



# The Eyberg Child Behavior Inventory: Developing an Adaptation for Toddler-aged Children

Sharon T. Phillips<sup>1</sup> · Lindsay R. Druskin<sup>1</sup> · Christa Lilly<sup>2</sup> · Felipa T. Chavez<sup>3</sup> · Sue Morgan<sup>4</sup> · Nancy Wallace<sup>4</sup> · Sara Cibralic<sup>5</sup> · Eva R. Kimonis<sup>6</sup> · David J. Hawes<sup>7</sup> · Valsamma Eapen<sup>5,8,9</sup> · Cheryl B. McNeil<sup>1,10</sup> · Jane Kohlhoff<sup>6,8,11</sup>

Accepted: 27 July 2023 / Published online: 30 August 2023  
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

## Abstract

Disruptive behavior problems among young children can result in long-term negative consequences, highlighting the importance of early interventions. While there have been recent developments in early interventions (e.g., Parent–Child Interaction Therapy–Toddler), there is a need for brief assessments for toddler disruptive behaviors. The current study aims to adapt the Eyberg Child Behavior Inventory (ECBI) for toddler-aged children. Archival data from toddler- and preschool-aged children underwent a three-step process for data reduction (qualitative content validation, missing data and low variability exclusion, and criterion-related validation against the preschool version of the Child Behavior Checklist [CBCL]). Exploratory and confirmatory factor analyses with measurement models on the reduced item set were conducted, with a final validation model against the CBCL. This procedure resulted in eight of the 36 items of the ECBI being retained. A metric measurement model of the 8 items, allowing for higher average scores for preschool children compared to toddlers, fit well (chi-square  $p=0.13$ ; SRMR=0.07, GFI=0.98, NFI=0.94). This study provides evidence for a developmental factor of the ECBI for toddler-aged youth that can be used in future clinical and research work. Items included on this factor support previous research that suggests toddler behavior problems are partially due to undeveloped emotion regulation skills and parent–child attachment concerns.

**Keywords** Toddler · Factor analysis · Assessment · Externalizing problems · Infant

## Introduction

The presence of disruptive behaviors (e.g., aggression, defiance, hyperactivity, emotional outbursts) is a common concern for caregivers of young children. Disruptive behaviors are among a constellation of symptoms consistent with several disorders including oppositional defiant disorder,

conduct disorder, attention-deficit/hyperactivity disorder, and disruptive mood dysregulation disorder. Disruptive behaviors can hinder development in several areas as children may experience difficulty making and maintaining peer relationships, experience conflict in home and school settings, struggle academically, and if severe enough, may even be expelled from school (Milledge et al., 2019; Stormshak

✉ Sharon T. Phillips  
stp00001@mix.wvu.edu

<sup>1</sup> Department, of Psychology, West Virginia University, 53 Campus Drive, Morgantown, WV 26506-6040, USA

<sup>2</sup> Department of Epidemiology and Biostatistics, West Virginia University, Morgantown, WV, USA

<sup>3</sup> Department of Psychology, Florida Institute of Technology, Melbourne, FL, USA

<sup>4</sup> Karitane, Sydney, Australia

<sup>5</sup> Ingham Institute for Applied Medical Research, Sydney, Australia

<sup>6</sup> School of Psychology, Parent-Child Research Clinic, University of New South Wales, Sydney, Australia

<sup>7</sup> School of Psychology, University of Sydney, Sydney, Australia

<sup>8</sup> Discipline of Psychiatry and Mental Health, University of New South Wales, Sydney, Australia

<sup>9</sup> Academic Unit of Child Psychiatry, South Western Sydney Local Health District, Sydney, Australia

<sup>10</sup> Department of Psychiatry, University of Florida, Gainesville, FL, USA

<sup>11</sup> Research Department, Karitane, Sydney, Australia

et al., 1998). Furthermore, without effective treatment, a host of negative outcomes across the lifespan have been associated with childhood disruptive behavior, including delinquency, unemployment, and substance use (Bongers et al., 2004; Copeland et al., 2007; Kokko & Pulkkinen, 2000; van Emmerik-van Oortmerssen et al., 2012). Thus, it is critical to accurately assess and characterize these behaviors in early childhood in order to implement early intervention.

The majority of the extant literature on disruptive behavior in children utilizes a parent report of child behavior (Epstein et al., 2015). Several common parent-report measures of child disruptive behavior include the Child Behavior Checklist (CBCL; ages 1.5 to 18 years, Achenbach, 1999), Strengths and Difficulties Questionnaire (SDQ, ages 4 to 17 years, Goodman, 1997), and the Eyberg Child Behavior Inventory (ECBI; ages 2 to 16 years old; Eyberg & Pincus, 1999). Each of these questionnaires has a complementary teacher-report version to capture disruptive behaviors in the school setting. Furthermore, a standardized behavioral observational system, such as the Dyadic Parent–Child Interaction Coding System (DPICS-IV, Eyberg et al., 2014) or the ADHD Behavior Coding System (Barkley, 1998), can be used to assess child disruptive behaviors observed across home or school settings. While there are several widely used measures of disruptive behavior in older children, there remains a dearth of measures to capture these behaviors among toddler-aged children (aged 1 to 2 years).

Disruptive behaviors in toddlers are conceptually different from disruptive behaviors in preschool- or school-aged children, which typically develop and are maintained by coercive parent–child interaction patterns (Patterson, 1982; Smith et al., 2012). This is because toddlers have not yet developed the higher-level cognitions to purposefully engage in these behaviors to gain the attention of their caregivers. While caregivers may experience a variety of challenging behaviors when caring for toddlers, the behaviors are likely to lack the instrumental quality that may be present in older children. Behaviors that may be perceived as “disruptive” in toddlers are often characterized by a lack of emotion regulation capabilities as this process is developing in this stage of life (Calkins & Johnson, 1998). As children explore their environment, learn to problem-solve, and develop self-regulatory capacities, it is developmentally appropriate for toddler-aged children to take toys from others, refuse to listen to caregiver requests, and experience difficulty managing emotional responses (Campbell et al., 2000; Wilks et al., 2010). Moreover, the problematic nature of child behavior is inherent to the developmental stage of the child, as a behavior considered normative for toddlers may be of clinical significance for an older child. Nevertheless, impulsivity, emotional reactivity, and aggressive behavior are common concerns for parents of toddlers seeking treatment. Research on toddler-aged children indicates that impairing disruptive behaviors

can begin at a very young age and that when severe behaviors are present across multiple domains (e.g., cognitive, social, emotional) as well as in multiple settings (e.g., home, school) it may be indicative of a negative developmental trajectory of behavior and conduct problems across later childhood and adulthood (Gardner et al., 2007; Reef et al., 2011). To address these behavior problems, developmental adaptations of parent training programs, such as the toddler adaptation of Parent–Child Interaction Therapy (PCIT-T), are being evaluated to determine their efficacy at treating a myriad of toddler and parental symptoms, including disruptive behaviors (Kohlhoff et al., 2020a, b, 2021). Consequently, there is a critical need for measures that can capture disruptive behavior in toddlers and assess their change in response to early interventions such as PCIT-T.

