Development and Psychopathology **31** (2019), 261–277

doi:10.1017/S0954579417001778

Refining the assessment of disrupted maternal communication: Using item response models to identify central indicators of disrupted behavior

JOHN D. HALTIGAN,[a] SHERI MADIGAN,[b,c] ELISA BRONFMAN,[d] HEIDI N. BAILEY,[e] CATHERINE BORLAND-KERR,[f] ROGER MILLS-KOONCE,[g] AND KARLEN LYONS-RUTH[d]

[a]*University of Toronto;* [b]*University of Calgary;* [c]*Alberta Children's Hospital Research Institute;* [d]*Harvard Medical School;* [e]*University of Guelph;* [f]*Family & Children's Services of Guelph & Wellington County; and* [g]*University of North Carolina at Greensboro*

Abstract

The Atypical Maternal Behavior Instrument for Assessment and Classification (AMBIANCE; Bronfman, Madigan, & Lyons-Ruth, 2009–2014; Bronfman, Parsons, & Lyons-Ruth, 1992–2004) is a widely used and well-validated measure for assessing disrupted forms of caregiver responsiveness within parent–child interactions. However, it requires evaluating approximately 150 behavioral items from videotape and extensive training to code, thus making its use impractical in most clinical contexts. Accordingly, the primary aim of the current study was to identify a reduced set of behavioral indicators most central to the AMBIANCE coding system using latent-trait item response theory (IRT) models. Observed mother–infant interaction data previously coded with the AMBIANCE was pooled from laboratories in both North America and Europe ($N = 343$). Using 2-parameter logistic IRT models, a reduced set of 45 AMBIANCE items was identified. Preliminary convergent and discriminant validity was evaluated in relation to classifications of maternal disrupted communication assigned using the full set of AMBIANCE indicators, to infant attachment disorganization, and to maternal sensitivity. The results supported the construct validity of the refined item set, opening the way for development of a brief screening measure for disrupted maternal communication. IRT models in clinical scale refinement and their potential for bridging clinical and research objectives in developmental psychopathology are discussed.

A large body of empirical work has established the importance of early disturbed care as a correlate of infant disorganized attachment (Cyr, Euser, Bakermans-Kranenburg, & van IJzendoorn, 2010). Disorganized attachment, in turn, is a reliable predictor of later maladaptation (e.g., Fearon, Bakermans-Kranenburg, van IJzendoorn, Lapsley, & Roisman, 2010; Madigan, Brumariu, Villani, Atkinson, & Lyons-Ruth, 2016; van IJzendoorn, Schuengel, & Bakermans-Kranenburg, 1999). However, the forms of disturbed care accompanying infant disorganization have proved more difficult to identify. Initial scales for parental sensitivity, while reliable predictors of organized forms of insecure attachment (i.e., avoidance and resistance) failed to provide good discrimination of disorganized attachment relationships (NICHD Early Child Care Research Network, 1997; van IJzendoorn et al., 1999; although see Bernier & Meins, 2008). To overcome the limitations of more global sensitivity rating scales, Lyons-Ruth and colleagues developed the Atypical Maternal Behavior Instrument for Assessment and Classification (AMBIANCE), which codes for the disrupted interactions more strongly associated with infant disorganization (Bronfman, Madigan, & Lyons-Ruth, 2009–2014; Bronfman, Parsons, & Lyons-Ruth, 1992–2008; Lyons-Ruth, Bronfman, & Parsons, 1999). The AMBIANCE is based on the premise that the parental response to infant distress must be predictable and responsive enough to allow the infant to develop a minimally effective attachment strategy for eliciting protection and care (Lyons-Ruth, Bronfman, & Atwood, 1999; Lyons-Ruth, Bronfman, & Parsons, 1999). The AMBIANCE measure includes indices of the frightening or frightened parental behavior discussed by Main and Hesse (1990), as well as additional indices of the caregiver's failure to help the infant regulate fearful or stressful arousal.

The resulting AMBIANCE coding system includes five higher order conceptual dimensions of disrupted caregiver behaviors: affective communication errors, role/boundary confusion, fearful/disorientation, intrusiveness/negativity, and withdrawing behavior. Within each of these five broadband dimensions, disrupted behaviors are further grouped according to subdimensions that reflect particular stylistic features and contexts within that dimension. In total, there are 15 subdimensions (see Table 1) that are thought to reflect relatively homogenous or unidimensional constructs. Trained coders record the number of disrupted behaviors displayed by a caregiver on each

We are grateful to our collaborators and colleagues who contributed data for this project and to the families who participated in the studies. Further work with the refined Atypical Maternal Behavior Instrument for Assessment and Classification item set presented in this manuscript is being supported by the Social Sciences and Humanities Research Council of Canada.

Address correspondence and reprint requests to: John D. Haltigan, Department of Psychiatry, University of Toronto, Toronto, Canada; E-mail: John.Haltigan@camh.ca.

Table 1. *AMBIANCE indicator descriptives and trimmed bootstrapped AMBIANCE item parameter estimates for fitted AMBIANCE maternal behavior subdimensions*

AMBIANCE Subdimension and Item Description	Proportion Occurrence (Raw Event Count)	Discrimination (a) (Mad)	Severity (b) (Mad)
Dimension 1: Affective Communication Errors			
Subdimension 1A: Contradictory signaling to child			
Sweet voice with derogatory message	.11 (38)	−7.40 (1.39)	−2.69 (1.59)
Invites approach verbally then distances	.02 (8)	0.32 (1.54)	0.89 (6.54)
Uses friendly tone, threatening posture	.02 (6)	−0.95 (2.63)	0.27 (6.19)
Directs infant to do, then not do something	.02 (6)	2.20 (1.50)	2.97 (1.67)
Offers then withdraws toy	.04 (13)	4.53 (1.51)	2.92 (1.28)
Holds affectionately, simultaneously withdraws/threatens infant	.01 (3)	−1.44 (1.28)	−1.48 (4.49)
Subdimension 1B: Failure to initiate responsive behavior to infant's cues			
Does not soothe infant when distressed	.24 (83)	13.08 (15.72)	0.78 (0.16)
Does not offer comfort when infant falls	.09 (32)	0.37 (0.23)	6.94 (3.16)
Fails to set appropriate safety limits	.02 (5)	0.56 (0.69)	4.23 (6.00)
Ignores cues for pickup	.38 (129)	6.72 (0.55)	0.48 (0.11)
Does not intervene when infant engages in dangerous behavior	.02 (6)	0.99 (0.78)	4.57 (1.93)
Does not respond to infant vocalization directed at caregiver	.30 (91)	0.06 (0.19)	2.05 (8.86)
Does not respond to infant cue	.26 (81)	0.14 (0.30)	1.26 (4.46)
Subdimension 1C: Inappropriate responding to infant's cues			
Laughs while infant crying/distressed	.17 (58)	2.36 (0.87)	1.25 (0.22)
Directs inauthentic affect toward infant	.04 (13)	−0.15 (0.63)	−0.15 (12.88)
Ignores infant cue for distance	.10 (31)	−0.15 (0.41)	−2.59 (12.14)
Ignores infant's "no"	.14 (49)	0.23 (0.28)	4.06 (3.94)
Mother smiles when infant angry, upset, afraid, or sad	.26 (67)	1.71 (0.46)	0.99 (0.19)
Minimize/discount infant's display of distress	.32 (111)	1.24 (0.34)	0.79 (0.18)
Dimension 2: Role/Boundary Confusion			
Subdimension 2A: Role confusion			
Elicits reassurance from infant	.09 (32)	1.00 (0.34)	2.80 (0.76)
Defers to infant	.01 (4)	0.29 (0.80)	3.30 (11.04)
Asks infant's permission to do something	.06 (19)	0.61 (0.34)	5.26 (2.18)
Demands affection from infant	.18 (60)	1.94 (0.62)	1.31 (0.25)
Seeks physical attention from infant while infant engaged in activity	.03 (10)	0.79 (0.58)	4.99 (2.08)
Prioritizes own needs over infant needs	.08 (28)	1.95 (0.69)	1.96 (0.38)
Repeats self-references	.35 (120)	0.42 (0.20)	1.67 (0.70)
Behaves as a child rather than parent	.09 (31)	0.88 (0.31)	3.11 (0.88)
Speaks in baby talk (not in response to infant)	.11 (38)	0.77 (0.28)	3.15 (0.96)
Uses "we" to describe self or infant	.06 (20)	0.37 (0.33)	6.76 (4.04)
Encourages infant to engage in negative behaviors	.02 (8)	0.17 (0.42)	3.79 (19.96)
Fake cries in response to infant–fake sadness	.02 (5)	1.40 (0.71)	4.00 (1.21)
Directs infant to self	.50 (153)	1.20 (0.40)	−0.02 (0.12)
Pleads with infant for attention	.04 (13)	0.73 (0.52)	4.60 (2.09)
Asks infant for reassurance around separation	.03 (9)	0.79 (0.49)	5.08 (2.09)
Threatens to cry	.01 (2)	1.95 (0.73)	4.39 (1.18)
Escalates infant's distress	.17 (45)	0.87 (0.31)	2.18 (0.61)
Subdimension 2B: Treats child as sexual/spousal partner			
Speaks in hushed intimate tones to infant	.09 (29)	17.45 (10.50)	1.43 (0.15)
Touches inappropriate body parts of infant	.01 (4)	−0.49 (1.01)	2.73 (6.41)
Behaves/speaks in manner more appropriate for spouse	.01 (4)	25.17 (4.38)	2.39 (0.33)
Kisses infant in sexualized manner	.03 (9)	1.55 (0.65)	3.15 (0.78)
Strokes in a sexualized manner	.01 (2)	20.55 (18.81)	2.60 (0.28)
Cups infant's face in hands with extended eye gaze	.01 (3)	0.00 (1.11)	2.16 (4.64)
Dimension 3: Fearful/Disorientation			
Subdimension 3A: Fearful behavior: appears frightened, apprehensive, or deferential in relation to the infant			
Exhibits frightened expression	.05 (17)	1.17 (0.65)	3.53 (1.13)
Handles infant in timid or helpless manner	.03 (11)	1.11 (1.25)	2.03 (1.09)
Exhibits smile with fear elements	.01 (4)	1.23 (1.29)	2.41 (1.19)

https://doi.org/10.1017/S0954579417001778 Published online by Cambridge University Press

Table 1 (*cont.*)

