3rd edition of the EAS (Bornstein et al., 2006a, 2006b). Our finding indicates that longer-term changes in EA, for example, over the course of a month or a year, can be interpreted as actual developmental change, because the EAS yields consistent scores at the *group* level.

Regarding gender differences in short-term test–retest reliability and continuity, we found that short-term reliability of responsiveness and structuring was stronger for parent–son dyads than for parent–daughter dyads, which is in line with one previous study (Lovas, 2005). Furthermore, girls' responsiveness increased over a one-week period. The group-level increase in girls' responsiveness over a one-week period could be an indication of the more rapid social development of girls compared to boys in infancy (McClure, 2000). Research has for example shown that 12-month-old girls were more responsive to mothers' emotional cues than boys (Rosen, Adamson, & Bakeman, 1992), which might be because infant girls are already better at processing facial expressions and other nonverbal cues than boys (McClure, 2000). However, the low rank-order stability in girls' responsiveness indicated that not all girls increased equally in responsiveness. Considering the dyadic nature of the EAS, it is possible that girls with parents who provided appropriate structuring of play in the first session, increased most in responsiveness in the second session. These girls in particular might have been reinforced for being responsive by the higher quality play they experienced with their parent. In addition, girls who received appropriate structuring in the first session might have remembered how to play with the toys that were again provided in the second session. As a consequence, parents of these girls might have reduced their structuring attempts, whereas parents of girls who did not remember how to play with the toys might have increased or changed their structuring attempts. More research is necessary to examine the mechanisms underlying differences in rank-order stability of structuring and responsiveness in parent–daughter and parent–son dyads. It is unlikely that structuring and responsiveness are easier to observe by coders in parent–son dyads than in parent–daughter dyads, as intercoder reliability was not different between parent–son and parent–daughter dyads.

Some limitations must be considered in the interpretation of the results. First, the generalizability of the results might be reduced because the majority of the sample was highly educated and Dutch. Relatedly, because our sample was highly normative, we could not reliably assess the EAS subscale nonhostility. Future studies in more at-risk samples should examine the psychometric properties of the nonhostility scale, because hostility is higher in at-risk samples, such as depressed parents (Low & Stocker, 2005). Second, only a few fathers were included in this study, because of which we are not able to draw a firm conclusion about the reliability of the EAS for fathers in particular. This is an important direction for

future research, as many parent–child relationship measures were developed originally for mothers, but are also used for fathers without knowledge of the reliability and validity of these measures for fathers (Cabrera, Volling, & Barr, 2018). Last, studies with larger samples are necessary to examine reliability and validity of the latent EA construct with more advanced structural equation modeling techniques which can take into account issues related to measurement error.

To conclude, this study demonstrated short-term reliability and continuity of the parent dimensions of the EAS, 4th edition, in a normative sample of parents and their infants. However, the child dimensions of the EAS could not be reliably assessed in infants, with the exception of boys' responsiveness. Short-term increases found in girls' responsiveness might, therefore, not reflect true developmental changes, but measurement error. The gender differences found in the reliability and continuity of the EAS, 4th edition, highlight the importance of examining gender differences in the psychometric properties of psychological measures. It is short-sighted to simply assume that psychological measures can be reliably used with both genders.

Acknowledgments

The authors would like to thank the parents and infants for participating and all employees at the Child Research Center of Utrecht University for help with data collection.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded by the KF Hein fund and by the Gravitation program of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.001.003), awarded to The Consortium on Individual Development (CID).

ORCID iD

Joyce J Endendijk https://orcid.org/0000-0001-8149-912X

Supplemental material

Supplemental material for this article is available online.

Note

1. We also conducted all analyses while excluding fathers ($n = 11$) and observations ($n = 3$) in the home setting, because only a few fathers participated in this study and only a few observations were conducted in the home. Results from these analyses were nearly identical to the results presented on the whole sample (see Supplementary file). Therefore, we presented the results of the complete sample.