Although there are several widely used measures of disruptive behavior for older children, there remains a dearth of such measures for use with toddler-aged children (aged one to two years). When evaluating toddler-aged children, two measures of social-emotional and behavioral difficulties are most commonly used: the Devereux Early Childhood Assessment (DECA; LeBuffe & Naglieri, 2003, 2009; Mackrain et al., 2007) and the Brief Infant–Toddler Social Emotional Assessment (BITSEA; Briggs-Gowan & Carter, 2002). These measures are not suitable for weekly monitoring of toddler’s disruptive behaviors during intervention due to their length and limited sensitivity to change (McClendon et al., 2010). Another measure, the Toddler and Preschool Behavior Scale (Holtz et al., 2008) was developed to capture disruptive behavior in young children; however, there is limited evidence supporting its psychometric properties, thus limiting its generalizability. A commonly used measure for monitoring treatment-related change in disruptive behaviors is the Eyberg Child Behavior Inventory (ECBI), a caregiver report measure of disruptive behaviors for children aged 2 to 18 years old (Eyberg & Pincus, 1999). The ECBI is relatively brief and also sensitive to change and thus one of its primary uses is in the context of Parent–Child Interaction Therapy (PCIT), an evidence-based treatment for children aged 2 to 7 years old with disruptive behavior problems. The ECBI is administered as a routine part of weekly PCIT sessions to monitor child disruptive behavior symptoms and guide treatment (Eyberg & Pincus, 1999). ECBI scores have been found to have high internal consistency, test–retest reliability, and convergent validity with scores on the Preschool Behavior Questionnaire in preschool-aged children (Burns et al., 1991; Funderburk et al., 2003; Morawska & Sanders, 2006). There is also evidence supporting the convergent and discriminant validity of ECBI scores in ethnically diverse samples (Machado, 2020). Weis and colleagues (2005) found that ECBI scores differentiated between clinic-referred children with and without externalizing symptoms. Overall, the ECBI appears to be a psychometrically sound measure of

child behavior problems, but its psychometric properties and clinical utility with toddlers has yet to be investigated. As the ECBI is only validated for children aged 2 and older, there is also a need for a validated measure of child disruptive behaviors similar to the ECBI for children younger than the age of two, particularly as developmental adaptations of PCIT for toddler-aged children are being developed and evaluated (Girard et al., 2018; Kohlhoff et al., 2021).

## Current Study

As many PCIT therapists become trained in PCIT adaptations, such as PCIT-T, the development of a toddler adaptation of the ECBI allows for streamlined assessment for PCIT therapists who are particularly familiar with the standard ECBI. The ECBI is brief, easy to score, and widely used as an assessment of child behavior concerns. The high degree of clinical utility makes the ECBI a reasonable choice for regular assessment during early intervention programs. In this study, we aimed to develop a developmentally appropriate version of the ECBI for toddlers aged 12–24 months. The current study aimed to 1) determine which ECBI items are appropriate for use in a toddler adaptation, 2) evaluate the content validity of scores on this toddler adaptation of the ECBI using expert data, and 3) evaluate the convergent validity evidence for the toddler adaptation of the ECBI by comparing its scores to scores on externalizing subscales of the CBCL 1.5–5. To this end, the ECBI underwent an initial data reduction against three criteria (content validation via qualitative expert survey, reducing items with high percent missing and lack of variability, and criterion-related validity by comparing scores on the toddler version of the ECBI against a validated measure of child externalizing behaviors [i.e., CBCL]) in a sample of toddlers. Factor analyses (exploratory, EFA, and confirmatory, CFA) were then conducted on the reduced ECBI item set, comparing toddlers against preschool children. Once the best-fitting measurement model was determined, we conducted a final model against the CBCL. We hypothesized that only a subset of items on the ECBI would load together onto a ‘toddler factor,’ but that factors on the ECBI would be similar for older and younger children, with older children having higher ECBI scores than younger children. We also hypothesized that experts would identify specific items on the ECBI that are inappropriate for assessing disruptive behavior in toddler-aged children.

## Method

This project has been approved by West Virginia University’s Institutional Review Board (protocol number: 2111474890).

## Sample 1

Participants in Sample 1 ( $n = 160$ ) were recruited as part of two randomized controlled trials evaluating the efficacy of PCIT-T (see Kohlhoff et al., 2020b and Kohlhoff et al., 2021 for full descriptions of the RCT protocols). Data collection occurred at the Karitane Toddler Clinic, a community-based child treatment center providing evidence-based parenting services for families with toddlers and young children in Australia. General exclusionary criteria for the Karitane Toddler Clinic include parents with current severe depression with suicidality, psychosis, or other serious mental health conditions causing severe impairment. Children included in this sample ranged from 14 to 24 months ( $M = 19.24$  months,  $SD = 2.85$ ) and 51.2% were boys ( $n = 82$ ). All caregivers were mothers, who averaged 32.55 years old ( $SD = 5.29$ ), and 72.5% were partnered (married or de-facto). The sample was ethnically diverse, with 31.9% speaking a language other than English in the home. For these participants, 48.7% were university educated. Henceforth in this paper, this sample of children will be referred to as the ‘young’ sample.

## Sample 2

Participants in Sample 2 ( $n = 100$ ) were recruited from the Karitane Toddler Clinic for a study examining treatment outcomes for young children presenting with subtypes of early childhood disruptive behaviors. This sample had the sample exclusionary criteria as Sample 1. Children included in this sample ranged from 2 to 4 years of age ( $M = 36.28$  months,  $SD = 7.73$ ) and 62% were boys ( $n = 62$ ). Mothers averaged 33.11 years ( $SD = 5.16$ ) and 77.8% were partnered (married or de-facto). Only one parent completed questionnaires per child and 85% of the questionnaire-completing parents were mothers. For these participants, 42.6% were university educated. Of the families in the study, 11.4% spoke a language other than English in the home. Henceforth in this paper, this sample of children will be referred to as the ‘old’ sample. Demographic information for both samples is provided in Table 1.

## Procedure

At the intake session for therapeutic services at the Karitane Toddler Clinic, written parental consent was obtained for data to be used for research purposes. Prior to beginning treatment, parents completed a battery of questionnaires about the child’s behavior.

Prior to conducting statistical analyses and after collecting all data, the researchers sent experts in the field of PCIT and toddler populations a survey to gather information on the content validity of items of the ECBI for the toddler population. The ‘experts’ were chosen because they worked in a

**Table 1** Demographics and descriptive statistics

Variable	Overall		Old		Young	
	<i>M</i> or <i>n</i>	<i>SD</i> or %	<i>M</i> or <i>n</i>	<i>SD</i> or %	<i>M</i> or <i>n</i>	<i>SD</i> or %
Age Group	229					
Old (> 2 years)	90	39.3%	--	--	--	--
Young ( $\leq$ 2 years)	139	60.7%	--	--	--	--
Age (in months)	<i>N</i> = 220					
	25.6	9.8	36.3	8.0	19.2	2.8
Child Gender	221					
Male	129	58.4%	54	65.1%	75	54.4%
Female	92	41.6%	29	34.9%	63	45.7%
CBCL Total	<i>N</i> = 209					
	57.8	24.9	66.8	25.6	52.7	23.0
CBCL Total Cut-off						
$\leq$ 61 ( <i>not</i> at risk)	123	58.9%	33	43.4%	90	67.7%
> 61 (at risk)	86	41.2%	43	56.6%	43	32.3%

CBCL Child behavior checklist

research program that published on the use of PCIT with this age range, were authors of the PCIT-Toddler book (Girard et al., 2018), or were clinicians that used PCIT with toddler-aged children. The survey was sent to 23 experts along with an email requesting their assistance with this project, and 21 experts completed the survey yielding a 91.3% completion rate. Experts were in three countries (United States of America, Australia, and New Zealand), were predominantly female (90.4%), and were primarily academic researchers (81%). Experts all completed at least one higher-education degree (one bachelor's-level individual, three masters-level experts, eight doctoral students, and nine doctoral-level experts). Responses from the expert survey were used in two ways: first, to guide a theoretical understanding of which items would be developmentally appropriate for toddler-aged children, and second, as one of the methods for the exclusion of items on the final measure (Criteria 3; Table 2).

## Measures

### The Child Behavior Checklist

The Child Behavior Checklist 1.5–5 (CBCL; Achenbach, 1999) is a measure of a broad array of child psychological symptoms and disorders for youth aged 1.5 years to 5 years old. This measure consists of 99 items that reflect child behaviors, and one item where caregivers can write in additional problems the child has that were not reflected in the items. Items are rated on a scale from 0 (Not True) to 2 (Very True or Often True) based on whether the child has displayed the behavior in the past two months or is currently displaying the behavior. This measure yields 14 subscale scores reflecting different syndromes and DSM disorders that children may present with, an externalizing problems

score, an internalizing problems score, and a total score. A systematic review of the psychometric evidence for the CBCL suggested that CBCL scores have good internal consistency and moderate levels of convergent validity (Gridley et al., 2019). In this study, Cronbach's alpha was high for the combined sample ( $\alpha = .944$ ), the younger sample ( $\alpha = .941$ ), and the older sample ( $\alpha = .941$ ).

