



Reliability and Validity of Behavior Observation Coding Systems in Child Maltreatment Risk Evaluation: A Systematic Review

Jelena Zumbach¹ · Anna Oster¹ · Annika Rademacher² · Ute Koglin²

Accepted: 25 October 2021 / Published online: 12 November 2021
© The Author(s) 2021

Abstract

Performing child maltreatment risk assessments is a challenging task that calls for valid and reliable measures. In child protection proceedings, mental health professionals conduct maltreatment assessments that often form an important basis for judicial decision making. Because parent–child interaction is a key construct in maltreatment risk evaluations, observational assessment measures are crucial. This systematic review aims to identify observational coding systems of parent–child interaction that are applicable for psychological evaluations of the risk of child maltreatment. The goal is to examine the potential of observational coding systems to discriminate behavior of parents who have versus have not engaged in child maltreatment. A systematic literature search led to the inclusion of 13 studies published in the United States and Europe that were then analyzed in detail. Across the 13 studies, this review identified 11 unique observational coding systems. Results are summarized systematically for study characteristics and outcomes. Additionally, the main characteristics of the observational coding systems are identified and analyzed, including the age range of the child, observation tasks, measured constructs, and reliability. The discussion focuses on the strengths and weaknesses of the individual observational coding systems in the context of child maltreatment risk assessments.

Keywords Child maltreatment · Parent–child interaction · Observational coding systems · Child protection proceedings · Systematic review

Highlights

- Behavioral observation that focuses systematically on specific behavioral dimensions may be a valid approach to assess the risk of child maltreatment.
- The largest body of evidence supports the conclusion that significantly lower levels of “parental sensitivity and responsiveness,” “developmentally appropriate behavior,” and “positive affect,” as well as significantly higher levels of “hostility and control” and “parental anger” differentiate parents who have from those who have not engaged in child maltreatment.
- The selection of an observational coding system within a child abuse risk assessment should take the position and value of parent–child observations into account within the entire risk evaluation.

The World Health Organization (WHO, 2016) defines child maltreatment as “all types of physical and/or emotional ill-treatment, sexual abuse, neglect or negligent treatment or commercial or other exploitation, resulting in actual or potential harm to the child’s health, survival, development or dignity in the context of a relationship of responsibility, trust or power.” Child maltreatment is a major social issue with an alarmingly high prevalence all over the world (see Finkelhor, 1994; Gilbert et al., 2009; Smith Slep et al., 2015). Because child maltreatment has serious long-term consequences for child development, scientific and practical

✉ Jelena Zumbach
j.zumbach@phb.de

¹ Family Law Psychology, Psychologische Hochschule Berlin, Am Koellnischen Park 2, 10179 Berlin, Germany

² Faculty of Educational and Social Sciences, Department of Pedagogic and Rehabilitation Psychology, University of Oldenburg, Ammerlaender Heerstr. 114–118, 26111 Oldenburg, Germany

efforts are needed to improve the selection, development, and use of methods to assess and prevent its risk (Cicchetti et al., 2006; Cyr & Alink, 2017; Toth & Cicchetti, 2013).

In child protection proceedings resulting from child abuse or neglect, family courts frequently commission mental health professionals to conduct child maltreatment risk evaluations. Mental health experts have a duty to “provide relevant, professionally sound results or opinions in matters where a child’s health and welfare may have been and/or may be harmed” (American Psychological Association, 2013, p. 22). Alongside the child’s well-being and psychological needs as well as wider family environmental factors, one of the crucial factors assessed by mental health professionals when concerns arise regarding a child’s endangerment is parenting capacity (see American Psychological Association, 2013; Zumbach & Koglin, 2015).

Parenting capacity is defined rather broadly as a construct referring to the ability of parents to meet the basic needs of a given child. Specifically, the parent’s skills and behaviors are assessed to ascertain how far that parent is capable of meeting the child’s needs. Because parenting capacity is based on fundamental parental abilities, which can manifest themselves in the interaction with a specific child, parenting behavior serves as an important indicator for parenting capacity (American Psychological Association, 2013; Zumbach & Oster, 2021).