AMBIANCE Subdimension and Item Description	Proportion Occurrence (Raw Event Count)	Discrimination (a) (Mad)	Severity (b) (Mad)
Exhibits highly vigilant posture in presence of infant	.01 (3)	1.52 (2.52)	1.35 (1.56)
Exhibits irrational fear regarding environment	.01 (4)	1.16 (1.35)	3.15 (2.21)
Startles to infant behavior without clear cause	.01 (3)	25.52 (24.47)	2.02 (0.42)
Treats infant as more powerful than self	.04 (12)	1.45 (1.07)	2.17 (0.83)
Hesitant, apprehensive, or stop-start movement in relation to infant	.04 (10)	0.19 (1.42)	1.23 (5.97)
Unexpected hesitancy/pause at moment of infant's bid for closeness/contact	.01 (3)	0.76 (1.56)	2.25 (3.78)
Approaches or moves away from infant in circuitous manner	.10 (33)	0.79 (0.98)	1.47 (1.16)
Approaches infant then quickly moves away	.02 (6)	0.86 (1.65)	1.84 (1.93)
Actively recoils from infant	.01 (2)	2.04 (3.02)	1.60 (1.35)
Fearful posture or expression (e.g., raised eyebrow, open mouth)	.10 (9)	0.83 (1.00)	2.22 (1.96)
Subdimension 3B: Disorientation or dissociative behavior			
Exhibits sudden change in mood unrelated to environment	.02 (6)	1.10 (0.69)	6.33 (3.24)
Handles infant as though inanimate	.06 (20)	0.69 (1.09)	1.36 (1.66)
Assumes trancelike posture or expression	.06 (19)	0.49 (1.18)	1.24 (2.86)
Deadened or flattened affect leaving empty feel to interaction (interaction)	.05 (15)	0.06 (1.57)	0.97 (3.83)
Exhibits sudden loss of affect	.03 (10)	1.38 (3.28)	0.73 (1.25)
Exhibits rapid shifts in affect unrelated to environment	.02 (6)	1.00 (1.75)	1.87 (2.44)
Exhibits disoriented or odd facial expression	.03 (10)	0.34 (1.47)	1.17 (6.48)
Sudden movement unrelated to environment	.03 (8)	0.60 (1.66)	1.03 (2.65)
Treats inanimate objects as animate	.02 (6)	1.19 (1.37)	2.69 (2.03)
Shifts rapidly from topic to topic or activity to activity	.13 (44)	0.05 (0.94)	1.20 (4.69)
Fails to finish movements	.01 (2)	−0.23 (1.92)	2.64 (8.06)
Subdimension 3C: Fearful or disoriented voices			
Exhibits haunted voice	.03 (10)	1.03 (0.69)	6.82 (3.67)
Exhibits frightened voice	.03 (8)	0.53 (2.12)	0.45 (4.90)
Exhibits sudden rise in intonation	.08 (26)	0.09 (0.50)	1.85 (11.92)
Exhibits stammering voice quality	.07 (23)	−0.95 (1.70)	0.54 (4.66)
Exhibits "ghost-like" whispering, stilted voice affectively disconnected	.19 (58)	−0.26 (0.37)	−2.57 (5.24)
Exhibits tense, high-pitched, squeaky voice tone such as at entry to room	.06 (18)	1.07 (2.10)	0.60 (2.60)
Exhibits sudden drop in pitch	.05 (18)	0.42 (1.43)	1.24 (5.53)
Exhibits sudden voice change, almost as if different person	.17 (51)	0.16 (0.86)	0.86 (5.06)
Affect or voice tone seems odd/unvarying in relation to environment	.01 (2)	0.06 (7.65)	0.80 (6.67)
Dimension 4: Intrusiveness/Negativity			
Subdimension 4A: Physical communications			
Pulls infant by wrist	.11 (36)	1.14 (0.34)	2.37 (0.51)
Looms	.12 (42)	1.01 (0.30)	2.36 (0.53)
Wipes infant's nose vigorously	.07 (24)	1.01 (0.42)	3.13 (0.86)
Pushes infant	.13 (43)	1.26 (0.33)	2.01 (0.35)
Attempts to grab infant	.06 (22)	1.88 (0.57)	2.15 (0.34)
Restrains infant	.10 (35)	1.20 (0.41)	2.31 (0.51)
Picks up or continues holding despite infant resistance	.09 (29)	1.35 (0.37)	2.32 (0.44)
Pulls infant into standing position	.03 (10)	1.21 (0.64)	3.57 (1.12)
Turns infant's head	.01 (3)	1.94 (0.77)	3.47 (0.90)
Behaves aggressively toward infant	.08 (26)	0.44 (0.34)	5.99 (2.80)
Touches infant in manner appearing affectionate but is irritating to infant	.18 (61)	1.35 (0.38)	1.53 (0.31)
Engages in rough physical play without enjoyment	.02 (8)	0.11 (0.56)	2.47 (14.27)
Tickles infant when infant resists	.01 (2)	6.37 (1.54)	3.54 (0.89)
Tosses toy or object at infant	.02 (7)	1.04 (0.58)	4.50 (1.73)
Physically crowds or hovers closely over infant	.12 (35)	1.12 (0.39)	2.22 (0.51)
Provides physical contact which offers no comfort	.09 (29)	0.63 (0.37)	4.11 (1.80)
Subdimension 4B: Verbal communications			
Mocks/teases infant	.11 (39)	1.58 (0.48)	1.85 (0.33)
Hushes crying infant (distinct from comforting sounds)	.16 (55)	0.88 (0.29)	2.27 (0.59)
Uses loud, sharp, or angry voice	.10 (33)	1.60 (0.45)	1.99 (0.33)
Disapproves, criticizes, or threatens	.25 (87)	2.32 (0.81)	0.86 (0.15)

https://doi.org/10.1017/S0954579417001778 Published online by Cambridge University Press

Table 1 (*cont.*)

AMBIANCE Subdimension and Item Description	Proportion Occurrence (Raw Event Count)	Discrimination (a) (Mad)	Severity (b) (Mad)
Plays frightening games such as chasing infant	.06 (22)	1.05 (0.41)	3.17 (0.92)
Makes negative comment about infant	.10 (35)	0.83 (0.27)	3.10 (0.81)
Laughs at infant	.03 (8)	0.67 (0.42)	5.72 (2.21)
Subdimension 4C: Inappropriately attributes negative feelings or motivation to infant			
Suggests negative motivation to innocuous behaviors	.02 (5)	−2.41 (1.07)	−5.51 (2.64)
Indicates infant's actions could have harmful consequences	.02 (7)	2.12 (3.72)	1.20 (1.47)
Personalizes infant behavior as negative	.06 (20)	−1.38 (3.74)	−0.58 (2.83)
Ascribes negative feelings to the infant	.02 (6)	1.34 (2.67)	0.76 (5.15)
Subdimension 4D: Exerts control using objects			
Removes toy from infant despite engagement	.15 (51)	1.74 (0.53)	1.52 (0.26)
Withholds toy from infant	.12 (42)	1.96 (0.63)	1.61 (0.27)
Directs infant to new activity while infant clearly immersed in playing with toy	.27 (91)	1.31 (0.43)	1.07 (0.25)
Deals with objects in an angry manner	.05 (12)	1.35 (0.57)	2.99 (0.75)
Ignores cue that activity is not liked, continued too long, or is too difficult for infant	.04 (13)	1.76 (0.60)	2.67 (0.51)
Dimension 5: Withdrawing Behavior			
Subdimension 5A: Creates a physical distance from infant			
Holds infant away from body with stiff arms	.06 (22)	0.98 (0.30)	3.27 (0.78)
Squats behind infant to play	.13 (43)	0.97 (0.28)	2.40 (0.53)
Backs away from infant	.08 (27)	1.77 (0.55)	2.06 (0.34)
Stands and looks down to interact with infant	.14 (47)	0.73 (0.26)	2.94 (0.83)
Turns infant away from body when holding	.18 (62)	1.06 (0.27)	1.77 (0.34)
Stands behind infant to lift	.02 (5)	0.85 (0.66)	5.17 (2.57)
Averts gaze	.09 (30)	0.54 (0.25)	4.93 (1.88)
Adopts posture designed to keep infant at a distance	.22 (75)	1.77 (0.39)	1.10 (0.15)
Maintains interaction at distance from infant	.16 (54)	0.46 (0.25)	4.18 (1.65)
Indicates touching infant uncomfortable/unpleasant	.02 (6)	1.17 (0.47)	4.26 (1.19)
Leaves area after infant approach	.04 (12)	0.86 (0.42)	4.63 (1.63)
Holds infant awkwardly	.07 (21)	0.89 (0.37)	3.52 (1.05)
Directs approaching infant away	.21 (62)	0.82 (0.24)	1.93 (0.48)
Distances when infant approaches	.05 (13)	0.95 (0.38)	3.79 (1.13)
Moves out of interaction to chair when infant clearly wants contact or interaction	.19 (56)	0.97 (0.27)	1.84 (0.39)
Puts infant down too soon before cue from infant	.36 (107)	1.70 (0.37)	0.54 (0.12)
Abrupt end to interaction	.02 (4)	1.37 (0.41)	3.90 (0.86)
Subdimension 5B: Use of verbal communication to maintain distance			
No interaction with infant	.01 (3)	1.81 (1.10)	10.01 (4.37)
Uses words to create distance	.03 (10)	0.14 (0.67)	1.93 (12.70)
Does not greet infant after separation	.45 (153)	2.14 (1.43)	−0.05 (0.41)
Interacts silently with infant	.36 (106)	−0.35 (1.07)	0.30 (1.42)
Leaves silently without speaking to infant	.09 (27)	−0.09 (0.71)	0.96 (7.38)
Subdimension 5C: Directs infant away from self via toys			
Steers infant toward toys from behind	.04 (15)	1.09 (0.48)	3.66 (1.25)
Redirects infant to toys not self as substitute for closer contact with parent	.47 (161)	1.19 (0.53)	0.14 (0.15)
Uses prop to keep infant at a distance	.09 (29)	3.72 (2.14)	1.84 (0.41)
Offers object to infant over unusual distance	.04 (12)	1.26 (0.76)	3.12 (0.89)

Note: The Atypical Maternal Behavior Instrument for Assessment and Classification (AMBIANCE) coding system descriptors taken from Bronfman et al. (2009–2014). Ordinary nonparametric bootstrap; 500 bootstrap replicates. Parameter estimates reflect the 10% trimmed mean across bootstrap replicates. Mad, median absolute deviation (from the median). For all subdimensions except 4C (see below), italicized items for each subdimension reflect the three strongest, positively discriminating items at the more severe end of the latent trait. Note that virtually all of these items also contain the most information (i.e., measurement precision) at the more severe end of the AMBIANCE latent trait, as item information/precision is related to the items discrimination value in the 2PLM model (see text). This set of items was then selected to constitute a potential screening version of the AMBIANCE for further analyses. Model solutions for Subdimensions 1A, 1B, 2B, and 4C were unstable across different random starts despite model convergence. As such, parameter estimates for these subdimensions should be considered as especially provisional.

subdimension during interactions with the infant, assign an overall rating (1–7) of the level of disrupted communication, and classify the caregiver behavior as either "disrupted" or "not disrupted" in communication with the infant (Bronfman et al., 1992–2008, 2009–2014). In the only major revision to the AMBIANCE, rating scales were added for each of the five dimensions of disrupted behavior, so that both continuous ratings as well as frequency data could be generated for each of the dimensions (Bronfman et al., 2009–2014). Prior to 2009, only very minor changes occurred in which a small number of behavioral items that were hard to define clearly and that coders then found difficult to code were deleted.

The AMBIANCE system has been used to code disrupted caregiver behaviors in both low- and high-risk samples among caregivers with children aged 4 months to 7 years (for review, see Lyons-Ruth & Jacobvitz, 2016). Meta-analytic work has confirmed an association of moderate effect size between disrupted maternal communication and disorganized attachment ($r = .35$; Madigan et al., 2006). Other work has provided evidence for the predictive and discriminative validity of disrupted caregiver behaviors in relation to disorganized, but not secure, infant attachment assessed 1 year later (e.g., Forbes, Evans, Moran, & Pederson, 2007). In addition, Madigan, Voci, and Benoit (2011) demonstrated that disrupted caregiver behaviors coded with the AMBIANCE were stable over a 6-year period. Finally, reduction in maternal disrupted communication was shown to be one mechanism mediating reduction in infant disorganized behavior in the context of a randomized intervention trial (Tereno et al., 2017).

Whereas the reliability and validity of the AMBIANCE coding system has been well documented, to date the measurement properties of the disrupted maternal behavior indicators that comprise the first-level of coding in the AMBIANCE system have not been investigated. The absence of a systematic investigation of these measurement properties is due in part to analytic and modeling challenges associated with frequency count variables (e.g., Madigan et al., 2006; Sterba et al., 2010). In addition, a reasonably large sample size is needed, given that most individual disrupted behaviors constitute low base-rate events. Nevertheless, as Madigan et al. (2006) noted almost a decade ago, more vigorous analyses of anomalous caregiving behaviors is needed to improve our understanding of the specific disrupted caregiving behaviors that best define atypical parenting. Moreover, given the high-fidelity nature of the AMBIANCE coding system, coding of particular caregiver behaviors is a laborious process, so that there is a significant demand from those working in clinical settings for a streamlined version that focuses on the most central indicators of disturbed interaction.