### Eyberg Child Behavior Inventory

The Eyberg Child Behavior Inventory (ECBI; Eyberg & Pincus, 1999) is a 36-item parent-report measure of child disruptive behaviors. The ECBI consists of two subscales: one that measures the intensity or frequency of 36 common childhood disruptive behaviors and the other that dichotomously measures whether each disruptive behavior is a problem for the caregiver. Each of the ECBI items reflects a different disruptive behavior; each item is rated both on their frequency using a Likert-type scale from 1 (Never) to 7 (Always) and on whether that particular behavior is a problem to the caregiver using a Yes/No response. As stated above, ECBI scores have strong internal consistency (Burns et al., 1991; Morawska & Sanders, 2006), good convergent and discriminant validity (Machado, 2020; Weis et al., 2005), and good levels of sensitivity and specificity (Gridley et al., 2019). In this study, Cronbach's alpha was high for the combined sample ( $\alpha = .927$ ), the younger sample ( $\alpha = .904$ ), and the older sample ( $\alpha = .915$ ).

### Qualitative Survey

For the purposes of the current study, a survey was developed to evaluate the appropriateness of ECBI items and sent to identified experts within the field. This survey requested

**Table 2** Decision tree for item inclusion

Item Number	Item Name	Criteria 1: Difference in Missing; Difference in Never + Missing Responses					Criteria 2: Related to Total and Externalizing CBCL Scores			Criteria 3: Compare to Experts	Total Criteria Met	Decision
		Old: % Missing	Young: % Missing	Old: Cumulative % (Missing + Never)	Young: Cumulative % (Missing + Never)	Fisher Exact p value	Total CBCL Cut-off, p value	Externalizing CBCL p value	Met Criteria 2			
1	Dawdles in getting dressed	1.11	5.04	6.67	25.9	.00	.73	.76	0.0	0.0	0.0	NO
2	Dawdles or lingers at mealtime	2.22	2.88	7.78	16.55	.07	.09	.20	0.0	1.0	2.0	NO
3	Has poor table manners	2.22	3.6	8.89	17.99	.08	.35	.07	0.0	0.0	1.0	NO
4	Refuses to eat food presented	2.22	2.88	6.67	10.07	.48	.03	.15	0.5	1.0	2.5	NO
5	Refuses to do chores when asked	2.22	9.35	4.44	32.37	<.0001	.29	.70	0.0	0.0	0.0	NO
6	Slow in getting ready for bed	2.22	5.04	10	33.81	<.0001	.07	.81	0.0	0.0	0.0	NO
7	Refuses to go to bed on time	1.11	3.6	11.11	26.62	<.001	.04	.45	0.5	1.0	1.5	NO
8	Does not obey house rules on own	3.33	7.91	7.78	28.78	<.0001	.07	<.001	0.5	0.0	0.5	NO
9	Refuses to obey until threatened with punishment	3.33	6.47	10	38.85	<.0001	.02	<.001	1.0	0.5	1.5	NO
10	Acts defiant when told to do something	6.67	6.47	7.78	14.39	.15	.01	<.001	1.0	1.0	3.0	YES
11	Argues with parents about rules	6.67	11.51	23.33	53.24	<.0001	.53	.31	0.0	0.0	0.0	NO
12	Gets angry when doesn't get own way	0	0.72	2.22	1.44	.65	<.0001	<.0001	1.0	1.0	3.0	YES
13	Has temper tantrums	0	0	2.22	3.6	.71	<.001	<.0001	1.0	1.0	3.0	YES
14	Sasses adults	13.33	12.95	33.33	51.08	.01	.10	.04	0.5	0.0	0.5	NO
15	Whines	2.22	3.6	5.56	10.07	.33	<.001	.10	0.5	1.0	2.5	NO
16	Cries easily	1.11	2.88	4.44	9.35	.20	<.0001	<.001	1.0	1.0	3.0	YES
17	Yells or screams	1.11	2.16	4.44	7.91	.41	<.0001	<.0001	1.0	1.0	3.0	YES
18	Hits parents	2.22	2.88	17.27	17.78	1.00	<.001	<.0001	1.0	1.0	3.0	YES
19	Destroys toys and other objects	1.11	1.44	17.78	27.34	.11	<.0001	<.0001	1.0	0.5	2.5	NO
20	Is careless with toys and other objects	1.11	2.16	15.56	19.42	.49	<.0001	<.0001	1.0	0.0	2.0	NO

Table 2 (continued)

Item Number	Item Name	Criteria 1: Difference in Missing; Difference in Never + Missing Responses					Criteria 2: Related to Total and Externalizing CBCL Scores			Criteria 3: Compare to Experts	Total Criteria Met	Decision
		Old: % Missing	Young: % Missing	Old: Cumulative % (Missing + Never)	Young: Cumulative % (Missing + Never)	Fisher Exact p value	Total CBCL Cut-off, p value	Externalizing CBCL p value	Met Criteria 2			
21	Steals	3.33	9.35	72.22	82.01	.10	.26	.58	0.0	0.0	1.0	NO
22	Lies	4.44	10.79	56.67	90.65	<.0001	.14	.10	0.0	0.0	0.0	NO
23	Teases or provokes other children	2.22	9.35	42.22	78.42	<.0001	.31	.24	0.0	0.5	0.5	NO
24	Verbally fights with friends own age	5.56	9.35	50	81.29	<.0001	.90	.57	0.0	0.0	0.0	NO
25	Verbally fights with sisters and brothers	6.67	12.95	54.44	76.98	<.001	.63	.99	0.0	0.0	0.0	NO
26	Physically fights with friends own age	3.33	6.47	45.56	61.87	.02	.01	<.001	1.0	1.0	2.0	NO
27	Physically fights with sisters and brothers	6.67	11.51	42.22	65.47	<.001	.28	.03	0.5	1.0	1.5	NO
28	Constantly seeks attention	1.11	2.88	5.56	10.07	.33	<.0001	<.001	1.0	1.0	3.0	YES
29	Interrupts	3.33	6.47	11.11	28.78	<.001	.01	.01	1.0	0.5	1.5	NO
30	Is easily distracted	3.33	0.72	6.67	6.47	1.00	<.001	<.001	1.0	0.5	2.5	NO
31	Has short attention span	2.22	3.6	5.56	15.11	.03	<.001	<.001	1.0	0.0	1.0	NO
32	Fails to finish tasks or projects	6.67	7.91	10	32.37	<.001	.03	.01	1.0	0.0	1.0	NO
33	Has difficulty entertaining self alone	2.22	2.88	20	17.27	.60	<.001	<.001	1.0	0.5	2.5	NO
34	Has difficulty concentrating on one thing	2.22	3.6	14.44	21.58	.23	.01	.02	1.0	0.5	2.5	NO
35	Is overactive or restless	1.11	3.6	6.67	14.39	.09	<.001	<.001	1.0	1.0	3.0	YES
36	Wets the bed	15.6	21.6	65.56	87.05	<.001	.93	.43	0.0	0.0	0.0	NO

CBCL Child behavior checklist, EFA Exploratory factor analysis

that the expert rate the appropriateness of each item of the ECBI on a scale from 1 (not at all appropriate) to 10 (very appropriate) for 12–24-month-olds. In addition to a numerical rating, experts were also asked to explain their reasoning for their rating for each ECBI item.