Assessing the risk of future child abuse and neglect is a demanding and multidimensional challenge that indispensably requires an approach that includes multiple methods of assessment. Alongside file analyses, interviews with parents and children, psychometric testing, and analyses of third-party reports, one of the core methods of assessment is the behavioral observation of parent–child interaction (American Psychological Association, 2013).

Critical components when assessing parenting capacity are parental responsiveness, parent–child interaction problems, and the overall parent–child relationship (Zumbach & Oster, 2021). From a diagnostic perspective, this assessment imposes several challenges. For example, there is limited evidence on the validity of self-reports on these aspects within an evaluation context. Given that a parental self-report may be biased by social desirability and that especially young children may be limited in their receptive and productive language competencies, self-report measures and investigative child interviews may not be sufficiently informative.

Therefore, observations of parent–child interactions in natural or structured settings are a crucial method with which to augment self-report assessments when evaluating such constructs as parenting skills, parent–child interaction problems, attachment, and parent–child relationships (Bennett et al., 2006; Harnett, 2007). One common assumption is that stable patterns in a relationship are

reflected in parent–child interactions. However, it is important to note that behavioral observation in this context is not expected to directly reveal child abuse such as physical or sexual abuse. It is unlikely that most parents will engage in frank abuse during a rather short observation or that they will display gross neglect of physical needs. Rather, the behavioral observation aims to provide data on potentially limited parenting skills, parent–child interaction problems, or a maladaptive parent–child relationship. These are commonly discussed in the literature as being among the strongest *risk factors* for predicting the future (re-)abuse of a child (Assink et al., 2019; Mulder et al., 2018; Sledjeski et al., 2008; Stith et al., 2009; Wilson et al., 2008).

In the given context, mental health professionals therefore focus on observing parent–child interactions in order to infer the *risk* of abuse and/or neglect from evidence that the parent’s behavior is adequate to meet the child’s needs. Such inadequacy may involve the presence of problematic parental behavior and/or the absence of needed parental behavior. Simultaneously, adaptive parenting behavior must be observed so as to perform an unbiased assessment (Zumbach & Oster, 2021).

This approach is part of a larger risk assessment that includes further factors such as parent-related risk factors (e.g., parent’s own victimization through child abuse, mental disorders, substance use problems, negative attitudes toward intervention or treatment) or family factors (e.g., socioeconomic stressors, lack of social support, intimate partner violence; see de Ruiter et al., 2020). When assessing these risk factors, other assessment techniques as mentioned above (e.g., interviews with parents, psychological testing, analyses of third-party reports) might play a more prominent role than behavioral observation. In the individual case, any final conclusion on the risk for future child abuse can be reached only after integrating and weighting the information obtained within each single step of the risk assessment (see American Psychological Association, 2013).

Given the wide range of parental behavior that may indicate child maltreatment risk, decisions about which observational measure to use for child maltreatment risk assessments are challenging (Bennett et al., 2006; Cerezo, 1997). Up to the present day, a number of observational coding systems have been developed to examine relevant constructs such as attachment, the parent–child relationship, or the interaction with a parent or caregiver. However, choosing a reliable and valid observational coding system is crucial. Even though a high number of observational coding systems exist to assess parent–child interaction behavior, not all of these may be appropriate to assess child maltreatment risk (Budd, 2001; Budd & Holdsworth, 1996).

A number of theoretical guidelines and empirical studies indicate that observational measures of parent–child interaction are frequently used in evaluation practice, and that

evaluators rate as an important technique. However, many guidelines and empirical studies on evaluation practice fail to specify whether to conduct an unstructured versus a structured observation, and they hardly ever name specific coding systems with which to analyze the observed interaction (Barber & Delfabbro, 2000; Budd, 2001; Garber, 2016; Kähkönen, 1999; Wilson et al., 2008; Ziegenhain et al., 2007; Zumbach & Koglin, 2015). Nonetheless, the literature does show that the systematic nature of behavioral observation has a large impact on its reliability and validity. This speaks for the application of structured and semi-structured observational coding systems in the given context (see Haynes & O'Brien, 2000).