Latent Trait Models and Item Response Theory (IRT)

The latent trait model is the analogue of the factor analysis model for binary observed data (Muthén, 1989; Rizopoulos, 2006). Within the latent trait purview, IRT has emerged as a powerful set of modeling techniques for the analysis of item-level data obtained to measure interindividual variation (e.g., mental health status; Edelen & Reeve, 2007). However, the IRT methodological tradition originated in the measurement of latent traits of scholastic ability (e.g., reading and arithmetic; Baker, 2001), and thus, it has been used less commonly in clinical and developmental psychological science, where classical test theory approaches to instrument evaluation have been the standard. Nevertheless, the benefits and utility of IRT methods have been increasingly applied in clinical and developmental research (Cole et al., 2011; Edelen & Reeve, 2007; Fraley, Waller, & Brennan, 2000; Gordon, 2015; Reise & Waller, 2009). In particular, IRT methods can be used to provide highly detailed information on the properties of existing coding systems and their indicators, which can then be used to optimally shorten the instrument to effectively reduce coding or response burden. Accordingly, IRT methods have been applied to measurement instruments assessing mental health symptoms (e.g., depression; Cole et al., 2011), alcohol and drug symptomatology (Krueger et al., 2004; Langenbucher et al., 2004), and psychopathy (Cooke & Michie, 1997).

There are a variety of different IRT models that can be fit to binary response data (for a review, see Gordon, 2015). However, the two-parameter logistic IRT (2PLM) is often applied (e.g., Krueger et al., 2004; Langenbucher et al., 2004). A key assumption of this model is that the latent trait under investigation is a unidimensional (i.e., single-factor) construct. Item trace lines or item characteristic curves (ICCs) are produced, which are S-shaped logistic functions that graphically relate item endorsement probabilities across latent trait values (Edelen & Reeve, 2007; Martin et al., 2006). These lines are described by two parameters, the location (b) parameter, and the slope (a) parameter. The b, or location, parameter is the point along the ICC at which the probability of a positive response for a dichotomous item is 50%. The larger the location parameter, the more of the measured construct (often denoted as θ) a respondent must possess for a particular item to be endorsed. When the construct of interest (i.e., the latent trait) is relevant to mental health problems or physical disease, this parameter can be cast as the "severity" parameter. The a, or discrimination parameter, reflects how well a particular item discriminates respondents or "participants" at contiguous points around (i.e., above and below) the location parameter. In other words, it is the slope of the ICC at the value of the location parameter and indicates the extent to which the item is related to the underlying construct or latent trait. This parameter is analogous to the factor loading in traditional factor analysis.

After estimating the parameters of an IRT model, researchers can investigate the fidelity by which items measure a given latent trait by examining the item's information. In the 2PLM, an item's information value is inversely related to the item's discrimination parameter and reflects the standard error of the indicator at its location on the latent trait. As such, examining an item's information value provides crucial insight into how well (i.e., the precision with which) an indicator is measuring the latent construct under consideration. Exploring how an item's information changes as a function of the latent trait level is one of the most widely cited motivations for using IRT

https://doi.org/10.1017/S0954579417001778 Published online by Cambridge University Press

in clinical measurement (Reise & Waller, 2009). Such information is especially useful in guiding efforts at reducing item measurement batteries so as to maximize their efficiency and precision. This contribution of IRT is especially useful in clinical and high-risk samples, where it is important to efficiently extract as much information as possible at the severe end of relevant traits in order to screen for mental and physical health concerns (Kim & Pilkonis, 1999). As Forero and Maydeu-Olivares (2009) note, there is significant demand from practitioners within medical settings for short assessment tools capable of gathering the maximum amount of information in the minimum possible time.

Study Overview

In the current study, we collected all known observations of disrupted maternal behaviors for which item-level data were available, in order to provide the first large-sample analysis of the item structure of the AMBIANCE using latent trait modeling under the IRT approach. As noted, the AMBIANCE is composed of 15 disrupted behavior subdimensions (Table 1). These subdimensions were originally conceptualized as unidimensional constructs reflecting particular stylistic patterns of disrupted maternal behavior. Thus, the latent trait IRT approach is well suited to address the chief objective of the current work, which is to assess the item properties for each AMBIANCE subdimension and identify the behavioral items most central to each subdimension. In so doing, we developed a preliminary and empirically informed, refined AMBIANCE item set and evaluated its convergent and discriminant validity with constructs in its nomological network (Cronbach & Meehl, 1955), including maternal sensitivity and infant attachment disorganization. Given the central importance of assessing disrupted parenting behaviors in clinical and child protective settings, identifying these empirically central items was seen as a critical first step toward the development of a more efficient clinical screening instrument for disrupted maternal behavior that maintained adequate conceptual and content coverage with maximum precision.

Method

Participants and procedure

Item-level AMBIANCE data, acquired from six subsamples (pooled $N = 343$) were used in the current project. Data were drawn from various parent studies conducted in the United States, Canada, and Great Britain. In the United States, AMBIANCE indicator-level data were obtained from the Harvard Longitudinal Study, a longitudinal investigation of the effects of social risk factors on child development ($n = 55$; Lyons-Ruth, Bronfman, & Parsons, 1999), and from a subset of participants in the NICHD Study of Early Child Care and Youth Development (SECCYD; $n = 219$; see NICHD Early Child Care Research Network, 2005, and the study website http://secc.rti.org). In Canada, AMBIANCE data were acquired from a larger study of preschool behavioral problems in healthy and pediatric medical conditions ($n = 39$; Goldberg, Gotowiec, & Simmons, 1995; Madigan et al., 2011), and in Great Britain data were acquired from a study investigating personal relatedness and attachment patterns in 12-month-old infants of mothers with and without borderline personality disorder ($n = 30$; Hobson et al., 2009).

AMBIANCE data were coded from mother–child interactions in a variety of standard interactive research paradigms, such as the Strange Situation Procedure, free play, and cleanup task, and included children from 12 months through 54 months of age (58% of children in the pooled sample were 15 months of age). Girls and boys were approximately equally distributed in the pooled sample. Individual study cohorts were heterogeneous with respect to demographic risk given the differing aims of the parent studies (for additional detail about the demographic characteristics of parent studies comprising the pooled sample, see parent study references cited above).

Measures

Disrupted Behavior Instrument for Assessment and Classification (AMBIANCE). The AMBIANCE coding system (Bronfman et al., 1992–2008, 2009–2014) is a detailed observational coding protocol that provides objective behavioral criteria for coding disrupted caregiver communication with the infant during videotaped caregiver–infant interactions. The coder first documents the frequency of approximately 150 behavioral items on 5 dimensions of disrupted maternal behavior. Frequency counts for individual items are then summed to yield total frequency scores for each of 15 subdimensions. A final overall level of disrupted communication (1–7) is assigned by the coder, based on both the frequency and the intensity of the disrupted communications displayed by the caregiver. The overall level of disrupted communication is scored as follows: 1 = *warm and sensitive communication*, 3 = *generally positive interaction with some evidence of disrupted communication*, 5 = *clear and repeated disruption in affective communication*, and 7 = *disrupted communication with few or no ameliorating behaviors*. Scores of 5 or above on the overall rating are classified as “disrupted” and scores of less than 5 are classified as “not disrupted” (Bronfman et al., 1992–2008, 2009–2014).

Reliability of AMBIANCE coding at the level of the overall rating and classification has been strong across all the studies whose item-level data are included here (see original parent study publications as follows: Hobson et al., 2009; Lyons-Ruth, Bronfman, & Parsons, 1999; Madigan et al., 2011). For the SECCYD subsample, reliability coefficients on $n = 62$ tapes (20%) between two coders was high, with ICCs for ratings on each of the five AMBIANCE dimensions all >0.80 (Mills-Koonce et al., 2017).

Although the exact number of AMBIANCE items has changed slightly over time (see Analytic Plan section), the total numbers of behavioral items for each AMBIANCE dimension available for inclusion in the current investigation were as follows: affective communication errors = 23 indicators; role/boundary confusion = 24 indicators; fearful/disoriented behaviors = 37 indicators, intrusive/negativity = 34 indicators; and withdrawing behavior = 29 indicators. Thus, a total of 147 behavioral items were available for potential inclusion in the current analyses.

Infant attachment disorganization. In all studies, infant attachment was assessed during the standard Strange Situation Procedure (Ainsworth, Blehar, Waters, & Wall, 1978). The Strange Situation Procedure is an observational procedure that contains eight brief episodes of increasing stress for the infant, including two mother–infant separations and reunions. All video recordings were coded for infant attachment behaviors and for the three attachment classifications as described by Ainsworth et al. (1978) and for disorganized/disoriented behaviors as described by Main and Solomon (1990). Reliability of attachment classifications were satisfactory within each of the parent samples (Goldberg et al., 1995; Lyons-Ruth, Connell, Grunebaum, & Botein, 1990; Hobson, Patrick, Crandell, & Garcia-Pérez, 2005; NICHD Early Child Care Research Network, 1997).

Maternal sensitivity. The maternal sensitivity measure was available only in the SECCYD subsample ($n = 197$). Early maternal sensitivity was assessed in the context of mother–child interactions that were videotaped during 15-min semistructured play procedures at 6, 15, 24, and 36 months. At 6 months, mothers and children were instructed to play together, first with toys available in the home (or none at all) and then with a standard set of toys. At 15, 24, and 36 months, mothers were asked to show their children age-appropriate toys in three containers in a set order. As in prior studies of this sample (e.g., NICHD Early Child Care Research Network, 2001), observations of maternal sensitivity from the first 3 years of life (6, 15, 24, and 36 months) were standardized and averaged to create a composite of the observed early sensitivity. At 6, 15, and 24 months, the a priori maternal sensitivity composites were constructed by summing ratings for sensitivity to nondistress, positive regard, and intrusiveness (reversed). At 36 months the supportive presence, respect for autonomy, and hostility (reversed) scales were composited (as reported in NICHD Early Child Care Research Network, 2001, internal consistencies of composites were 0.75, 0.70, 0.79, and 0.78 for the 6-, 15-, 24-, and 36-month composites, respectively, and intercoder reliabilities on scales >.80; for additional details on the sensitivity composite, see NICHD Early Child Care Research Network, 2001, 2004).

Analytic plan

We first transformed the 147 AMBIANCE indictors from their original count scales to a dichotomous (0 = *behavior did not occur*, 1 = *behavior did occur*) scale to ease modeling burden and permit IRT modeling under the 2PLM. This transformation resulted in relatively little loss of information because most indicators had extremely low base rates due to their atypical nature.

Over time, the exact number of items in the coding manual has changed slightly, because some items were found to be difficult to code and were deleted, while new behaviors of particular import were observed in subsequent samples and added to the item list. In the current analyses, we included all AMBIANCE items that were coded in *any* of the parent studies. Next, to improve model estimation tractability (i.e., the ability to generate stable model solutions and parameter estimates), from the total pool of 147 AMBIANCE indicators available for consideration in the current analyses, indicators with zero variance (i.e., no event occurrences) and/or minimal variability (i.e., only one event occurrence) were excluded. Percentages of items with zero or minimum variability were generally evenly distributed across the five primary dimensions of the AMBIANCE: affective communication errors (17%); role/boundary communication (19%); fearful/disorientation (17%); intrusiveness/negativity (6%); and withdrawing behavior (10%). In light of model estimation concerns, if greater than 90% of the data for a specific AMBIANCE indicator were missing (e.g., due to removal from subsequent versions of the manual), it was also removed from consideration in analyses.[1] This resulted in a total AMBIANCE indicator pool of 133 items for inclusion in the IRT analyses. Missing or unavailable data on remaining AMBIANCE indicators across study cohorts ranged from none to 29%.

Using the *ltm* package (Rizopoulos, 2006) in the R environment for statistical computing (R Core Team, 2016), we then fitted latent trait models for each of the 15 subdimensions comprising the AMBIANCE maternal behavior system. The ltm package uses marginal maximum likelihood estimation (Bock & Aitkin, 1981), which is a commonly used iterative estimation procedure that provides maximum likelihood estimates of severity and discrimination parameters. Under marginal maximum likelihood estimation, all missing data are treated as missing at random, and all available cases are used in model estimation taking into account the observed part of sample units with missing data. We also specified start.val = "random" to allow for inspection of local maxima issues in likelihood surfaces (i.e., replication of the best log-likelihood across different start values).