## Analysis

### Quantitative Data

All quantitative data analyses were conducted in SAS 9.4. Variables for the ECBI went through a multi-dimensional refinement process prior to factor analysis. This approach was taken in response to i) observations of a large amount of missing data on a number of ECBI items, and ii) feedback from experts suggested that many of the ECBI items may not accurately capture a target behavior or be developmentally appropriate for toddler aged children. After noting that the mean scores for all ECBI items were higher for older than younger children, we refined our variable inclusion chart to meet three criteria: 1) Sufficient variability and similar missing data trends to older children; 2) a relationship to two “gold standard” clinical scales, the CBCL total and externalizing subscale; and 3) a quantitative component of whether experts agreed the item content was appropriate (explained below). Previous work on missing data indicated that reasons for item nonresponse commonly include items not being applicable to the respondent (Huissman, 1999; De Leeuw et al., 2003). As such, the researchers decided to utilize data on missingness as one of the criteria for keeping or removing an item. The first criterion was tested by binary coding each item to 1: missing or “never” v. 0: all other responses, and tested by group using a separate chi-square fisher’s exact test for each item, appropriate for small cell counts. The CBCL is a commonly used, validated, general behavioral assessment for children in the toddler age range. Thus, researchers determined that examining the relationships between an ECBI item and the CBCL scores would be an appropriate way to differentiate between items that are indicative of behavior problems in toddlerhood and items that are not. The second criterion was tested using separate logistic regressions (on the young sample only) with the ECBI item as the independent variable (IV) and the CBCL cut-offs (using clinical cut-offs for the total score > 61, externalizing behavior subscale > 25; Achenbach, 1999) as the dependent variables. The third criterion was chosen to provide a check for content validity for this toddler factor. The third criterion was tested by summing how many experts agreed the item was appropriate for capturing disruptive behaviors in toddlers with an item meeting the criterion if the majority of experts rated it as appropriate. In order to be included in subsequent analyses, items had to meet all three criteria.

Next, an exploratory factor analysis (EFA) was run on the reduced item set from the young sample, using a maximum likelihood method, squared multiple correlation priors, and setting the preferred factor number to 1 (O’Rourke & Hatcher, 2013). As these were ordinal response items, the items were first combined into a polychoric correlation matrix for the entire sample and by age group, and all subsequent analyses run on these polychoric correlation matrices unless otherwise noted. EFA model fit included the eigenvalue and proportion, accompanying squared canonical correlation, the Chi-Square, Akaike’s Information Criterion (AIC; smaller values preferred), Schwarz’s Bayesian Criterion (BIC; smaller values preferred), and the Tucker and Lewis’s Reliability Coefficient (preferred value closer to 1). Additionally factor loadings were reported for each item (desired values > 0.4). Based on EFA results, a single-factor CFA was then conducted to determine appropriateness of allowing correlated errors between items using modification fit indices.

Next, three sets of CFA measurement models were run to determine how consistently or inconsistently these items loaded onto a single factor, allowing for the correlated residuals, overall and by group (young and old; O’Rourke & Hatcher, 2013). This was performed as a set of models from least restrictive to most restrictive. First, a configural model was run with intercepts, parameter estimates (also called slopes or factor loadings), and error variances all allowed to differ for each group. Next, a metric model was run, allowing intercepts and error variances only to differ. Finally, a scalar model was run where both intercept and parameter estimates were forced to be the same between groups, but residual error variances allowed to differ. More restrictive models are typically preferred, as they are more parsimonious and suggest fewer differences between age groups (Lilly, 2022). A final model was selected based on parsimony and model fit criteria that either were better or did not significantly decrease model fit (e.g., SRMR, closer to 0 preferred; goodness-of-fit index (GFI; closer to 1 preferred), and Bentler-Bonett normed fit index (NNFI; closer to 1 preferred; O’Rourke & Hatcher, 2013). Factor loadings for each model are given for the intercepts and slopes, and correlations for the errors allowed to correlate.

Finally, aim 3 compared the final measurement model against the gold standard CBCL using a full structural equation model (SEM) again for the whole sample and by group. The CBCL cut-offs for the total score are presented (using clinical cut-offs of total score > 61). Model fit criteria included RMSEA (preferred < 0.08), SRMR (preferred < 0.08), AGFI and GFI (preferred > 0.90), NNFI (preferred > 0.90; O’Rourke & Hatcher, 2013). Factor loadings are presented for each item.

## Qualitative Data

In experts' solicitation for quantitative appropriateness of ECBI items, they were also asked to offer potential explanations regarding suitability of items for assessing toddlers. Those qualitative narratives were subsequently independently reviewed by two raters assessing for emerging thematic content. This content was ultimately broken down into "Yes" or "No" endorsement as a criterion for inclusion in the EFA based on developmental appropriateness for toddlers. Subsequently, count scores were tabulated across all items for all expert participants. Items were determined as either appropriate for inclusion (i.e., "Yes") or not appropriate (i.e., "No") based on the majority of expert endorsements. Quantitative responses for each item were averaged, with higher scores indicating the item was regarded as more developmentally appropriate for 12–24-month-old children. For aim 2, further qualitative analysis of the expert data was then conducted to clarify expert perceptions about each item, assessing for thematic content across the highly endorsed (i.e., appropriate) and low endorsed (i.e., not appropriate) items. Qualitative responses were coded by two of the study personnel using a conventional content analysis approach, in which thematic categories were defined after an initial review of the qualitative data without preconceived categories (Braun & Clarke, 2006; Hsieh & Shannon, 2005). This approach allowed study authors to include meaningful, data-driven codes and themes. Both researchers who conducted this content analysis had previous experience with qualitative analysis. Data from all experts were reviewed independently before the researchers came together to discuss qualitative codes before re-coding the full dataset. In place of a formal reliability analysis, all qualitative data were double-coded by study authors. Consensus was reached through discussion to resolve any coding disagreements.

## Results

Table 1 depicts the study demographics, including child gender, child age, and descriptive statistics of key study variables. In the total sample, 41.2% of children had CBCL total scores in the clinically significant range. For participants under the age of 2 years, 32.3% had CBCL total scores in the clinically significant range, while 56.6% of participants over the age of 2 years had CBCL total scores in the clinically significant range.

### Aim 1. To Adapt the ECBI for Toddler-Aged Children

#### ECBI Item Inclusion

The first criteria for inclusion of an ECBI items was that a combined 'Never + Missing' score on the Intensity scale for the young sample (i.e., a score of 0 was used when the parent

rated "never" OR if the parent did not answer than item at all; shown in Table 2) was not significantly different than the older children. This decision was made because it was observed that there were both large percentages of missing data and lack of variability on select items (e.g., the young sample had 21.6% missing on the item "wets the bed" and when combined with the "never" response, this total percentage increased to 87.1% of the young sample). Eighteen of the original 36 items met this criterion (see Table 2).

The second criteria for inclusion of an ECBI item in the factor analysis was that the mean score for that item in the young sample needed to be significantly associated with the mean CBCL total and externalizing behavior subscale scores, tested via logistic regression. Eighteen of the original 36 items met this criterion, *p*-values reported in Table 2.

Third, we examined 21 expert responses on the survey about the developmental appropriateness of each item of the ECBI for expert content validation. Averages for individual items ranged from 2.05 to 9.71, while the average standard deviation was 2.33 (ranging from 0.56 to 3.22; see Supplementary Table S1 for descriptive statistics for each item). Fourteen of the original 36 items met quantitative expert agreement for inclusion (see Table 2). In total, eight of the 36 ECBI items met all three criteria and were included in the following steps.

#### EFA

Model fit was generally adequate for the EFA (Table 3), with all items loading above 0.38, and a single factor accounting for 97% of the variability. The factor loadings were deemed appropriate given the breadth of target behavior in young children. For example, items that were focused on more specific or circumscribed behaviors such as "Is overactive or restless" and "acts defiant when told to do something" had lower factor loadings compared to items that were conceptually more clearly related disruptive behavior "gets angry when doesn't get own way" and "has temper tantrums". The squared canonical correlation was appropriately high (0.95) and the Tucker-Lewis reliability coefficient 0.87.

#### CFA

Confirmatory factor analyses (CFAs) were run by age group to determine the appropriateness of allowing correlated errors. Fit modification indices indicated that for the older sample, errors should be correlated between items 12 and 17 ( $p < .0001$ ) and 10 and 12 ( $p = .0012$ ), and for young children the errors should be correlated between items 12 and 13 ( $p < .0001$ ) and 28 and 35 ( $p = .0002$ ). All subsequent models were run allowing these four correlated errors for improved model fit.