The overall aim of this systematic review is to provide professionals performing child abuse/neglect risk evaluations with information about the utility of different observational coding systems for parent–child interactions. The goal is to examine the potential of observational coding systems to discriminate the behavior of parents who have versus have not engaged in child maltreatment. Therefore, we systematically identified studies that analyze observational coding systems and examined their potential to detect differences in the behavior of parents or to discriminate parents who have engaged in child maltreatment from those who have not.

Method

We conducted a systematic literature search in bibliographic databases (Web of Science and PsycINFO) between April and June 2019 to find relevant studies published in English. Additionally, we checked literature references of studies meeting our inclusion criteria as well as reference lists from literature reviews and meta-analyses. The methods used in this systematic review follow the PRISMA guidelines (Moher et al. [PRISMA Group], 2009). For a transparent, structured literature management with a systematic screening and exclusion of the identified data, we used EPPI reviewer software (EPPI-Reviewer 4; Thomas et al., 2010).

Search Strategy and Inclusion Criteria

Based on our preliminary theoretical research, we applied an open search term strategy using the following keywords:

[parent–child relation OR parenting] AND [scales OR coding OR assessment OR measures OR observ* OR interac*] AND [maltreatment OR abuse OR neglect]

Studies had to meet the following inclusion criteria:

1. Observational measures on parent–child interaction were used to (a) detect differences in parenting behavior in parents who did versus did not engage in child maltreatment, or to (b) classify parents as

engaging in or not engaging in child maltreatment.

2. Maltreatment had to be defined by the individual studies as parents or children being documented as having a substantiated history of maltreatment (e.g., by child protection services) and/or being at risk at or having lost custody of their child.
3. Detailed information on the specific parent–child observational coding system had to be provided. The observational coding system for parent–child interaction was defined as an observational measure of either parental behavior or parental and child behavior in procedural interactions.
4. The observational coding had to be conducted by a mental health professional and/or a trained researcher and not a layperson.
5. The study had to be published in the English language in journals applying a peer-review process. English-language dissertations were included if they were listed in bibliographic databases.

This review excluded reports that present the results of literature reviews and reports about coding systems of parent–child interactions that did not involve observational methods (e.g., self-report checklists). The flow chart in Fig. 1 displays the study selection process for the literature search.