It is well known that maximum likelihood estimation of latent (trait) models with binary or ordered categorical data present modeling challenges (Albanese & Knott, 1994; Sterba et al., 2010). This problem is further magnified in rare-event behavioral data, where many indicators demonstrate a preponderance of nonevents (i.e., the disrupted maternal behavior does not occur). Sparse data response patterns may lead to extreme parameter and/or standard error estimates, which are unstable (de Menezes, 1999) or frequently drift into inadmissible regions (Swaminathan, Hambleton, Sireci, Xing, & Rizavi, 2003). Albanese and Knott (1994)

1. Full descriptions of the 14 items removed from analyses are available from the first author upon request.

showed that estimated asymptotic variances of the parameter estimates in a one-factor model for binary data are unreliable. They arrived at a better idea of the sampling distribution of the parameter estimates by bootstrapping. Given the model estimation characteristics of the current data set, we also performed 500 bootstraps of each latent trait model using the *boot* package (Canty & Ripley, 2016) to arrive at a better approximation of the *sampling behavior* of the estimators (i.e., more precise values of the estimators and their standard errors).

For each of the 15 disrupted maternal behavior latent traits, we decided a priori to select, whenever possible, the three items possessing the strongest positive discrimination parameters on the severe end of the trait. This approach was taken to maintain the empirical meaning of each latent trait while also maintaining adequate content coverage of the AMBIANCE as a whole. These three items from each subdimension were then used to construct a 45-item refined AMBIANCE summary measure. We then evaluated the construct validity of this refined measure in relation to the full AMBIANCE coding protocol, as well as in relation to infant attachment disorganization and maternal sensitivity.

Results

Descriptive data

The proportions of observed occurrence and the raw frequency counts for all binary AMBIANCE indicators included in the present analyses are presented in Table 1. Base rates ranged from 0.01 to 0.50. As can be seen, means (i.e., proportions of occurrence) of AMBIANCE binary indicators indicated that most are rare behaviors, with low base rates of occurrence. The relative rarity of these observed behaviors is consistent with the goal of the AMBIANCE coding system to detect atypical maternal behavior.

Latent trait models for AMBIANCE subdimensions

To assess the unidimensionality of each AMBIANCE subdimension, we performed a likelihood ratio test evaluating the fit of one- and two-factor models for each subdimension using the ANOVA function in the ltm package (Rizopoulos, 2006). With the exception of the physical communications (4A) and creates physical distance from infant (5A) subdimensions, a two-factor model did *not* provide a significantly better fit to the data than did a unidimensional model (for model comparisons, all $ps \geq .05$). This supports the notion that for 13 of 15 AMBIANCE subdimensions, indicators for each disrupted maternal behavior should be conceptualized as indicators of a single, dominant underlying dimension.

Although the likelihood ratio test above suggested that the physical communications (4A) and creates physical distance from the infant (5A) subdimensions may be better explained by two factors rather than a single factor, we examined them under the 2PLM IRT unidimensional assumption in this investigation for two reasons. First, these subdimensions were two of the three subdimensions in the current analyses with 16 or more indicators (see Table 1). With relatively large numbers of items, there are many opportunities for subsets of items to have shared method variance reflecting inconsequential multidimensionality not accounted for by the dominant trait (Cook, Kallen, & Amtmann, 2009; see Floyd & Widaman, 1995). Second, confirmatory model fit indices for the physical communications subdimension (4A) were adequate (i.e., confirmatory fit index ≥ 0.93, root mean square error of approximation < 0.05, weighted root mean square residual < 1.0), suggesting the tenability of a unidimensional latent trait.[2]

The interdecile (i.e., 0.10) trimmed mean bootstrap IRT discrimination and location parameter estimates and their median absolute deviation from the median for each latent trait model are also listed in Table 1. Recall that, within the context of the current investigation, an item's location refers to the point on each latent disrupted maternal behavior trait (i.e., each subdimension) at which there is a .5 probability of that item's being observed (vs. not). An item's discrimination reflects its ability to discriminate individuals around the item's location. Exemplar item characteristic and item information curves for selected subdimensions and items are presented in Figures 1–5.[3]

For 11 of the 15 subdimensions (excluded subdimensions are discussed below), the trimmed mean discrimination (slope) estimate across the AMBIANCE subdimensions was 1.23 (range = 0.24–2.97), which corresponds to a correlation of ~.59 between a particular disrupted maternal behavior item and the underlying disrupted maternal behavior latent trait continuum. This suggests that the AMBIANCE items were generally moderately to strongly associated with their underlying latent trait abstractions. The trimmed mean severity parameter estimate across these same AMBIANCE subdimensions was 2.27 (range = 0.73–3.43). Recall that the severity parameter is scaled on a standard (z-score) metric and therefore can be directly referenced to the underlying latent trait continuum. In the present investigation, zero indicates the average level of the particular disrupted maternal behavior in the sample on the underlying latent disrupted maternal behavior trait. Thus, the AMBIANCE maternal behavior indicators are, in general, measuring the higher or more severe end of their associated disrupted maternal behavior latent traits. For example, the strongest average within-subdimension severity was observed for the role confusion subdimension (2A), suggesting that the disrupted maternal behavior indicators for this construct were measuring its

2. Categorical data factor analyses were performed in Mplus V. 7.11 (Muthén & Muthén, 1998–2012) using a robust weighted least squares estimator. Note that categorical data factor analyses and IRT are equivalent parameterizations of the same underlying model. We return to the issue of the unidimensionality of AMBIANCE subdimensions in the Discussion section.

3. Additional item characteristic and information curves can be requested from the first author.

https://doi.org/10.1017/S0954579417001778 Published online by Cambridge University Press

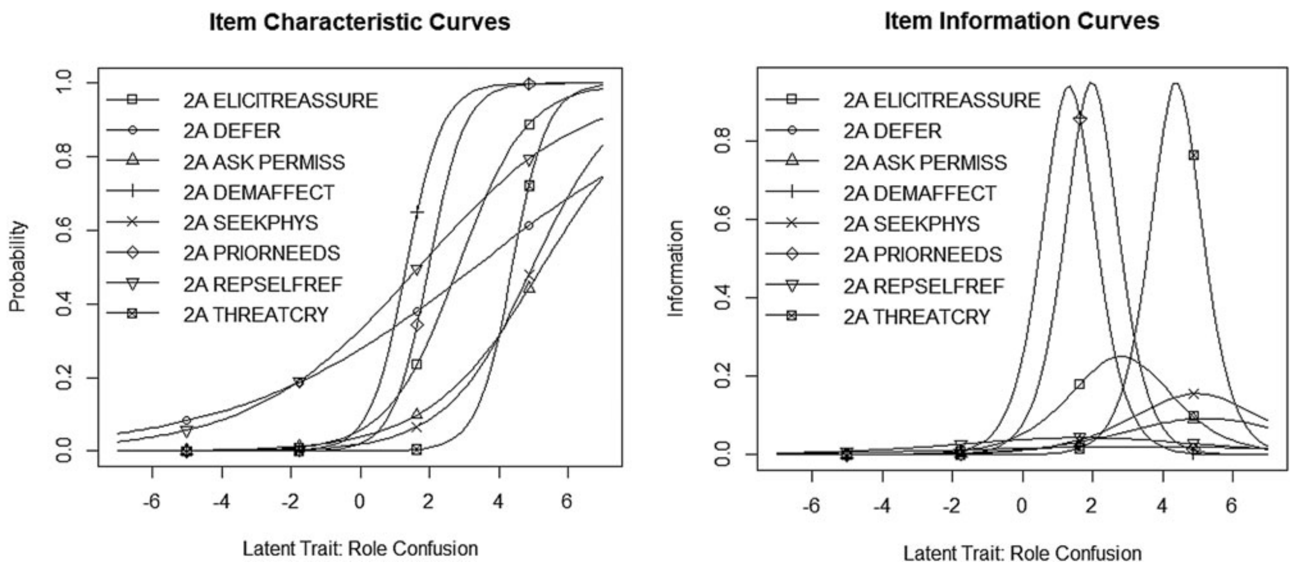

Figure 1. Item characteristic and item information curves for eight indicators of AMBIANCE subdimension 2A: "role confusion." Full item code descriptors are provided in Table 1. Chosen indicators were selected to facilitate visual interpretation.

more severe end particularly well. This is visually reflected in the right shift of the item characteristic and information curves on the latent trait (see Figure 1), as well as in the test information function for this subdimension (see Figure 6), where the largest portion of the area under the test information curve is at the more extreme end of the role confusion latent trait.

For the remaining 4 of the 15 AMBIANCE subdimensions (1A, 1B, 2B, and 4C; Table 1), model estimation was tentative. For each of these latent trait models, we were unable to replicate the best log-likelihood value three times in initial (nonbootstrapped) runs. In addition, some random starting values produced unstable individual model solutions in bootstrapped models as indicated by the absence of positive definite Hessian matrix at convergence. Finally, for some items bootstrapped standard error estimates were excessively large, indicating the extreme degree of instability for these particular parameter estimates (implications for model interpretation are discussed more fully below).

Another indicator of how well AMBIANCE items are indexing the higher or more severe end of disrupted maternal behavior latent traits is their precision (i.e., reliability) in doing so. An item's precision in IRT is reflected via its information. Item information may be thought of as the reliability of the items with respect to their ability to distinguish between respondents at *a given level of the latent trait.* AMBIANCE

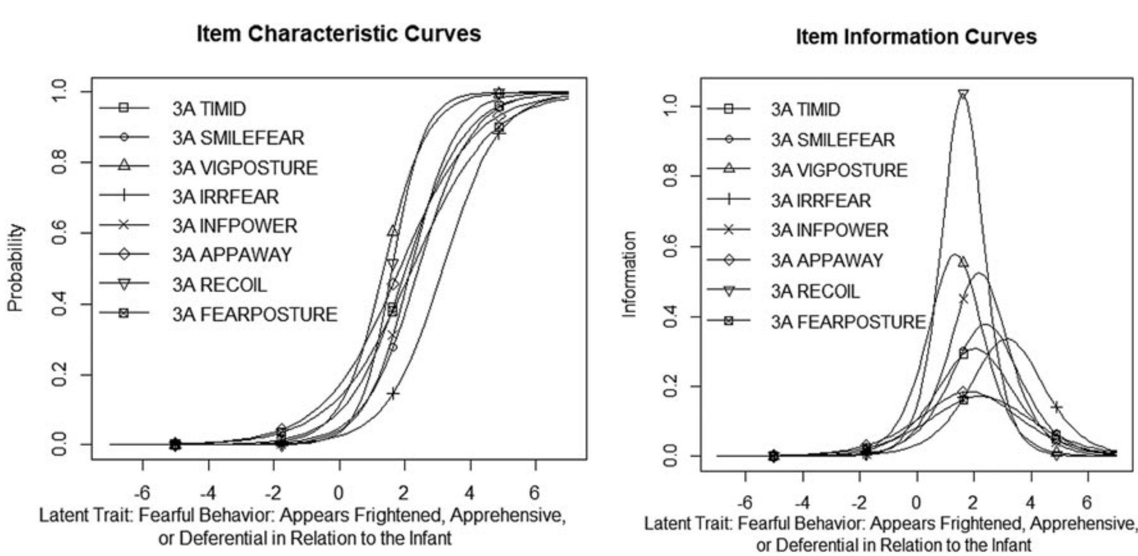

Figure 2. Item characteristic and item information curves for eight indicators of AMBIANCE subdimension 3A: "fearful behavior." Full item code descriptors are provided in Table 1. Chosen indicators were selected to facilitate visual interpretation.

https://doi.org/10.1017/S0954579417001778 Published online by Cambridge University Press