We hypothesized the metric model would be the best fitting model, as it suggests the factor loadings are the same between old and young groups, but that older children could have higher on average scores (i.e., intercepts) and that old and young children may have differences driving the variability of residuals. As seen in Table 3, the metric model presents with the best fit or not substantially worse fit than the configural model, with a non-significant chi-square ( $p = .13$ ;  $p > .05$  preferred), smaller fit functions than scalar (smaller is preferred), SRMR at 0.07 (below 0.1 preferred), GFI and NFI identical to configural at 0.98 and 0.94 (close to 1 preferred).

The best fitting model, metric model, allowed for higher average scores for older children compared to young children, with intercepts for older children ranging from 4.18 to 5.86 and intercepts for young children ranging from 3.93 to 5.41. Unstandardized parameter estimates ranged from 0.66 (lowest for “is overactive or restless”) to 1.44 (highest for “yells or screams”).

### **Aim 2. To Evaluate the Content Validity for this Developmental Adaptation of the ECBI Using Expert Data**

A number of general themes emerged from analysis of the expert data: typicality/developmentally appropriate, emotional regulatory issues, child autonomy, parent education about limit setting (education regarding parental limit setting), temperament, intensity/frequency, and developmentally inappropriate. Qualitative themes and example comments for each theme are presented in Table 4. These themes provided a rationale for exclusion of certain items from the toddler scale, which was indicated by overwhelming consensus (ranging from  $n = 16$  to 20) by the total number of experts ( $n = 21$ ). Expert comments may have been coded as being in more than one category if the content of the comment spanned across multiple topics.

### **Aim 3. To Evaluate the Convergent Validity Evidence for the Developmental Adaptation of the ECBI by Comparing those Scores to Scores on Externalizing Subscales of the CBCL**

#### **SEM**

Finally, a full structural equation model (SEM) was conducted using the best fitting measurement model and including CBCL cut-off scores as the outcome. Both CBCL cut-offs for the total score and externalizing behavior scores were run; models were consistent between the two outcomes and for simplicity only CBCL total cut-offs are presented in Table 5 and Fig. 1. Excellent to good SEM model fits were found, including AGFI and

CFI close to 1 (0.96 for both), SRMR below 0.1 (0.07), RMSEA with a confidence interval including the preferred 0.05 (RMSEA: 0.07, 95% CL: 0.04, 0.10). Separate fit functions by group suggested the model fits slightly better for younger than older children (0.34 v. 0.62). Specifically, when examining the relationship of the factor to the outcome of the CBCL total cut-off, the R-square for young children was 0.49 and for older children was 0.42.

## **Discussion**

As adaptations to PCIT continue to develop, it is important to assess the applicability of the ECBI for these novel populations. The current study assessed which items of the ECBI would be applicable for toddler-aged children in response to the development of PCIT-T in 2018 (Girard et al., 2018). This adaptation of the ECBI could also be useful for clinicians and researchers to monitor toddlers' disruptive behavior outcomes when utilizing other early intervention parenting programs, such as Triple P (Sanders, 2012) and Circle of Security (Marvin et al., 2002). Findings from this mixed methods study suggest that eight items from the original 36-item ECBI are relevant when assessing disruptive behaviors in toddler-aged children. These eight items demonstrated preliminary content validity through an analysis of qualitative and quantitative information from experts. Additionally, scores on the 8-item toddler version of the ECBI demonstrated convergent validity via comparisons to the CBCL, an already well-validated measure for similarly aged children. Qualitative responses from experts also highlighted that many of the behaviors listed in ECBI items are present during the toddler-aged period but would not be considered a clinical problem during this developmental stage (e.g., “dawdles in getting dressed”).

A toddler factor from the ECBI provides a comprehensive, yet brief, assessment that is short enough to be administered quickly prior to every session. This tool may overcome the limitations of current validated measures of toddler behavior problems, such as the DECA, CBCL, and BITSEA, which do not capture the entire toddler age range (i.e., 12 months to 24 months) and may be too lengthy to be feasibly administered at every session (McClendon et al., 2010). Moreover, the standard ECBI is sensitive to weekly changes unlike the DECA and the CBCL, which further supports clinician capacity to easily collect weekly data to inform the tailoring of treatment to problems that the family experienced in the past week (McClendon et al., 2010). However, more research would be needed to determine if the toddler scale of the ECBI is also sensitive to weekly changes. While briefer than the DECA and CBCL, the BITSEA captures a broader range of problems than may be

**Table 3** EFA and Measurement Models,  $n = 130$  for young,  $n = 84$  for older children

Model Fit	EFA	Measurement models	Configural (intercepts, slopes, error variance differ)			Metric (intercepts and error variance differ)			Scalar (error variance differ)		
			Overall	Old	Young	Overall	Old	Young	Overall	Old	Young
Eigenvalue	13.79	Chi-square	48.78			54.78			91.03		
Eigenvalue Proportion	0.97	Chi-square df	36			44			52		
Squared Canonical Correlation	0.95	Chi-square $p$ value	0.0757			0.1278			0.0007		
Chi-Square without Bartlett's Correction	71.33	Fit Function	0.23	0.19	0.26	0.26	0.23	0.28	0.43	0.45	0.42
Akaike's Information Criterion	31.33	SRMR	0.04	0.04	0.04	0.07	0.08	0.06	0.12	0.14	0.10
Schwarz's Bayesian Criterion	-26.02	GFI	0.98	0.99	0.98	0.98	0.99	0.98	0.98	0.98	0.97
Tucker and Lewis's Reliability Coefficient	0.87	Bentler-Bonett NFI	0.94	0.95	0.94	0.94	0.94	0.94	0.90	0.88	0.90
<b>Items</b>	<b>Factor Loadings</b>	<b>Old Intercept Slope</b>	<b>Young Intercept Slope</b>	<b>Old Intercept Slope</b>	<b>Young Intercept Slope</b>	<b>Old Intercept Slope</b>	<b>Young Intercept Slope</b>	<b>Old Intercept Slope</b>	<b>Young Intercept Slope</b>	<b>Old Intercept Slope</b>	<b>Young Intercept Slope</b>
10. Acts defiant when told to do something	0.45	5.02	4.00	0.79	4.00	5.02	4.00	0.74	4.52	0.84	0.84
12. Gets angry when doesn't get own way	0.96	5.86	5.41	1.11	5.41	5.86	5.41	1.20	5.59	1.22	1.22
13. Has temper tantrums	0.91	5.57	5.20	1.14	5.20	5.57	5.20	1.19	5.34	1.20	1.20
16. Cries easily	0.66	4.69	4.33	0.81	4.33	4.69	4.33	1.10	4.47	1.11	1.11
17. Yells or screams	0.75	5.31	4.62	1.46	4.62	5.31	4.62	1.44	4.90	1.48	1.48
18. Hits parents	0.59	4.18	3.93	1.17	3.93	4.18	3.93	1.21	4.04	1.21	1.21
28. Constantly seeks attention	0.52	5.16	4.30	0.82	4.30	5.16	4.30	0.92	4.63	0.98	0.98
35. Is overactive or restless	0.38	5.00	4.11	0.58	4.11	5.00	4.11	0.66	4.47	0.74	0.74
<b>Correlated errors</b>			<b>Estimate</b>			<b>Estimate</b>			<b>Estimate</b>		
	E10 E12	0.30				0.28			0.28		0.28
	E12 E13	0.47				0.48			0.46		0.46
	E12 E17	-0.001				-0.02			-0.01		-0.01
	E28 E35	0.77				0.84			0.78		0.78