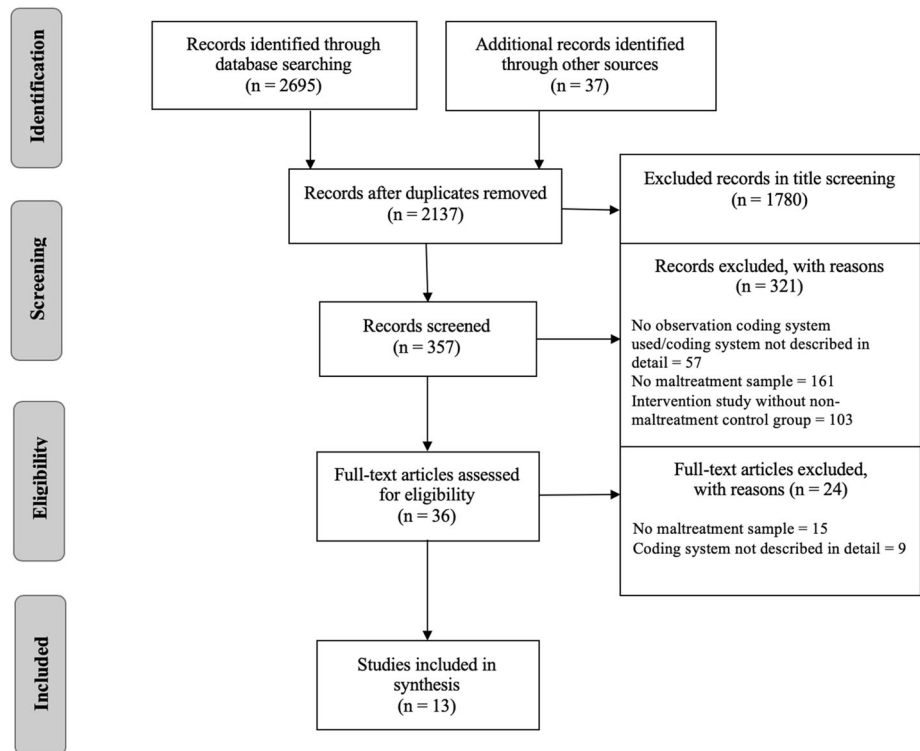
Data Extraction

Study findings were synthesized in a descriptive approach. We selected the following parameters to analyze our database: aim of the study, observational coding system, observation task, sample, and outcome. If reported in the original study, means and standard deviations were extracted to calculate effect sizes (Cohen's *d*) for outcomes of studies analyzing differences in parenting behavior in parents who had engaged versus not engaged in child maltreatment. We also analyzed information given on the sensitivity and specificity of study outcomes in terms of the classification of parents as engaging versus not engaging in child maltreatment.

Results

After removing duplicates, we could identify 2137 potential studies in our literature search. A large number were excluded after title and abstract screening, because many studies did not meet content criteria and sample characteristics or did not include an empirical analysis. Overall, 13 studies corresponded to the inclusion criteria. Although the literature search was conducted for a broad range of applicable studies published up to 2019, the 13 studies that met the inclusion

Fig. 1 Flow chart of the study selection process



criteria were all published between 1993 and 2013. Twelve of the included studies were conducted in the United States, and one was conducted in Spain. Publications were drawn from bibliographic databases. If full-text publications were not available in databases, authors were contacted via email.

Methodological Aspects of Included Studies

Main study characteristics and outcomes are displayed in detail in Tables 1 and 2. Eleven out of the 13 studies aimed to detect behavioral differences between parents who had versus had not engaged in child maltreatment, whereas two out of 13 studies aimed to classify parents as engaging versus not engaging in child maltreatment.

Overall, the individual studies applied 11 different observational coding systems. The Structural Analysis of Social Behavior (SASB; Benjamin, 1974) and the Qualitative Ratings of Parent–Child Interactions (Cox, 1997) were applied in multiple studies. The following coding systems were applied in one study each: the Child–Adult Relationship Experimental Index (CARE-Index, Crittenden 2004; Crittenden, 2006), the Maternal Behavior Q-Sort (MBQS; Pederson et al., 1999), the Parent/Caregiver Involvement Scale (P/CIS; Farran et al., 1986), the Parent–Child Interaction Procedure (Heller et al., 1998), the Psychological Maltreatment Rating Scales (PMRS; Brassard et al., 1993), the Standardized Observation Codes (SOC-III; Cerezo et al., 1986), the UCLA Parent–Child Coding System

(Valeri et al., 2001), the Coder Impressions Inventory (CII; Capaldi & Patterson, 1989), and the Dyadic Parent–Child Interaction Coding System-II (DPICS-II; Eyberg et al., 1994).

Trained research assistants blind to the maltreatment status conducted the majority of the coding across studies. Most studies reported interrater reliability coefficients ranging from satisfactory to high.

Sample Characteristics of Individual Studies

The mean age of the children within the samples ranged from 13.31 months to 8.04 years. In 6 of the 13 studies, the majority of the sample was Caucasian, whereas in 6 further studies, the sample consisted mainly of persons belonging to ethnic minorities. One study did not report ethnic characteristics of the sample (Cerezo et al., 1996).