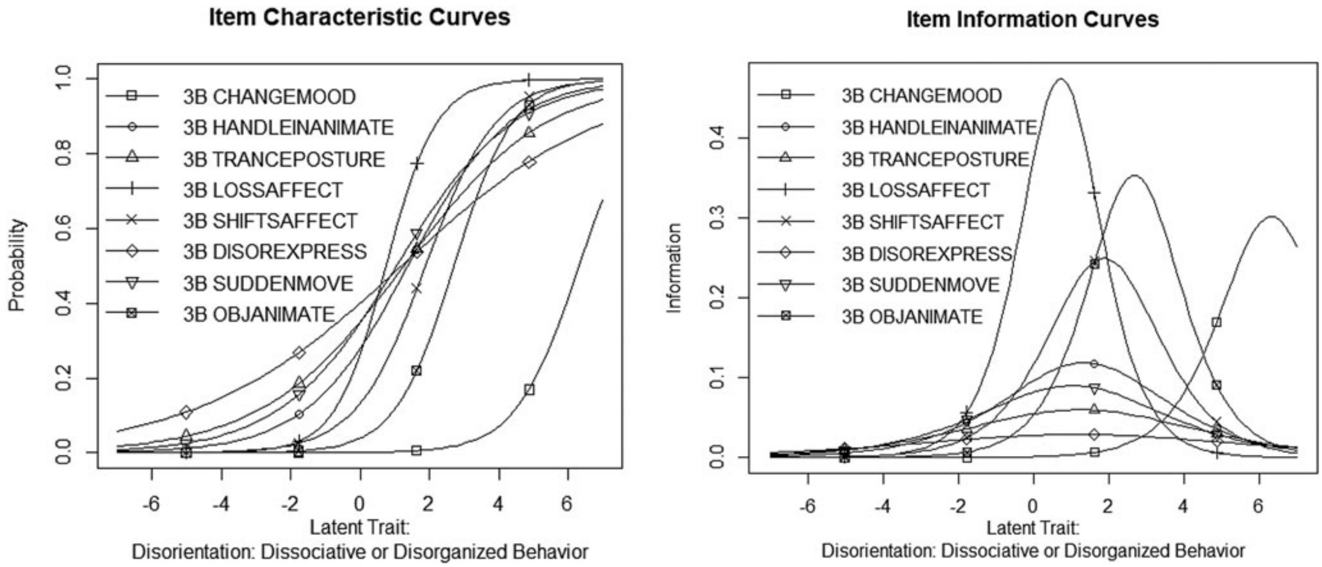

Figure 3. Item characteristic and item information curves for AMBIANCE subdimension 3B: "disorientation: disorganized or dissociative behavior." Full item code descriptors are provided in Table 1. Chosen indicators were selected to facilitate visual interpretation.

maternal behavior indicators generally demonstrated satisfactory reliability within the selected range at the more severe end of the disrupted maternal behavior latent trait (see Figure 6 for the role confusion test information function). Approximate total information values for each AMBIANCE subdimension (excluding the four subdimensions noted previously) on the latent trait continuum from 0 and +5 (i.e., the more severe end) were as follows: 1C (4.29), 2A (11.51), 3A (28.34), 3B (12.43), 3C (9.68), 4A (16.53), 4B (7.2), 4D (7.2), 5A (13.23), 5B (1.36), and 5C (5.1). These total information values correspond to IRT reliability (i.e., internal consistency) approximations of (0.77, 0.91, 0.97, 0.92, 0.90, 0.94, 0.86, 0.86, 0.92, 0.27 and 0.80, respectively), where reliability is calculated as 1 minus the squared reciprocal of the square root of the information for scores in that severity range of the disrupted latent trait (i.e., their error variance).[4] Thus, total information value provides a metric of how precisely the AMBIANCE items are measuring the more severe end of disrupted maternal behavior latent traits. The relatively smaller value of 0.27 corresponded to the subdimension use of verbal communication to maintain distance (5B). Visual inspection of the total information curve for this subdimension (not shown) revealed that the lower reliability value for this subdimension was because total item precision was most concentrated (i.e., the bulk of the item information) between –2 and +3 on the latent trait continuum (i.e., across the midpoint of the latent trait), rather than disproportionately at the severe end. This suggests that for this subdimension, items are more precisely measuring less severe and more benign aspects of this trait, rather than those mostly at the severe end.

4. The reciprocal of the square root of the information value provides an estimate of the standard error of latent trait ability measurement in the specified latent trait range, which, when squared, provides an index of error variance in the same latent trait range.

Selecting a reduced indicator set for further evaluation as a screening measure

IRT parameter estimates were used to inform our selection of a reduced set of AMBIANCE indicators. As noted above, each of the 15 disrupted maternal behavior latent traits, we decided a priori to select, whenever possible the three items that possessed the strongest positive discrimination parameters on the severe end of the trait (recall that these items also possess the most information or precision). This approach was taken to maintain adequate coverage of the empirical meaning of each latent trait (i.e., item discrimination parameters are analogous to factor loadings), as well as to maintain adequate content coverage of the full AMBIANCE protocol (Edelen & Reeve, 2007), while also ensuring a clinically practical reduced set of indicators.

Note that we also selected items from the four AMBIANCE subdimensions (1A, 1B, 2B, 4C) with tentative model solutions per the a priori selection rule described above. As further discussed below, we selected items from these models despite tentative model solutions because omitting items from AMBIANCE subdimensions considered clinically and conceptually important would have substantively altered the meaning and conceptual content of the reduced item set in relation to the full AMBIANCE system.

Construct validity

Following identification of the 45-item set, a unit-weighted sum was computed ($M = 4.59$, $SD = 3.33$) and used as a re-

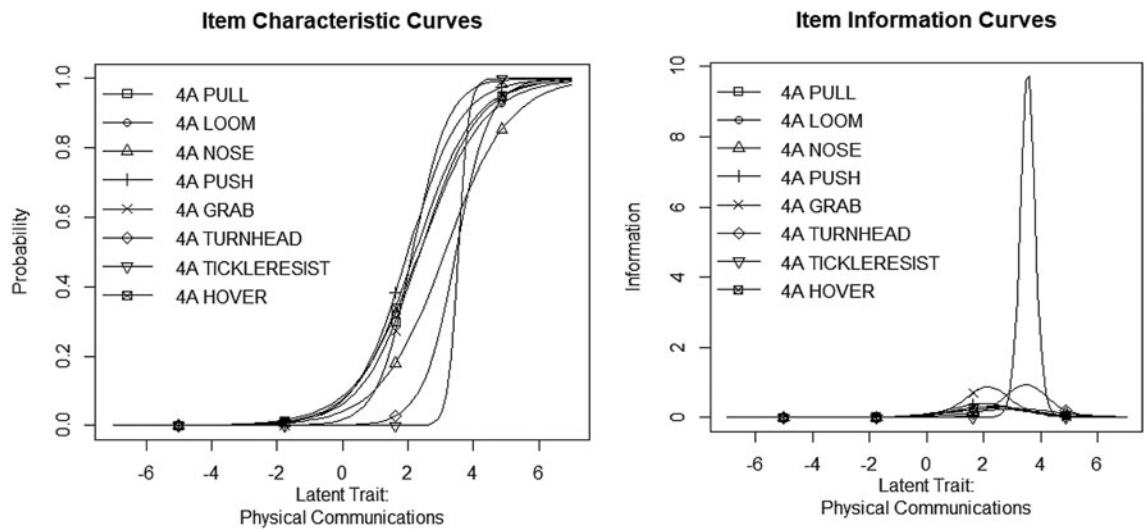

Figure 4. Item characteristic and item information curves for AMBIANCE subdimension 4A: "physical communications." Full item code descriptors are provided in Table 1. Chosen indicators were selected to facilitate visual interpretation.

fined index of disrupted maternal behavior. The unit-weighted summary score for the 45-item refined AMBIANCE was strongly convergent with a unit-weighted sum of the full AMBIANCE item pool (133 items) available in the present study ($r = .89$, $p < .001$). We evaluated the construct validity of the refined AMBIANCE 45-item set in several ways. First, we used receiver operating characteristic (ROC) analyses to explore the clinical utility of the AMBIANCE 45-item set with respect to the final overall maternal disrupted classification status that is assigned using the full AMBIANCE. To make a final classification, trained coders use the full AMBIANCE protocol to take into consideration the frequency and severity of all disrupted maternal behaviors and rate the overall level of disrupted communication observed in the parent–child interaction. Ratings of 5 or above on the overall level result in a classification as disrupted. Thus, the disrupted classification is based on a broader coder judgment than frequency counts alone.

ROC analyses are similar to logistic regression in that one can use the strength with which a set of explanatory variables predicts a given binary outcome to calibrate the precision of a measure by plotting the range of classification accuracy at different thresholds of the predictor variables. The resultant area under the curve (AUC) can then be quantified by a value representing the likelihood that a random chosen positive case (i.e., a mother classified as disrupted) will exceed the result for a randomly chosen negative case (thus ranging from

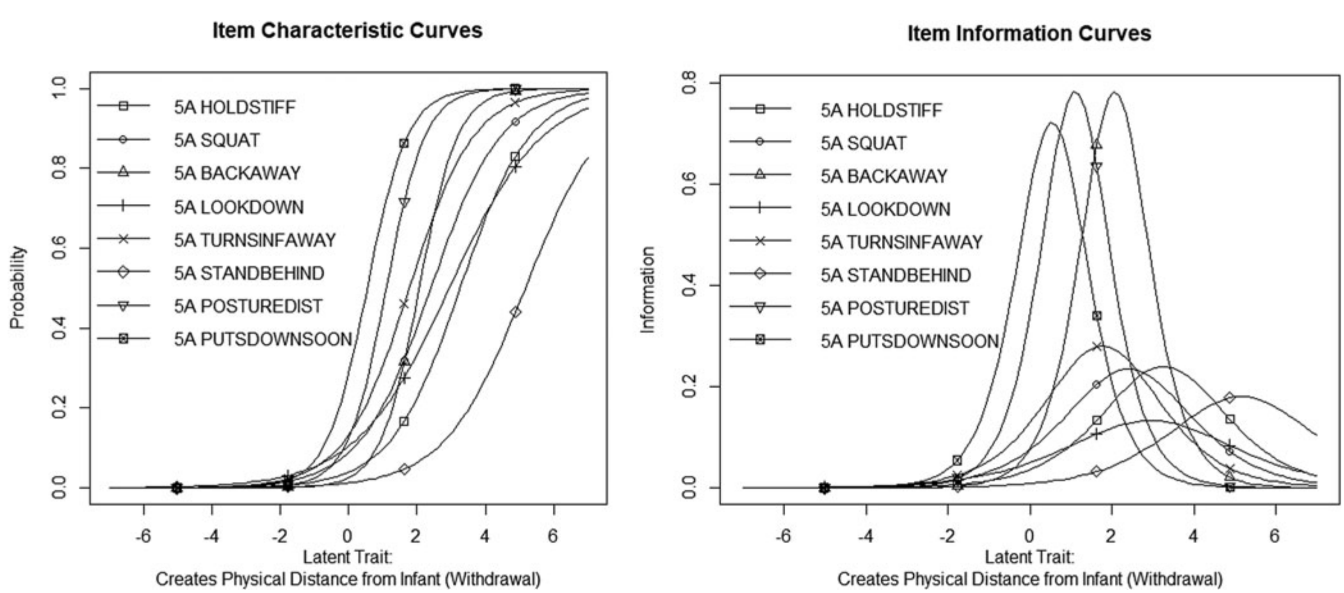

Figure 5. Item characteristic and item information curves for eight indicators of AMBIANCE subdimension 5A: "creates physical distance from infant." Full item code descriptors are provided in Table 1. Chosen indicators were selected to facilitate visual interpretation.

https://doi.org/10.1017/S0954579417001778 Published online by Cambridge University Press

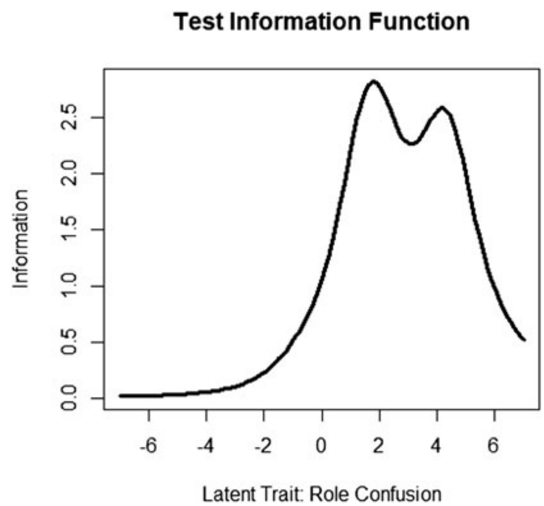

Figure 6. Test information function for AMBIANCE subdimension 2A: "role confusion." Note that the test information function is reflective of the total set of AMBIANCE indicators for a given latent trait.