References

Ainsworth, M. D. S., & Bell, S. M. (1970). Attachment, exploration, and separation: Illustrated by the behavior of one-year-olds in a strange situation. *Child Development*, *41*, 49–67. doi:10.2307/1127388

Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the strange situation*. Hillsdale, NJ: Lawrence Erlbaum.

Belsky, J., & Jaffee, S. R. (2006). The multiple determinants of parenting. In D. Ciccetti & D. J. Cohen (Eds.), *Developmental psychopathology: Risk, disorder, and adaptation* (pp. 38–85). Hoboken, NJ: Wiley.

Biringen, Z. (2008). *The Emotional Availability (EA) Scales, 4th ed. Infancy/Early Childhood Version (child age: 0–5 years)*. Retrieved from www.emotionalavailability.com.

Biringen, Z., Derscheid, D., Vliegen, N., Closson, L., & Easterbrooks, M. A. (2014). Emotional availability (EA): Theoretical background, empirical research using the EA Scales, and clinical applications. *Developmental Review*, *34*, 114–167. doi:0.1016/j.dr.2014.01.002

Biringen, Z., Emde, R. N., Brown, D., Lowe, L., Myers, S., & Nelson, D. (1999). Emotional availability and emotion communication in naturalistic mother-infant interactions: Evidence for gender relations. *Journal of Social Behavior and Personality*, *14*, 463–478.

Biringen, Z., Matheny, A., Bretherton, I., Renouf, A., & Sherman, M. (2000). Maternal representation of the self as parent: Connections with maternal sensitivity and maternal structuring. *Attachment & Human Development*, *2*, 218–232. doi:10.1080/14616730050085572

Biringen, Z., Robinson, J. L., & Emde, R. N. (1998). *Emotional availability scales* (3rd ed.). Unpublished manual, Colorado State University, Department on Human Development and Family Studies, Fort Collins, CO.

Bornstein, M. H., Gini, M., Putnick, D. L., Haynes, O. M., Painter, K. M., & Suwalsky, A. J. T. (2006a). Short-term reliability and continuity of emotional availability in mother–child dyads across contexts of observation. *Infancy*, *10*, 1–16. doi:10.1207/s15327078in1001_1

Bornstein, M. H., Gini, M., Suwalsky, J. T., Putnick, D. L., & Haynes, O. M. (2006b). Emotional availability in mother-child dyads: Short-term stability and continuity from variable-centered and person-centered perspectives. *Merrill-Palmer Quarterly*, *52*, 547–571.

Bornstein, M. H., Suwalsky, J. T., Putnick, D. L., Gini, M., Venuti, P., de Falco, S., . . . Zingman de Galperín, C. (2010). Developmental continuity and stability of emotional availability in the family: Two ages and two genders in child-mother dyads from two regions in three countries. *International Journal of Behavioral Development*, *34*, 385–397. doi:10.1177/0165025409339080

Bowlby, J. (1988). *A secure base: Parent–child attachment and healthy human development*. London, UK: Routledge.

Cabrera, N. J., Volling, B. L., & Barr, R. (2018). Fathers are parents, too! Widening the lens on parenting for children's development. *Child Development Perspectives*, *12*, 152–157. doi:10.1111/ cdep.12275.

Consortium on Individual Development. (n.d.). Retrieved September 3, 2018, from https://individualdevelopment.nl/

Emde, R. N. (1980). Emotional availability: A reciprocal reward system for infants and parents with implications for prevention of psychosocial disorders. In P. M. Taylor (Ed.), *Parent–infant relationships* (pp. 87–115). Orlando, FL: Grune & Stratton.