The majority of parents included in the original samples were mothers with five studies involving other types of caregivers such as fathers, stepfathers, or grandparents. Only eight of the studies (61.5%) had sample sizes larger than 100 caregiver–child dyads. The largest sample was analyzed by Hurlburt et al. (2013) with 481 dyads.

In most studies, maltreatment was classified as parent–child dyads having a substantiated history of maltreatment recorded in child protection service reports (Brassard et al., 1993; Cerezo et al., 1996; Cicchetti et al., 2006; Cipriano-Essel et al., 2013; Fagan & Dore, 1993;

Table 1 Included studies detecting differences in behavior of parents who engage versus do not in child maltreatment ($N = 11$)

Study	Aim	Observational coding system	Observation task	Coder information	Sample	Significant group differences in parenting behavior of maltreating versus nonmaltreating parents	Effect size (Cohen's d)
Cicchetti et al. (2006) US	Detect differences in the efficacy of three different preventive interventions in mother-child dyads that do versus do not engage in maltreatment	MBQS	3-hr home observation	Two trained coders Intraclass correlations between pairs of observers 0.71 to 0.72	$N = 189$ mother-child dyads, $n = 137$ dyads with substantiated history of abuse Child age in months $M = 13.31$ ($SD = 0.81$); Est. range: 10.88–15.74 months 74.1% ethnic minority	Significantly lower sensitivity scores for mothers engaging in maltreatment ($M = 0.27$, $SD = 0.36$) than mothers not engaging in maltreatment ($M = 0.48$, $SD = 0.29$).	$d = -0.39$
Cipriano-Essel et al., (2013) US	Detect differences in parent and child behavior by comparing mother-child dyads who do versus do not engage in maltreatment	SASB	3–5 min moderately challenging joint teaching task	Trained coders (75 h), blind to maltreatment status Weighted kappa 0.53 to 1.0 ($M = 0.74$)	$N = 118$ mother-child dyads, $n = 64$ dyads with substantiated history of abuse Child age in years: $M = 3.7$ ($SD = 0.74$); Range: 3–5 years 79.8% Caucasian, 2.5% African American, 16.2% multiracial	Mothers engaging in maltreatment ($M = 0.31$, $SD = 0.21$) show significantly more strict/hostile control than mothers not engaging in maltreatment ($M = 0.23$, $SD = 0.21$).	$d = 0.38$
Fagan & Dore (1993) US	Examining differences in mother-child play between mother-child dyads who do versus do not engage in neglect	P/CIS	30-min free play session at home	Trained research assistants and experienced child welfare professionals, blind to the maltreatment status and purpose of the study	$N = 27$ mother-child dyads; $n = 17$ neglected children Child age in months $M = 27.5$ (SD and range not reported) 62% African American, 38% Caucasian	Significantly lower maternal positive responsiveness for mothers engaging in neglect ($M = 16.00$) than for mothers not engaging in neglect ($M = 20.54$).	n. a. (SD not reported)
Haskett et al. (2008) US	Examining protective and vulnerability factors for physically abused children	Qualitative ratings of parent-child interactions	Three different tasks, 30 min	Trained coders, uninformed of maltreatment status $\kappa = 0.80$	$N = 153$ parent-child dyads; $n = 78$ dyads with substantiated history of abuse Child age in years: $M = 7.2$ ($SD = 1.4$); Range: 3.0–11.4 69% African American, 31% Caucasian	Significantly lower maternal development appropriateness for mothers engaging in neglect ($M = 37.29$) than for mothers not engaging in neglect ($M = 53.89$).	n. a. (SD not reported)
Hurlburt et al. (2013) US	Examining efficacy of an intervention program in families with self-reported history of child maltreatment	DPICS-II & CII	30-min home observation	Highly experienced or extensively trained coders Intraclass correlations 0.70 to 0.87	$N = 481$ parent-child dyads; $n = 100$ dyads with self-reported history of abuse Child age in years: $M = 4.7$ ($SD = 0.36$); Est. range: 3.62–5.78 years 60.3% Caucasian, 8.4% African American, 4.6% Hispanic/Latino, 1.0% Asian American, 1.5% Native American, 14.2% multiracial	Significantly lower scores for parental sensitivity in parents engaging in maltreatment ($M = 3.57$, $SD = 1.3$) than in parents not engaging in maltreatment ($M = 3.98$, $SD = 1.3$).	$d = -0.36$
						DPICS-II: Significantly more total critical statements in parents engaging in maltreatment ($M = 25.48$, $SD = 19.30$) than in parents not engaging in maltreatment ($M = 19.07$, $SD = 17.36$).	$d = 0.36$
						CII: Significantly less nurturing/supportive behavior in parents engaging in maltreatment ($M = 2.29$, $SD = 0.53$) versus parents not engaging in maltreatment ($M = 2.47$, $SD = 0.45$).	$d = -0.39$
						CII: Significantly more harsh/critical behavior in parents engaging in maltreatment ($M = 1.60$, $SD = 0.58$) versus parents not engaging in maltreatment ($M = 1.42$, $SD = 0.46$).	$d = 0.37$