0.50 to 1.00; Ondersma, Chaffin, Mullins, & Lebreton, 2005).

ROC analyses for the AMBIANCE IRT-based 45-item set (see Figure 7) showed that the AUC value and standard error was significant ($p < .01$), with an AUC of 0.85. These results indicate that the 45-item set showed good diagnostic accuracy with respect to coder classifications of maternal disrupted communication status using the complete AMBIANCE item set and coding protocol.

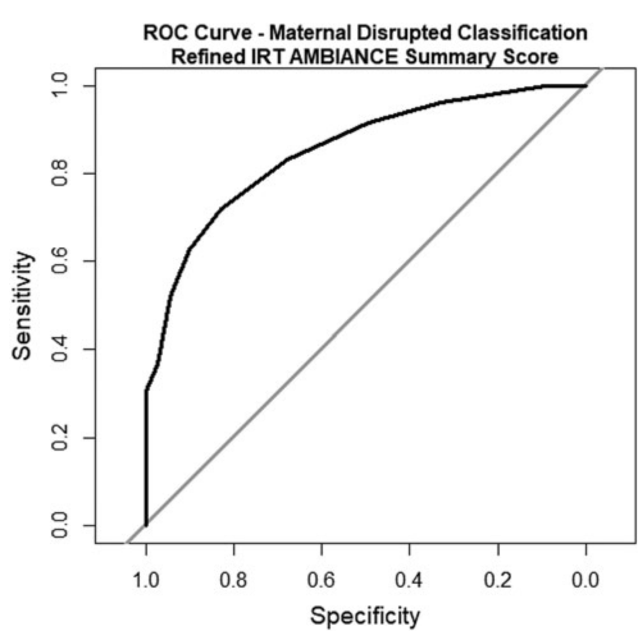

Figure 7. Receiver operating characteristic (ROC) analysis curve for the item response theory (IRT) based refined AMBIANCE summary score in detecting maternal disrupted classification status as defined using the AMBIANCE protocol. Area under the curve = 0.85, $p < .01$.

Second, we examined whether the full and the reduced AMBIANCE item sets demonstrated similar positive associations with infant attachment disorganization. Mothers of disorganized infants demonstrated significantly higher scores on the refined 45-item AMBIANCE measure (M = 5.67, SD = 3.42) compared to mothers of nondisorganized infants (M = 4.45, SD = 3.40), t (254) = –2.50, $p < .02$, d = ~0.36. These results were quite similar to those using the full AMBIANCE item set (disorganized M = 14.11, SD = 7.27; nondisorganized M = 11.48, SD = 7.28), t (254) = –2.52, $p < .02$, d = ~0.36; see Figure 8 for raw data, descriptive statistics, and inferential statistics plot of refined 45-item summary scores by attachment disorganization. Thus, scores obtained with the IRT-based 45-item set demonstrated virtually identical associations with infant attachment disorganization as did the full AMBIANCE summary score, with effect size magnitudes (i.e., strength of associations) intermediate in degree between small and medium-sized effects (Cohen, 1992).

To assess discriminative validity, we also compared the magnitude of the association between the AMBIANCE 45-item set and infant attachment disorganization to the magnitude of the association between maternal sensitivity (assessed at infant ages 6, 15, 24, and 36 months) and infant disorganization at 15 months of age in the SECCYD subsample (n = 197). We tested this comparison to assess whether the refined 45-item summary measure of disrupted communication would show a stronger relation to infant disorganization

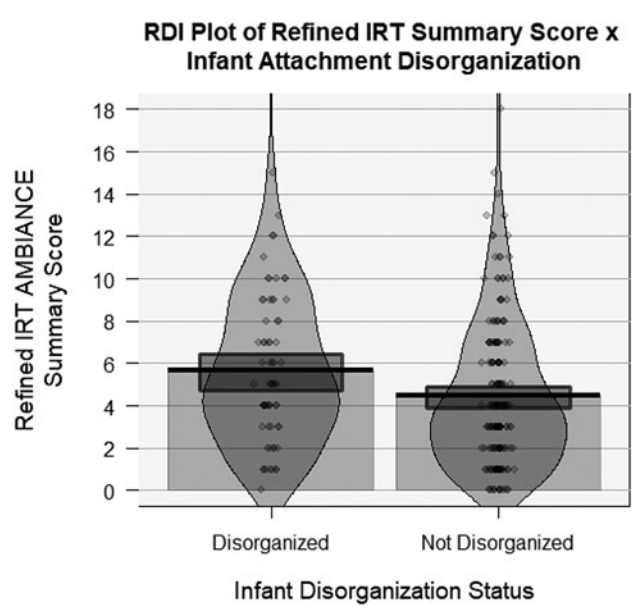

Figure 8. Raw data, descriptive statistics, and inferential statistics (RDI) plot of significant mean difference in item response theory (IRT) based refined AMBIANCE summary scores between disorganized and not disorganized infants. Note that the RDI plot contains four main elements that allow for greater empirical resolution into the patterning of data (relative, e.g., to bar plots): points reflected by darkened dots reflect raw data points; the vertical shaded bar reflects central tendencies; the bean reflects a smoothed density; and the shaded rectangle reflects an inference interval (e.g., frequentist confidence interval, as in this example).

https://doi.org/10.1017/S0954579417001778 Published online by Cambridge University Press

than maternal sensitivity. As noted earlier, meta-analytic data have generated a mean effect size of $r = .10$ for the association between maternal (in)sensitivity and disorganized attachment (van IJzendoorn et al., 1999) and $r = .35$ for the association between disrupted communication and disorganized attachment (Madigan et al., 2006). Both the full ($r = .22$) and the refined ($r = .23$) AMBIANCE scores were significantly ($p < .01$) associated with infant attachment disorganization in the SECCYD subsample, whereas maternal sensitivity was not ($r = –.06$, *ns*). Note that the association between maternal sensitivity and infant disorganization in this subsample of SECCYD data was virtually identical to the zero-order association using the full SECCYD cohort ($r = –.05$; Haltigan & Roisman, 2015). We also assessed the reliability of the difference between the correlation of disorganization with sensitivity and the correlation of disorganization with the refined AMBIANCE using the Fisher r to z transformation. The correlation of the refined (45-item) AMBIANCE summary measure with infant disorganization was significantly stronger than the association of maternal sensitivity with infant disorganization ($z = 2.90$, $p < .01$). In a final analysis, we also examined the 9-point continuous attachment disorganization scale. These analyses yielded a similar pattern of effects as those using the categorical measure of infant disorganization described above (full AMBIANCE $r = .25$, 45-item refined AMBIANCE $r = .24$, both $ps < .01$; maternal sensitivity $r = –.11$, *ns*; for the difference between refined AMBIANCE and maternal sensitivity associations with continuous infant disorganization scores, see the Steiger, 1980, test for dependent correlations, $t = 3.18$, $p < .01$, case A).

Discussion

To the best of our knowledge, this is the first investigation to attempt to extend the IRT methodology to the study of observed maternal caregiving behavior. Using IRT modeling, we examined the functioning of the individual behavioral items included in the AMBIANCE coding system to assess which items were the most informative indicators of the 15 latent dimensions of disturbed communication. Identifying the most informative indicators is important to honing our understanding of disrupted maternal communication, as well as a critical first step toward the related goal of developing an efficient screening instrument for clinical use. The results of our latent trait analyses for each of the 15 AMBIANCE subdimensions provided the necessary psychometric information to cull a refined set of 45 indicators that were maximally informative. Furthermore, the results of our 2PLM models suggested that the identified items were moderately to strongly related to their underlying latent traits and were reliably measuring the severe end of those latent traits.

In addition, the latent trait analyses yielded valuable insights into the defining characteristics of each dimension. For example, both “threatens to cry” and “prioritizes own [parent] needs over infant needs” were identified by the latent trait analyses as possessing high discriminatory value at the severe end of the role confusion latent trait. In the current sample, these items are indexing codable behaviors that carry maximal information regarding the broader theoretical construct of role confusion (Macfie, McElwain, Houts, & Cox, 2005; Mayseless, Bartholomew, Henderson, & Trinke, 2004; Sroufe, Jacobvitiz, Mangelsdorf, DeAngelo, & Ward, 1985). Similarly, “exhibits highly vigilant posture in presence of infant” and “Exhibits sudden loss of affect,” each of which emerged as possessing relatively high discriminatory value on the subdimensions measuring fearful and disoriented behavior, respectively, are central to theoretical constructs regarding the frightened or dissociative parental behaviors thought to contribute to infant disorganization (Main & Hesse, 1990). Frightened or frightening parental behavior is hypothesized to be a key mechanism mediating the link between parental unresolved trauma on the Adult Attachment Interview and attachment disorganization in the child. Thus, these analyses inform a more molecular understanding of the aspects of parental frightened or frightening behavior that may be central to the construct. Similarly, items identified across the physical communications latent trait, including “attempts to grab infant,” “turns infant’s head,” and “tickles infant when infant resists” offer insight into the core defining features of specific parenting behaviors that may underlie current conceptualizations of the construct of parental intrusiveness in the context of early parent–child interaction (Egeland, Pianta, & O’Brien, 1993; Haltigan, Leerkes, Supple, & Calkins, 2013; Lyons-Ruth, Alpern, & Repacholi, 1993). What is striking to us about these particular behaviors is their physically invasive, and coarse kinesthetic nature. Finally, the very different items on the withdrawing behavior dimension, such as “backs away from the infant,” “puts infant down too soon before cue from infant,” and “adopts posture designed to keep infant at a distance” index a maternal stance that has been differentially associated with poor self-regulation in late adolescence, including borderline features, suicidality, and antisocial personality disorder (Lyons-Ruth et al., 2013; Shi, Bureau, Easterbrooks, Zhao, & Lyons-Ruth, 2012). The identification of such particularly informative disrupted behaviors for each subdimension offers much more specificity to our conceptualizations of both the threatening and emotionally neglecting caregiving contexts that have been highlighted as potential contributors to risk for infant and child psychopathology and maladaptation, as well as altered trajectories of brain functioning (Lyons-Ruth, Pechtel, Yoon, Anderson, & Teicher, 2016; Sheridan & McLaughlin, 2014).

The results of the IRT models thus provide important points of basic and applied departure for further empirical work exploring central features of disturbed caregiving. It is also notable that initial qualitative review of the identified items by experienced clinicians and AMBIANCE coders has confirmed that the identified behaviors were all behaviors that were theoretically and clinically central to the construct being coded. This initial qualitative review must be further confirmed by studies on clinical utility, but it is important that the item set selected by the IRT analyses was felt to be clinically rich and informative on initial review.

Further, analyses exploring the initial construct validity of the refined 45-item set suggested that it performed similarly to the full AMBIANCE item set in relation to key validity criteria. First, the refined 45-item set demonstrated good diagnostic accuracy with respect to discriminating between disrupted and not disrupted maternal classifications generated from the full AMBIANCE coding protocol. Thus, the results of our ROC analyses indicate that the IRT models were performing well in identifying highly discriminating items at the more severe end of the latent traits, in somewhat the same way that coder judgment would be used to capture severity on the overall rating scale for level of disrupted communication. Second, in relation to infant attachment disorganization, the refined AMBIANCE 45-item set generated associations similar to those using the full AMBIANCE item set, for both continuous and categorical measures of disorganization. Third, in the SECCYD subsample, both the full AMBIANCE and the smaller IRT-refined item set showed discriminant validity in relation to maternal sensitivity, in that they were significantly more strongly associated with infant disorganization than was the sensitivity measure. This latter finding is consistent with meta-analytic evidence that disrupted maternal behavior demonstrates stronger associations with infant disorganization (Madigan et al., 2006) than does maternal (in)sensitivity in normative-risk populations (Haltigan & Roisman, 2015; van IJzendoorn et al., 1999).