Table 4 Themes present in high and low endorsed items from experts

Qualitative Theme	Theme Description	Example Quote	Range of Experts' Endorsed Per Question
<i>High Endorsed</i>	<i>Comments related to the item being appropriate for toddlers because ...</i>		
Typicality/ Developmentally Appropriate	... the item is in line with developmental milestones and typical development as toddlers	"This is relevant as the presenting problems for this age range are usually around managing big emotions when they aren't getting what they want"	10–19
Emotional Regulation	... the item captures concerns related to important variations in emotion regulation abilities	"Toddlers have found their voice and yelling/screaming while having a big emotion is common and helpful to know"	1–10
Getting Own Way	... the item captures child behaviors related to the child wanting to exercise their own will and autonomy	"If the child has been taught not to do that and they're doing it ... to get their way (not just when they're having fun and squealing) then I think it could be applicable"	1
Parent Education	... the item captures a need for increased psychoeducation about developmental appropriateness and how to manage challenging behaviors	"This is important to help caregivers know how to set limits and provide consistency in managing big emotions"	1–3
Temperament	... the item captures themes related to child emotion dysregulation that was indicative of the child's innate temperament	"While tantrums are common at this age, they also do serve as good indicators of temperament"	1–3
Intensity/ Frequency	... the item indicates experts' sense that while the behavior was typical, ascertaining the intensity and frequency would be useful for assessing the problematic nature	"It's going to be typical but will give us a reference for intensity"	1–4
<i>Low Endorsed</i>	<i>Comments related to the item being inappropriate for toddlers because ...</i>		
Atypicality/ Developmentally Inappropriate	... the item is not in line with developmental milestones and typical development as toddlers	"Not developmentally appropriate because most 12–24-month old [children] do not have the language skills to argue"	8–20
Variability with Age	... the item may not capture all toddlers, only those on either end of the age range	"I think this might only be appropriate for the upper end of the toddler age range"	2–6
Normative Behavior	... the item is normal for the toddler age range and thus would not provide information on those in need of intervention	"Most toddlers are not toilet trained so I wouldn't expect wetting the bed to be indicative of a behavior problem even though the behavior itself is common in this age range"	1–10

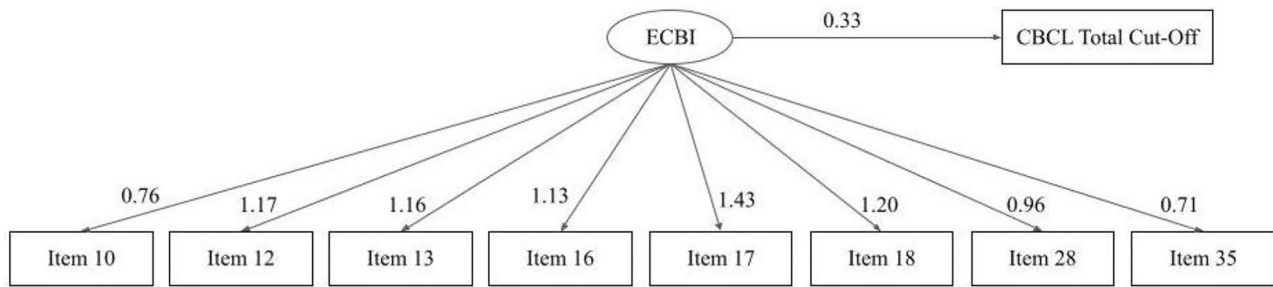
needed for treatment monitoring (e.g., social emotional competencies), and thus may not be appropriate for weekly use (Briggs-Gowan & Carter, 2002). In sum, this toddler-based factor analysis of the ECBI may provide an efficient assessment tool for capturing specific toddler behavior problems in a manner that is sensitive enough to use for continuous monitoring of treatment progress, though additional replication is needed.

Expert data collected as part of this study provided important qualitative information about why certain ECBI items were/were not suitable for the ECBI toddler scale. Results indicated that toddler behavior problems may be

difficult to distinguish from behaviors that are due to typical developmental processes. For example, the ECBI item “wets the bed” may occur during toddlerhood but is not a suitable item for a measure of problematic behavior in this age range as most toddlers would still be wearing diapers overnight or having expected overnight accidents. The qualitative results also suggested that there were some items that were appropriate for inclusion in the ECBI toddler scale due to the impact that they can have on parents, despite them being developmentally expected. An example of this was the item “has temper tantrums.” Experts suggested that this is an example of a behavior that frequently occurs in toddlers and

**Table 5** Structural Equation Model Using Best Fitting Measurement Model (Metric) with Old and Young Children, Predicting CBCL Cut-off Scores, with Correlated Errors

Model Fit	Overall				
Chi-square	91.06				
Chi-square df	59				
Chi-square <i>p</i> value	.005				
AGFI	0.96				
CFI	0.96				
SRMR	0.07				
RMSEA	0.07				
RMSEA LL	0.04				
RMSEA UL	0.10				
<i>p</i> value close fit	.11				
	Overall	Old	Young		
Fit Function	0.45	0.62	0.34		
SRMR	0.07	0.09	0.06		
GFI	0.98	0.97	0.98		
Bentler-Bonett NFI	0.91	0.86	0.93		
Cronbach's alpha	0.84	0.79	0.86		
Items	Factor Loadings	Old Intercept	Young Intercept	R-square, Old	R-square, Young
10. Acts defiant when told to do something	0.76	5.02	4.00	0.33	0.21
12. Gets angry when doesn't get own way	1.17	5.86	5.41	0.65	0.68
13. Has temper tantrums	1.16	5.57	5.20	0.56	0.58
16. Cries easily	1.13	4.69	4.33	0.38	0.47
17. Yells or screams	1.43	5.31	4.62	0.78	0.70
18. Hits parents	1.20	4.18	3.93	0.36	0.42
28. Constantly seeks attention	0.96	5.16	4.30	0.31	0.33
35. Is overactive or restless	0.71	5.00	4.11	0.16	0.15
<b>Outcome</b>					
CBCL Total Cut-off	0.33	0.57	0.32	0.42	0.49
Correlated Errors	Old Estimate	Old <i>P</i> Value	Young Estimate	Young <i>P</i> Value	
E10 E12	0.14	<.0001	0.11	<.0001	
E12 E13	0.24	<.0001	0.24	<.0001	
E12 E17	0.01	.785	0.01	.785	
E28 E35	0.23	<.0001	0.23	<.0001	



**Fig. 1** Visualization of full structural equation model using best fitting measurement model (metric) with old and young children, predicting CBCL cut-off scores. Note. *CBCL* Child Behavior Checklist, *ECBI* Eyberg Child Behavior Inventory

that may be considered developmentally typical but can also be a problem for families. In this study, expert data provided a good framework for determining whether behaviors were problematic, typical, or both.

The specific items in the toddler factor of the ECBI identified in this study support the theory that behavior problems in this age range are related to undeveloped emotion regulation skills (Calkins & Johnson, 1998). Several of the ECBI-Toddler factor items assessed problems with emotion regulation (i.e., “has temper tantrums,” “yells or screams,” “hits parents,” “gets angry when doesn’t get own way,” and “cries easily”). A seminal study of toddler emotion regulation suggests that toddlers’ difficulty with regulating frustration is associated with aggression and disruptive behaviors (Calkins & Johnson, 1998). Expert qualitative data from this study further support Calkins and Johnson’s (1998) findings by emphasizing the importance of emotion regulation development during this stage of toddlerhood and the relation between emotion regulation and behavior problems. Girard et al. (2018) distinguish between “big emotions” (i.e., emotional outbursts that are due to a toddler’s lack of emotion regulation skills) and tantrums (i.e., outbursts related to defiance and disruptive behavior problems). Previous research also indicated that parental behaviors significantly impacted a toddler’s ability to regulate their emotions, suggesting that parents are important models for emotion regulation development (Calkins & Johnson, 1998; Ekas et al., 2011). Overall, emotion regulation is a critical area for assessment and intervention in toddler-aged children with behavior concerns, and the identified 8-item toddler factor of the ECBI in the current study appears to capture behaviors that can be attributed to emotion regulation problems.