Table 1 (continued)

Study	Aim	Observational coding system	Observation task	Coder information	Sample	Significant group differences in parenting behavior of maltreating versus nonmaltreating parents	Effect size (Cohen's <i>d</i>)
Sabourin Ward & Haskett (2008) US	Exploration and validation of clusters of children who suffered from physical abuse	Qualitative ratings of parent-child interactions	Three different tasks, 30 min	Trained, uninformed of maltreatment status $\kappa = 0.80$ for reliability between raters	$N = 175$ parent-child dyads; $n = 98$ dyads with substantiated history of abuse Child age in years: $M = 7.33$ ($SD = 1.54$); Range: 5–10 years 70% African American, 27% European American, 3% Hispanic or biracial	Parents engaging in maltreatment ($M = 6.93$, $SD = 2.6$) showed significantly less sensitivity than parents not engaging in maltreatment ($M = 8.09$, $SD = 2.6$).	$d = -0.45$
Skowron et al. (2011) US	Investigate mother and child physiological regulation (RSA) and their associations to observed parent-child interaction	SASB	Mother-child joint task	Trained coders (75 hrs. of training), blind to family status κ from 0.64 to 0.84 ($M = 0.74$) for reliability between raters	$N = 76$ children and their mothers; $n = 17$ suffered from physical abuse; $n = 26$ suffered from physical neglect Child age in years: $M = 3.7$ ($SD = 0.73$); Range: 3–5 years 77.3% Caucasian, 18.7% multiracial, 4.0% African American	Mothers engaging in physical abuse ($M = 0.24$, $SD = 0.15$) were significantly more likely to react with strict control than mothers who did not engage in maltreatment ($M = 0.13$, $SD = 0.13$). Mothers engaging in neglect ($M = 0.23$, $SD = 0.16$) were significantly more likely to react with strict control than mothers who did not engage in maltreatment ($M = 0.13$, $SD = 0.13$).	$d = 0.80$
Skowron et al. (2013) US	Examine the relationship between concurrent and lagged physiological regulation (RSA) and parenting behavior of maltreating and nonmaltreating mothers	SASB	Mother-child teaching task	Trained coders (75 h of training), blind to maltreatment status κ from 0.53 to 1.0 ($M = 0.74$) for reliability between raters	$N = 141$ mothers and their children; $n = 62$ maltreating/neglecting mothers Child age in years: $M = 3.7$ ($SD = 0.72$); Range: 3–5 years 94.2% Caucasian	Mothers engaging in physical abuse ($M = 0.12$, $SD = 0.09$) were significantly less likely to affirm autonomy to their children than mothers who did not engage in maltreatment ($M = 0.22$, $SD = 0.13$). Mothers engaging in abuse ($M = 32.98$) and mothers engaging in neglect ($M = 35.28$) showed less positive parenting than mothers who did not engage in maltreatment ($M = 42.52$). Mothers engaging in abuse ($M = 0.34$) and mothers engaging in neglect ($M = 0.32$) mothers showed more strict/hostile control than mothers who did not engage in maltreatment ($M = 0.20$).	$d = -0.85$ n. a. (SD not reported)