Limitations

As we have noted, the use of latent trait analyses with dichotomous items poses several modeling challenges, in particular when there is a large preponderance of extremely low base-rate items. To address this issue and mitigate the possibility that extreme sampling properties may unduly influence the accuracy and precision of parameter estimates, we utilized bootstrapping procedures to generate latent trait parameter estimates, presenting trimmed mean estimate values as our point estimates for all relevant parameters. Although all latent trait models converged, the direction and magnitude of parameter estimates from 4 of the 15 models (1A, 1B, 2B, and 4C) should be considered especially provisional given that these models did not reliably converge at a consistent log-likelihood value. However, we have included results for these subdimensions because, given the absence of other available item-level frequency data, we considered it important to include parameter estimates for all subdimensions of the coding system, rather than omitting 4 clinically important subdimensions entirely. Omitting subdimensions considered clinically and conceptually important from the refined 45-item set would have substantively changed the meaning of the refined set in relation to the full AMBIANCE system. Moreover, by providing these estimates while also noting the issue of their imprecision (Maxwell, 2004), we allow for comparison with future work examining the latent trait structure of the AMBIANCE coding system and these subdimensions in particular.

It is also possible that, in samples with higher rates of occurrence of a given disrupted caregiving behavior, latent trait models for these 15 subdimensions would yield different absolute magnitudes for IRT parameter estimates, as well as different relative rankings of AMBIANCE indicators with respect to the magnitude of their discrimination parameter estimates. As such, pending replication efforts, we anticipate there may be some fluidity in the specific indicators from these subdimensions that ultimately emerge as those consistently demonstrated to be the most strongly related to their underlying latent traits. Similarly, the somewhat ambiguous findings regarding the unidimensionality of AMBIANCE subdimensions 4A and 5A, indexing physical intrusiveness and physical withdrawing behaviors, respectively, warrant additional research investigating their dimensionality before firm conclusions regarding their latent trait structure are suggested.

Related to the above points, it is important to note that relatively large discrimination and location parameters for some of the latent trait models are not surprising, given the rare nature of the behaviors and the consequent sparse number of observations for each item. Consequently, their absolute magnitudes should be considered cautiously in view of the nature of the data. It is likely that the absolute magnitude of these discrimination parameters may be smaller in samples where there are more events per variable for particular AMBIANCE indicators. However, given that our chief aim was not definitive parameter generalization to the population but data refinement and reduction by selecting the most informative items in the larger set, different levels of precision may be acceptable based on the nature and intent of the investigation (Edelen & Reeve, 2007). Parameter estimates need not be accepted uncritically as highly precise to argue that their relative magnitude, and thus their relative importance for measuring the disrupted maternal behavior latent trait, reflect the best available information concerning these disrupted maternal behavior dimensions. To this end, all existing AMBIANCE data sets with item-level data that we are aware of were included in these analyses. Although serving as an empirical starting point, it will be critically important to assess the replicability of these parameter estimates and continue to refine them as new data is acquired.

A final limitation to stress is that this refined set of indicators is not yet ready for widespread use in research settings. The presence of a larger pool of disrupted behaviors, as currently described in the coding manual, may be important in helping coders to develop a template of the dimension being coded in a way that the inclusion of only three items could not. This is an empirical question that needs to be addressed. A refined AMBIANCE measure would also need to be assessed in relation to its association with a variety of other relevant maternal risk factors and child outcomes. While the larger AMBIANCE measure now has a sound track record of reliability, stability over time, and convergent validity with relevant maternal and child constructs, the reduced measure has not yet received this level of scrutiny.

Despite these issues, we believe these analyses resulted in a well-chosen set of indicators, with good initial validity, that warrant taking further steps toward a more efficient measure of maternal disrupted communication for use in clinical settings. We might expect such a reduced measure to have somewhat less precision and prediction than the full measure, while still offering a more standard clinical training format and yielding information with higher validity and specificity than the varying judgments of individual clinicians. Thus, a next important step in moving toward a clinically efficient instrument will be to evaluate the reliability, validity, and prediction of the reduced item set, and any associated training procedures, when used in clinical settings.

Future directions

Results of the current work also offer a number of potential future directions for clinicians and researchers interested in studying disrupted caregiving behaviors and their predictive significance for child maladaptation. Examinations of the convergent and predictive validity of the refined AMBIANCE item set in relation to additional domains, such as maternal risk factors and measures of child psychopathology, are now needed in independent samples. In addition, additional independent large-sample studies using *item-level* AMBIANCE data are especially needed to continue to refine our understanding of the latent structure of disturbed maternal communication. Because coding systems designed to quantify disrupted parental communication often require extended training and are labor intensive to code, the development of large enough data sets at the item level presents a challenge. In future work, it will be important for researchers to develop consortia that allow for the pooling of data sets, to allow more granular analysis of the process of parent–child communication and to yield more stable parameter estimates of underlying constructs. Moreover, with appropriate sample sizes, modeling the five higher order AMBIANCE dimensions as a function of lower order subdimensions and their indictors will also be an important objective for future research investigating the latent structure of disrupted parenting.

From a measurement science perspective, new developments in IRT mixture modeling (Finkelman, Green, Gruber, & Zaslavsky, 2011; Wall, Park, & Moustaki, 2015) and Bayesian IRT modeling (Swaminathan et al., 2003) offer additional strategies to deal with low-base rate behaviors, which presents challenges when modeling the latent structure of disrupted caregiving traits. In addition, it will be important to assess whether particular disrupted caregiving behaviors show the same degree of severity and discrimination on their relevant latent traits when considering parent–child interaction at different child ages, across different geographic subsamples, and across other demographic characteristics (e.g., sex and ethnicity; Haltigan et al., 2014). Differential item functioning in the IRT framework is well suited to address these issues (Osterlind & Everson, 2009). The full AMBIANCE coding protocol has been used successfully among mothers with children ranging in age from 4 months to 7 years. However, the item-level data used in the IRT analyses here were only available for mothers of children aged 12–54 months.

The need for the translation of research to practice within developmental psychopathology is crucial for the discipline to realize its full potential. The present work represents an initial step in the development of shorter protocols, which are more efficient for clinical and applied use. From a clinical perspective, it will be important to determine whether the specific items identified herein can be coded accurately in real time, which is a practical requirement of many agencies working with high-risk families. This work is currently under way in a group of collaborating clinical agencies. Methodologically, there is great potential for the IRT framework to bridge research and clinical objectives. Understanding the salience of discrete caregiving behaviors for the developing human organism in the earliest years of life is of immense importance. We believe the current work represents an important step in a larger enterprise aimed at identifying the neurobiological and psychological signatures of early social–environmental experiences. For example, this work is especially relevant to recent research conceptualizing childhood adversity around dimensions of deprivation and threat (Sheridan & McLaughin, 2014; Teicher, Samson, Anderson, & Ohashi, 2016), because it offers a fine-grained examination of disrupted caregiving behaviors characterizing both maternal withdrawal and maternal intrusiveness, which may be thought of as downward extensions of concepts of deprivation and threat to infancy and early childhood. The behaviors identified in these analyses, then, are anchored in strong conceptual frameworks from which to launch focused investigations of the effects of early social experience on neural and physiological development.

References

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

Albanese, M. T., & Knott, M. (1994). Bootstrapping latent variable models for binary response. *British Journal of Mathematical and Statistical Psychology*, *47*, 235–246. doi:10.1111/j.2044-8317.1994.tb01036.x

Baker, F. B. (2001). *The basics of item response theory*. College Park, MD: University of Maryland, ERIC Clearinghouse on Assessment and Evaluation.

Bernier, A., & Meins, E. (2008). A threshold approach to understanding the origins of attachment disorganization. *Developmental Psychology*, *44*, 969–982. doi:10.1037/0012-1649.44.4.969

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of EM algorithm. *Psychometrika*, *46*, 443–459. doi:10.1007/BF02293801

Bronfman, E., Madigan, S., & Lyons-Ruth, K. (2009–2014). *Disrupted Maternal Behavior Instrument for Assessment and Classification (AMBIANCE): Manual for coding disrupted affective communication* (2nd ed.). Unpublished manuscript, Harvard University Medical School.

Bronfman, E. T., Parsons, E., & Lyons-Ruth, K. (1992–2008). *Disrupted Maternal Behavior Instrument for Assessment and Classification (AMBIANCE): Manual for coding disrupted affective communication* (1st ed.). Unpublished manuscript, Harvard University Medical School.

Bronfman, E. T., Parsons, E., & Lyons-Ruth, K. (1992–2004). *Disrupted Maternal Behavior Instrument for Assessment and Classification (AMBIANCE), Version 2.1*. Unpublished manuscript, Harvard Medical School.

Canty, A., & Ripley, B. (2016). boot: Bootstrap R (S-Plus) Functions. R package version 1.3-18 [Computer software]. Vienna: R Foundation for Statistical Computing.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159. doi:10.1037//0033-2909.112.1.155

Cole, D. A., Cai, L., Martin, N. C., Findling, R. L., Youngstrom, E. A., Garber, J., . . . Forehand, R. (2011). Structure and measurement of depression in youths: Applying response theory to clinical data. *Psychological Assessment*, *23*, 819–833. doi:10.1037/a0023518

Cook, K. F., Kallen, M. A., & Amtmann, D. (2009). Having a fit: Impact of the number of items and distribution of data on traditional criteria for assisting IRT's unidimensionality assumption. *Quality of Life Research*, *18*, 447–460. 10.1007/s11136-009-9464-4

Cooke, D. J., & Michie, C. (1997). An item response theory analyses of the Hare Psychopathy Checklist—Revised. *Psychological Assessment*, *9*, 3–14. doi:10.1037/1040-3590.9.1.3

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302. doi:10.1037/h0040957

Cyr, C., Euser, E. M., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2010). Attachment security and disorganization in maltreating and high-risk families: A series of meta-analyses. *Development and Psychopathology*, *22*, 87–108. doi:10.1017/S0954579409990289

de Menezes, L. M. (1999). On fitting latent class models for binary data: The estimation of standard errors. *British Journal of Mathematical and Statistical Psychology*, *52*, 149–168. doi:10.1348/000711099159026

Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research*, *16*, 5–18. doi:10.1007/s11136-007-9198-0

Egeland, B., Pianta, R., & O'Brien, M. (1993). Maternal intrusiveness in infancy and child maladaptation in early school years. *Development and Psychopathology*, *5*, 359–370. doi:10.1017/S0954579400004466

Fearon, R. P., Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., Lapsley, A.-M., & Roisman, G. I. (2010). The significance of insecure attachment and disorganization in the development of children's externalizing behavior: A meta-analytic study. *Child Development*, *81*, 435–456. doi:10.1111/j.1467-8624.2009.01405.x

Finkelman, M. D., Green, J. G., Gruber, M. J., & Zaslavsky, A. M. (2011). A zero- and K-inflated mixture model for health questionnaire data. *Statistics in Medicine*, *30*, 1028–1043. doi:10.1002/sim.4217

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, *7*, 286–299. doi:10.1037/1040-3590.7.3.286

Forbes, L. M., Evans, E. M., Moran, G., & Pederson, D. R. (2007). Change in disrupted maternal behavior predicts change in attachment disorganization from 12 to 24 months in a high-risk sample. *Child Development*, *78*, 955–971. doi:10.1111/j.1467-8624.2007.01043.x

Forero, C. G., & Maydeu-Olivares, A. (2009). Estimation of IRT graded response models: Limited versus full information methods. *Psychological Methods*, *14*, 275–299. doi:10.1037/a0015825