Previous research suggests that toddler behavior problems also can be attributed to parent–child interactions and attachment (Diemer et al., 2021; Ekas et al., 2011). Some of the items retained on the toddler-based factor may be capturing these attachment-related or relationship-based concerns (i.e., “hits parent,” “acts defiant,” and “constantly seeks attention”). Diemer and colleagues (2021) found that intrusive

parenting was associated with more behavior problems and lower emotion regulation in toddlers. When considering the eight items retained in the factor, these findings suggest that interventions that target parenting behaviors, such as Parent–Child Interaction Therapy-Toddler, are essential for addressing toddler problem behaviors. As the factor identified in the current study appears to assess behavioral manifestations of attachment problems, this factor could be useful for monitoring treatment that focuses on strengthening parent–child relationships as the mechanism for improving toddler behaviors.

### Strengths, Limitations, and Future Directions

The current study has several strengths to be noted. Data used in this study were gathered prior to treatment beginning and thus, were not impacted by treatment attrition or treatment effects. Additionally, archival data included both the CBCL, a broadband measure of child behavior problems, and the ECBI allowing for comparisons of the new factor of the ECBI to an already well-validated toddler assessment. Another strength was the high response rate of experts in the field on the survey about the ECBI that allowed for the integration of rich qualitative data. The current study’s mixed-method approach allowed for more certainty about the content validity of the new factor, as well as important insights into why certain items are or are not appropriate for toddler-aged children.

There are also several limitations of the current study to note. Despite pre-treatment data collection, there was still a high rate of missingness on certain ECBI items. After discussion with research staff and clinicians and closer examination of the data, it was discovered that participating families tended to skip items that they did not find applicable to their child. The high rate of missing items also prevented the comparison of scores on the full ECBI with scores on the toddler factor. Future research would be needed to assess the added benefits of the toddler factor when compared to the full ECBI for this population.

An additional limitation is that the data used were archival in nature. Thus, the researchers could not obtain certain demographic information, such as race and ethnicity, that were not collected in the original randomized controlled trial. Data for this factor analysis were solely gathered in Australia and may not be generalizable to other countries or populations. Finally, participants in this study did not represent the lower end of the toddler age range as children were only included if they were 14 months or older, and thus this work should be replicated with toddlers ages 12 to 13 months.

An additional limitation of this study is that the EFA and CFA were run on the same samples, which may lead to the models overfitting to the samples used in this study. While we were able to obtain an older additional sample to the younger sample, this does not provide cross-validation with a younger sample. Thus, results might not be generalizable to other study populations. The younger sample included 130 children (older sample included 84 children), which, while sufficiently powered for the study, was too small for training-validation-test split.

To evaluate the generalization of the findings of the current study, future researchers should replicate this investigation using an international sample of children to assess the psychometric properties of the factor across different cultures and language versions. Other possible future directions could include gathering qualitative and survey data from parents of toddler-aged children to determine whether these items fit with caregivers' perceptions of behavior problems for this age range or if additional items should be added to fully capture the construct. After determining whether this factor captures the full range of toddler behavior problems, it would also be important for future research to develop new norms with this adaptation of the ECBI for toddler-aged children to make the measure more meaningful for clinical and research use. Finally, future research should also examine whether this new factor is sensitive to weekly change and whether elevations on items included in the factor are predictive of future behavior problems.

## Conclusion

In conclusion, this factor analysis is a first step in developing a brief assessment tool that can be used throughout treatment to monitor toddler disruptive behaviors. The development of the toddler factor in the ECBI may provide clinicians and researchers with a measure for assessing outcomes in toddler-based interventions that fills the gap left by already validated measures for toddlers, though replications are needed. Additionally, this study provides preliminary evidence for the content validity of this factor through the integration of qualitative data from expert toddler researchers

and clinicians, further supporting its future use in clinical and research settings.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10862-023-10078-0>.

**Data Availability** Please contact the first-author to inquire about the data used in this project.

## Declarations

**Conflict of Interest** Sharon T. Phillips, Lindsay R. Druskin, Christa Lilly, Felipa T. Chavez, Sue Morgan, Nancy Wallace, Sara Cibralic, Eva R. Kimonis, David J. Hawes, Valsamma Eapen, Cheryl B. McNeil, and Jane Kohlhoff declare that they have no conflict of interest.

**Experiment Participants** This project was approved by the Institutional Review Board at West Virginia University.

## References

- Achenbach, T. M. (1999). The Child Behavior Checklist and related instruments. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment* (pp. 429–466). Lawrence Erlbaum Associates Publishers.
- Barkley, R. (1998). *Attention-deficit hyperactivity disorder: A handbook for diagnosis* (2nd ed.). Guilford.
- Bongers, I. L., Koot, H. M., Van der Ende, J., & Verhulst, F. C. (2004). Developmental trajectories of externalizing behaviors in childhood and adolescence. *Child Development*, *75*, 1523–1537. <https://doi.org/10.1111/j.1467-8624.2004.00755.x>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*, 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- Briggs-Gowan, M. J., & Carter, A. S. (2002). *Brief Infant-Toddler Social and Emotional Assessment (BITSEA) manual, version 2.0*. New Haven, CT: Yale University.
- Burns, G. L., Patterson, D. R., Nussbaum, B. R., & Parker, C. M. (1991). Disruptive behaviors in an outpatient pediatric population: Additional standardization data on the Eyberg Child Behavior Inventory. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *3*(2), 202. <https://doi.org/10.1037/1040-3590.3.2.202>
- Calkins, S. D., & Johnson, M. C. (1998). Toddler regulation of distress to frustrating events: Temperamental and maternal correlates. *Infant Behavior and Development*, *21*, 379–395. [https://doi.org/10.1016/S0163-6383\(98\)90015-7](https://doi.org/10.1016/S0163-6383(98)90015-7)
- Campbell, S. B., Shaw, D. S., & Gilliom, M. (2000). Early externalizing behavior problems: Toddlers and preschoolers at risk for later maladjustment. *Development and Psychopathology*, *12*, 467–488. <https://doi.org/10.1017/S0954579400003114>
- Copeland, W. E., Miller-Johnson, S., Keeler, G., Angold, A., & Costello, E. J. (2007). Childhood psychiatric disorders and young adult crime: A prospective, population-based study. *American Journal of Psychiatry*, *164*, 1668–1675. <https://doi.org/10.1176/appi.ajp.2007.06122026>
- De Leeuw, E. D., Hox, J. J., & Huisman, M. (2003). Prevention and treatment of item nonresponse. *Journal of Official Statistics*, *19*, 153–176.
- Diemer, M. C., Treviño, M. S., & Gerstein, E. D. (2021). Contextualizing the role of intrusive parenting in toddler behavior problems and emotion regulation: Is more always worse?