Groh, A. M., Roisman, G. I., van IJzendoorn, M. H., Bakermans-Kranenburg, M. J., & Fearon, R. (2012). The significance of insecure and disorganized attachment for children's internalizing symptoms: A meta-analytic study. *Child Development*, *83*, 591–610. doi:10.1111/j.1467-8624.2011.01711.x

Hallers-Haalboom, E. T., Groeneveld, M. G., van Berkel, S. R., Endendijk, J. J., van der Pol, L. D., Linting, M., . . . Mesman, J. (2017). Mothers' and fathers' sensitivity with their two children:

A longitudinal study from infancy to early childhood. *Developmental Psychology*, *53*, 860–872. doi:10.1037/dev0000293

Hallers-Haalboom, E. T., Mesman, J., Groeneveld, M. G., Endendijk, J. J., van Berkel, S. R., van der Pol, L. D., & Bakermans-Kranenburg, M. J. (2014). Mothers, fathers, sons and daughters: Parental sensitivity in families with two children. *Journal of Family Psychology*, *28*, 138–147. doi:10.1037/a0036004

Howes, C., & Obregon, N. B. (2009). Emotional availability in Mexican-heritage low-income mothers and children: Infancy through preschool. *Parenting: Science and Practice*, *9*, 260–276. doi:10.1080/15295190902844589

Kenrick, D. T., Griskevicius, V., Neuberg, S. L., & Schaller, M. (2010). Renovating the pyramid of needs: Contemporary extensions built upon ancient foundations. *Perspectives on Psychological Science*, *5*, 292–314. doi:10.1177/1745691610369469

Lovas, G. S. (2005). Gender and patterns of emotional availability in mother–toddler and father–toddler dyads. *Infant Mental Health Journal*, *26*, 327–353. doi:10.1002/imhj.20056

Low, S. M., & Stocker, C. (2005). Family functioning and children's adjustment: associations among parents' depressed mood, marital hostility, parent–child hostility, and children's adjustment. *Journal of Family Psychology*, *19*, 394–403. doi:10.1037/0893-3200.19.3.394

McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin*, *126*, 424–453. doi:10.1037/0033-2909.126.3.424

McElwain, N. L., & Booth-LaForce, C. (2006). Maternal sensitivity to infant distress and nondistress as predictors of infant-mother attachment security. *Journal of family Psychology*, *20*, 247–255. doi:10.1037/0893-3200.20.2.247

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

Pillai Riddell, R., Campbell, L., Flora, D. B., Racine, N., Osmun, L. D., Garfield, H., & Greenberg, S. (2011). The relationship between caregiver sensitivity and infant pain behaviors across the first year of life. *Pain*, *152*, 2819–2826. doi:10.1016/j.pain.2011.09.011

Rosen, W. D., Adamson, L. B., & Bakeman, R. (1992). An experimental investigation of infant social referencing: Mothers' messages and gender differences. *Developmental Psychology*, *28*, 1172–1178. doi:10.1037/0012-1649.28.6.1172

Rosenblum, K. L., Dayton, C. J., & Muzik, M. (2009). Infant social and emotional development: Emerging competence in a relational context. In C. H. ZeanahJr., (Ed.), *Handbook of infant mental health* (3rd ed., pp. 80–103). New York, NY: Guilford Press.

Saunders, H., Biringen, Z., Benton, J., Closson, L., Herndon, E., & Prosser, J. L. (2017). Emotional Availability and Emotional Availability Zones (EA-Z): From assessment to intervention and universal prevention. *Perspectives in infant mental health* (pp. 1–16). Retrieved from https://perspectives.waimh.org/wp-content/uploads/sites/9/2017/05/12-16_Saundersetal_2017_1_Perspectives_IMH.pdf

Schoppe-Sullivan, S. J., Diener, M. L., Mangelsdorf, S. C., Brown, G. L., McHale, J. L., & Frosch, C. A. (2006). Attachment and sensitivity in family context: The roles of parent and infant gender. *Infant and Child Development*, *15*, 367–385. doi:10.1002/icd.449

Tamis-LeMonda, C. S., Shannon, J. D., Cabrera, N. J., & Lamb, M. E. (2004). Fathers and mothers at play with their 2-and 3-year-olds: Contributions to language and cognitive development. *Child Development*, *75*, 1806–1820. doi:10.1111/j.1467-8624.2004.00818.x