Fraley, R. C., Waller, N. G., & Brennan, K. A. (2000). An item response theory analysis of self-report measures of adult attachment. *Journal of Personality and Social Psychology*, *78*, 350–365. doi:10.1037//0022-3514.78.2.350

Goldberg, S., Gotowiec, A., & Simmons, R. J. (1995). Infant-mother attachment and behavior problems in healthy and chronically ill preschoolers. *Developmental and Psychopathology*, *7*, 267–282. doi:10.1017/S0954579400006490

Gordon, R. A. (2015). Measuring constructs in family science: How can item response theory improve precision and validity? *Journal of Marriage and Family*, *77*, 147–176. doi:10.1111/jomf.12157

Haltigan, J. D., Leerkes, E., Supple, A. J., & Calkins, S. D. (2013). Infant negative affect and maternal interactive behavior during the still-face procedure: The moderating role of adult attachment states of mind. *Attachment and Human Development*, *16*, 149–173. doi:10.1080/14616734.2013.863734

Haltigan, J. D., Leerkes, E., Wong, M. S., Fortuna, K., Roisman, G. I., Supple, A. J., . . . Plamondon, A. (2014). Adult attachment states of mind: Measurement invariance across ethnicity and associations with maternal sensitivity. *Child Development*, *85*, 1019–1035. doi:10.1111/cdev.12180

Haltigan, J. D., & Roisman, G. I. (2015). Infant attachment insecurity and dissociative symptomatology: Findings from the NICHD study of early child care and youth development. *Infant Mental Health Journal*, *36*, 30–41. doi:10.1002/imhj.21479

Hobson, R. P., Patrick, M., Crandell, L., García-Pérez, R., & Lee, A. (2005). Personal relatedness and attachment in infants of mothers with borderline personality disorder. *Development and Psychopathology*, *17*, 329–347. doi:10.1017/S0954579405050169

Hobson, R. P., Patrick, M., Hobson, J. A., Crandell, L., Bronfman, E., & Lyons-Ruth, K. (2009). How mothers with borderline personality disorder relate to their one-year-old infants. *British Journal of Psychiatry*, *195*, 1–6. doi:bjp.bp.108.060624

Kim, Y., & Pilkonis, P. A. (1999). Selecting the most informative items in the IIP scales for personality disorders: An application of item response theory. *Journal of Personality Disorders*, *13*, 157–174. doi:10.1521/pedi.1999.13.2.157

Krueger, R. F., Nichol, P. E., Hicks, B. M., Markon, K. E., Patrick, C. J., Iacono, W. G., & McGue, M. (2004). Using latent trait modeling to conceptualize an alcohol problems continuum. *Psychological Assessment*, *16*, 107–119. doi:10.1037/1040-3590.16.2.107

Langenbucher, J. W., Labouvie, E., Martin, C. S., Sanjuan, P. M., Bavly, L., Kirisci, L., & Chung, T. (2004). An application of item response theory analysis to alcohol, cannabis, and cocaine criteria in DSM-IV. *Journal of Abnormal Psychology*, *113*, 72–80. doi:10.1037/0021-843X.113.1.72

Lyons-Ruth, K., Alpern, L., & Repacholi, B. (1993). Disorganized infant attachment classification and maternal psychosocial problems as predictors of hostile-aggressive behavior in the preschool classroom. *Child Development*, *64*, 572–585. doi:10.2307/1131270

Lyons-Ruth, K., Bronfman, E., & Atwood, G. (1999). A relational diathesis model of hostile-helpless states of mind: Expressions in mother-infant interaction. In J. Solomon & C. George (Eds.), *Attachment disorganization* (pp. 33–70). New York: Guilford Press.

Lyons-Ruth, K., Bronfman, E., & Parsons, E. (1999). Maternal frightened, frightening or disrupted behavior and disorganized infant attachment patterns. *Monographs of the Society for Research in Child Development*, *64*(3, Serial No. 258).

Lyons-Ruth, K., Bureau, J.-F., Holmes, B., Easterbrooks, A., & Brooks, N. (2013). Borderline symptoms and suicidality/self-injury in late adolescence: Prospectively observed relationship correlates in infancy and childhood. *Psychiatry Research*, *206*, 273–281. doi:10.1016/j.psychres.2012.09.030

Lyons-Ruth, K., Connell, D., Grunebaum, H., & Botein, S. (1990). Infants at social risk: Maternal depression and family support services as mediators of infant development and security of attachment. *Child Development*, *61*, 85–98. doi:10.1111/j.1467-8624.1990.tb02762.x

Lyons-Ruth, K., & Jacobvitz, D. (2016). Attachment disorganization from infancy to adulthood: Neurobiological correlates, parenting context, and pathways to disorder. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (3rd ed., pp. 667–695). New York: Guilford Press.

Lyons-Ruth, K., Pechtel, P., Yoon, S. A., Anderson, C. M., & Teicher, M. H. (2016). Disorganized attachment in infancy predicts greater amygdala volume in adulthood. *Behavioural Brain Research*, *308*, 83–93. doi:10.1016/j.bbr.2016.03.050

Macfie, J., McElwain, N. L., Houts, R. M., & Cox, M. J. (2005). Intergenerational transmission of role reversal between parent and child: Dyadic and family systems internal working models. *Attachment and Human Development*, *7*, 51–65. doi:10.1080/14616730500039663

Madigan, S., Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., Moran, G., Pederson, D. R., & Benoit, D. (2006). Unresolved states of mind, anomalous parental behavior, and disorganized attachment: A review and meta-analysis of a transmission gap. *Attachment and Human Development*, *8*, 89–111. doi:10.1080/14616730600774458

Madigan, S., Brumariu, L. E., Villani, V., Atkinson, L., & Lyons-Ruth, K. (2016). Representational and questionnaire measures of attachment: A meta-analysis of relations to child internalizing and externalizing problems. *Psychological Bulletin*, *142*, 367–399. doi:10.1037/bul0000029

Madigan, S., Voci, S., & Benoit, D. (2011). Stability of disrupted caregiver behaviors over six years and associations with disorganized infant-caregiver attachment. *Attachment and Human Development*, *13*, 237–252. doi:10.1080/14616734.2011.562410

Main, M., & Hesse, E. (1990). Parents' unresolved traumatic experiences are related to infant disorganized status: Is frightened and/or frightening parental behavior the linking mechanism? In M. T. Greenberg, D. Cicchetti, & E. M. Cummings (Eds.), *Attachment in the preschool years: Theory, research, and intervention* (pp. 161–182). Chicago: University of Chicago Press.

Main, M., & Solomon, J. (1990). Procedures for identifying infants as disorganized/disoriented during the Ainsworth Strange Situation. In M. T. Greenberg, D. Cicchetti, & M. Cummings (Eds.), *Attachment in the preschool years: Theory, research, and intervention* (pp. 121–160). Chicago: University of Chicago Press.

Martin, C. S., Chung, T., Kirisci, L., & Langenbucher, J. W. (2006). Item response theory analysis of diagnostic criteria for alcohol and cannabis use disorders in adolescents: Implications for DSM-V. *Journal of Abnormal Psychology*, *115*, 807–814. doi:10.1037/0021-843X.115.4.807

Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: Causes, consequences, and remedies. *Psychological Methods*, *9*, 147–163. doi:10.1037/1082-989X.9.2.147

Mayseless, O., Bartholomew, K., Henderson, A., & Trinke, S. (2004). "I was more her mom than she was mine": Role reversal in a community sample. *Family Relations*, *53*, 78–86. doi:10.1111/j.1741-3729.2004.00011.x

Mills-Koonce, W. R., Willoughby, M. T., Rehder, P. D., Gottfredson, N., Bronfman, E., Madigan, S., & Lyons-Ruth, K. (2017). *Associations between disrupted maternal caregiving and oppositional defiant and callous-unemotional behaviors differ for boys versus girls*. Manuscript submitted for publication.

Muthén, B. O. (1989). Dichotomous factor analysis of symptom data. *Sociological Methods and Research*, *18*, 19–65.

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles: Author.

NICHD Early Child Care Research Network. (1997). The effects of infant childcare on infant-mother attachment security: Results of the NICHD study of early childcare. *Child Development*, *68*, 860–879. doi:10.2307/1132038

NICHD Early Child Care Research Network. (2001). Child-care and family predictors of preschool attachment and stability from infancy. *Developmental Psychology*, *37*, 847–862. doi:10.1037/0012-1649.37.6.847

NICHD Early Child Care Research Network. (2004). Father's and mother's parenting behavior and beliefs as predictors of child social adjustment in the transition to school. *Journal of Family Psychology*, *18*, 628–638. doi:10.1037/0893-3200.18.4.628

NICHD Early Child Care Research Network. (2005). *Child care and child development*. New York: Guilford Press.

Ondersma, S. J., Chaffin, M. J., Mullins, S. M., & LeBreton, J. M. (2005). A brief form of the child abuse potential inventory: Development and validation. *Journal of Clinical Child and Adolescent Psychology*, *34*, 300–311. doi:10.1207/s15374424jccp3402_9

Osterlind, S. J., & Everson, H. T. (2009). *Differential item functioning* (Vol. 161). Thousand Oaks, CA: Sage.

R Core Team. (2016). R: A language and environment for statistical computing [Computer software]. Vienna: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology*, *5*, 27–48. doi:10.1146/annurev.clinpsy.032408.153553

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, *17*, 1–25.

Sheridan, M. A., & McLauglin, K. A. (2014). Dimensions of early experience and neural development: Deprivation and threat. *Trends in Cognitive Sciences*, *18*, 580–585. doi:10.1016/j.tics.2014.09.001

Shi, Z., Bureau, J.-F., Easterbrooks, M. A., Zhao, X., & Lyons-Ruth, K. (2012). Childhood maltreatment and prospectively observed quality of early care as predictors of antisocial personality disorder features. *Infant Mental Health Journal*, *33*, 55–69. doi:10.1002/imhj.20295

Sroufe, L. A., Jacobvitz, D., Mangelsdorf, S., DeAngelo, E., & Ward, M. J. (1985). Generational boundary dissolution between mothers and their preschool children: A relationship systems approach. *Child Development*, *56*, 317–325. doi:10.2307/1129722

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*, 245–251. doi:10.1037/0033-2909.87.2.245

Sterba, S. K., Copeland, W., Egger, H. L., Costello, E. J., Erkanli, A., & Angold, A. (2010). Longitudinal dimensionality of adolescent psychopathology: Testing the differentiation hypothesis. *Journal of Child Psychology and Psychiatry*, *51*, 871–884. doi:10.1111/j.1469-7610.2010.02234.x

Swaminathan, H., Hambleton, R. K., Sireci, S. G., Xing, D., & Rizavi, S. M. (2003). Small sample estimation in dichotomous item response models: Effect of priors based on judgmental information on the accuracy of item parameter estimates. *Applied Psychological Measurement*, *27*, 27–51. doi:10.1177/0146621602239475

Teicher, M., Samson, J., Anderson, C., & Ohashi, K. (2016). The effects of childhood maltreatment on brain structure, function and connectivity. *Nature Reviews Neuroscience*, *17*, 652–666. doi:10.1038/nrn.2016.111

Tereno, S., Madigan, S., Lyons-Ruth, K., Plamondon, A., Atkinson, L., Guedeney, N., . . . Guedeney, A. (2017). Assessing mechanisms of effect in a randomized home-visiting trial: Reduced disrupted maternal communication decreases infant disorganization. *Development and Psychopathology*, *29*, 637–649. doi:10.1017/S0954579417000232

van IJzendoorn, M. H., Schuengel, C., & Bakermans-Kranenburg, M. J. (1999). Disorganized attachment in early childhood: Meta-analysis of precursors, concomitants, and sequalae. *Development and Psychopathology*, *11*, 225–249. doi:10.1017/S0954579499002035

Wall, M. M., Park, J. Y., & Moustaki, I. (2015). IRT modeling in the presence of zero-inflation with application to psychiatric disorder severity. *Applied Psychological Measurement*, *39*, 583–597. doi:10.1177/0146621615588184