- Developmental Psychology*, 57(8), 1242–1253. <https://doi.org/10.1037/dev0001231>
- Ekas, N. V., Braungart-Rieker, J. M., Lickenbrock, D. M., Zentall, S. R., & Maxwell, S. M. (2011). Toddler emotion regulation with mothers and fathers: Temporal associations between negative affect and behavioral strategies. *Infancy*, 16, 266–294. <https://doi.org/10.1111/j.1532-7078.2010.00042.x>
- Epstein, R. A., Fonnesebeck, C., Potter, S., Rizzone, K. H., & McPheeters, M. (2015). Psychosocial interventions for child disruptive behaviors: A meta-analysis. *Pediatrics*, 136(5), 947–960. <https://doi.org/10.1542/peds.2015-2577>
- Eyberg, S. M., Chase, R. M., Fernandez, M. A., & Nelson, M. M. (2014). *Dyadic Parent-Child Interaction Coding System (DPICS) Clinical Manual (4th Edition)*. PCIT International, Inc.
- Eyberg, S. M., & Pincus, D. (1999). Eyberg Child Behavior Inventory and Sutter-Eyberg Student Behavior Inventory-Revised (ECBI/SESBI-R) professional manual. *Psychological Assessment Resources*.
- Funderburk, & B. W., Eyberg, S. M., Rich, B. A., & Behar, L. (2003). Further psychometric evaluation of the Eyberg and Behar rating scales for parents and teachers of preschoolers. *Early Education and Development*, 14(1), 67–82. [https://doi.org/10.1207/s15566935eed1401\\_5](https://doi.org/10.1207/s15566935eed1401_5)
- Gardner, F., Shaw, D. S., Dishion, T. J., Burton, J., & Supplee, L. (2007). Randomized prevention trial for early conduct problems: Effects on proactive parenting and links to toddler disruptive behaviors. *Journal of Family Psychology*, 2(3), 398–406. <https://doi.org/10.1037/0893-3200.21.3.398>
- Girard, E. I., Wallace, N. M., Kohlhoff, J. R., Morgan, S. S. J., & McNeil, C. B. (2018). *Parent-child interaction therapy with toddlers: Improving attachment and emotion regulation*. Springer Nature Switzerland AG.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire; A research note. *Journal of Child Psychology and Psychiatry*, 38, 581–586. <https://doi.org/10.1111/j.1469-7610.1997.tb01545.x>
- Gridley, N., Blower, S., Dunn, A., Bywater, T., & Bryant, M. (2019). Psychometric properties of child (0-5 years) outcome measures as used in randomized controlled trials of parent programs: A systematic review. *Clinical Child and Family Psychology Review*, 22, 388–405. <https://doi.org/10.1007/s10567-019-00277-1>
- Holtz, C. A., Barber, B. N., & Jarrett, K. M. (2008). Toddler and preschool behavior scale: Preliminary psychometric evaluation of a new scale. *Graduate Journal of Counseling Psychology*, 1(1). <http://epublications.marquette.edu/gjcp/vol1/iss1/3>
- Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288. <https://doi.org/10.1177/1049732305276687>
- Huissman, M. (1999). *Item nonresponse: Occurrence, causes, and imputation of missing answers to test items*. DSWO Press.
- Kokko, K., & Pulkkinen, L. (2000). Aggression in childhood and long-term unemployment in adulthood: A cycle of maladaptation and some protective factors. *Developmental Psychology*, 36(4), 463–472. <https://doi.org/10.1037/0012-1649.36.4.463>
- Kohlhoff, J., Cibralic, S., Wallace, N., Morgan, S., McMahon, C., Hawkins, E., Eapen, V., Briggs, N., & McNeil, C. (2020a). A randomized controlled trial comparing parent-child interaction therapy-toddler, circle of security-parenting and waitlist controls in the treatment of disruptive behaviors for children aged 14–24 months: Study protocol. *BMC Psychology*, 8, 93. <https://doi.org/10.1186/s40359-020-00457-7>
- Kohlhoff, J., Morgan, S., Briggs, N., Egan, R., & Niec, L. (2020b). Parent-child interaction therapy with toddlers in a community-based setting: Improvements in emotional availability, child behaviour and attachment. *Infant Mental Health Journal*, 41(4), 543–562. <https://doi.org/10.1002/imhj.21864>
- Kohlhoff, J., Morgan, S., Briggs, N., Egan, R., & Niec, L. (2021). Parent-child interaction therapy with toddlers: A community-based randomized controlled trial with children aged 14–24 months. *Journal of Clinical Child and Adolescent Psychology*. <https://doi.org/10.1080/15374416.2020.1723599>
- LeBuffe, P. A., & Naglieri, J. A. (2003). *The Devereux Early Childhood Assessment Clinical Form (DECA-C): A measure of behaviors related to risk and resilience in preschool children*. Kaplan Press.
- LeBuffe, P. A., & Naglieri, J. A. (2009). The Devereux Early Childhood Assessment (DECA): A measure of within-child protective factors in preschool children. *NHSA Dialog*, 3(1), 75–80. [https://doi.org/10.1207/s19309325nhsa0301\\_10](https://doi.org/10.1207/s19309325nhsa0301_10)
- Lilly, C. L. (2022). A game of guess who? Testing measurement models between different groups. *The Journal of Pediatrics*, 250, 113–115.
- Machado, E. (2020). Psychometric evaluation of the Eyberg child behavior inventory in an ethnically diverse sample. [Doctoral dissertation, Nova Southeastern University] ProQuest Dissertations and Theses Global.
- Mackrain, M., LeBuffe, P., & Powell, G. (2007). *Devereux Early Childhood Assessment infants and toddlers*. Lewisville, NC: Kaplan Early Learning Company.
- Marvin, R. S., Cooper, G., Hoffman, K., & Powell, B. (2002). The circle of security project: Attachment-based intervention with caregiver-pre-school child dyads. *Attachment & Human Development*, 4, 107–124. <https://doi.org/10.1080/14616730252982491>
- McClendon, D. T., Warren, J. S., Green, K. M., Burlingame, G. M., Eggert, D. L., & McClendon, R. J. (2010). Sensitivity to change of youth treatment outcome measures: A comparison of the CBCL, BASC-2, and Y-OQ. *Journal of Clinical Psychology*, 67(1), 111–125. <https://doi.org/10.1002/jclp.20746>
- Milledge, S. V., Cortese, S., Thompson, M., McEwan, F., Rolt, M., Meyer, B., Sonuga-Barke, E., & Eisenbarth, H. (2019). Peer relationships and prosocial behaviour differences across disruptive behaviours. *European Child & Adolescent Psychiatry*, 28, 781–793. <https://doi.org/10.1007/s00787-018-1249-2>
- Morawska, A., & Sanders, M. R. (2006). Self-administered behavioural family intervention for parents of toddlers: Effectiveness and dissemination. *Behavior Research and Therapy*, 44(12), 1839–1848. <https://doi.org/10.1016/j.brat.2005.11.015>
- O'Rourke, N., & Hatcher, L. (2013). *A step by step approach to using SAS for factor analysis and structural equation modeling* (2nd ed.). SAS Institute Inc.
- Patterson, G. R. (1982). *Coercive family process*. Eugene, OR: Castalia.
- Reef, J., Diamantopoulou, S., van Meurs, I., Verhulst, F. C., & van der Ende, J. (2011). Developmental trajectories of child to adolescent externalizing behavior and adult DSM-IV disorder: Result of a 24-year longitudinal study. *Social Psychiatry and Psychiatric Epidemiology*, 46, 1233–1241. <https://doi.org/10.1007/s00127-10-0297-9>
- Sanders, M. R. (2012). Development, evaluation, and multinational dissemination of the Triple P-Positive Parenting Program. *Annual Review of Clinical Psychology*, 8, 1–35. <https://doi.org/10.1146/annurev-clinpsy-032511-143104>
- Smith, J. D., Dishion, T. J., Shaw, D. S., Wilson, M. N., Winter, C. C., & Patterson, G. R. (2012). Coercive family process and early-onset conduct problems from age 2 to school-entry. *Development and Psychopathology*, 26, 917–932. <https://doi.org/10.1017/S0954579414000169>
- Stormshak, E. A., Bierman, K. L., & The Conduct Problems Prevention Research Group. (1998). The implications of different developmental patterns of disruptive behavior problems for school adjustment. *Development and Psychopathology*, 10, 451–467. <https://doi.org/10.1017/S0954579498001692>
- Van Emmerik-van Oortmerssen, K., van de Glind, G., van den Brink, W., Smit, F., Crunelle, C. L., Swets, M., & Schoevers, R. A. (2012). Prevalence of attention-deficit hyperactivity disorder in substance use disorder patients: A meta-analysis and

- meta-regression analysis. *Dependence*, 122(1), 11–19. <https://doi.org/10.1016/j.drugalcdep.2011.12.007>
- Wilks, T., Gerber, R. J., & Erdie-Lalena, C. (2010). Developmental milestones: Cognitive development. *Pediatrics in Review*, 31(9), 364–367. <https://doi.org/10.1542/pir.31-9-364>
- Weis, R., Lovejoy, M. C., & Lundahl, B. W. (2005). Factor structure and discriminative validity of the Eyberg Child behavior Inventory with young children. *Journal of Psychopathology and Behavioral Assessment*, 27(4), 269–278. <https://doi.org/10.1007/s10862-005-2407-7>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.