Notes. n. a. = not applicable; est. range = estimated age range (± 3 standard deviations, calculated if age range was not reported in original study); CARE-Index = Child-Adult Relationship Experimental Index; CII = Coder Impressions Inventory; CM = Child Maltreatment; DPICS-II = Dyadic Parent-Child Interaction Coding System-II; MBQS = Maternal Behavior Q-Sort; P/CIS = Parent/Caregiver Involvement Scale; RSA = respiratory sinus arrhythmia; SASB = Structural Analysis of Social Behavior

Table 2 Included studies classifying parents in those who engage in child maltreatment versus those who do not ($N = 2$)

Study	Aim	Observational coding system	Observation task	Coder information	Sample	Classification of parents as maltreating versus nonmaltreating	Sensitivity and specificity of classification
Brassard et al. (1993) US	Evaluate the factor structure and validity of the Psychological Maltreatment Rating Scales (PMRS)	PMRS	15-min teaching task	Trained research assistants	$N = 49$ mother–child dyads; $n = 25$ children who suffered from maltreatment Child age in years: $M = 6.67$ (SD not reported); Range: 4–8.92 years 46% Caucasian, 41% African American, 13% Hispanic	Coding system correctly classified 81.63% of the mothers as engaging in psychological maltreatment versus not engaging in psychological maltreatment	Sensitivity = 0.92 Specificity = 0.71
Cerezo et al. (1996) Spain	Detecting the differentiating aspects of abusive and nonabusive mothers' behavior during parent–child interaction	SOC-III	Several 1-hr unstructured interactions in home	Trained undergraduate psychology students (average 20 hr of training)	$N = 47$ mother–child dyads; $n = 23$ children who suffered from maltreatment Sample with maltreatment history: Child age in years: $M = 8.04$ ($SD = 2.78$); Est. range: 0–16.38 years Sample without maltreatment history: Child mean age in years: $M = 8.0$ ($SD = 2.16$); Est. range: 1.52–14.48 years Ethnicity: n. a.	Maternal behavior scale correctly identified 70.21% of the mothers as engaging in maltreatment versus not engaging in maltreatment Child behavior scale correctly identified 74.47% of the mothers as engaging in maltreatment versus not engaging in maltreatment Mother–child interactional behavior scale correctly identified 78.72% of the mothers as engaging in maltreatment versus not engaging in maltreatment	Sensitivity = 0.79 Specificity = 0.61 Sensitivity = 0.88 Specificity = 0.61 Sensitivity = 0.88 Specificity = 0.61
						The whole model correctly identified 82.98% of the mothers as engaging in maltreatment versus not engaging in maltreatment	Sensitivity = 0.88 Specificity = 0.78

Notes. n. a. = not applicable; est. range = estimated age range (± 3 standard deviations, calculated if age range was not reported in original study); PMRS = Psychological Maltreatment Rating Scales; SOC-III = Standardized observation codes

Haskett et al., 2008; Lau et al., 2006; Robinson et al., 2009; Sabourin Ward & Haskett, 2008; Skowron et al., 2011; Skowron et al., 2013). Ostler (2010) classified maltreatment as mothers having lost custody during child protection proceedings aiming to measure the risk for reoccurrence of child maltreatment. Hurlburt et al. (2013) assessed a child maltreatment history via parent's self-report.

Differences in Behavior of Parents Who Did Versus Did Not Engage in Child Maltreatment

Eleven of the 13 studies analyzed differences in behavior of parents who did versus did not engage in child maltreatment. Across studies, the effect sizes (Cohen's d) calculated for outcomes of studies analyzing differences in parenting behavior in parents who did versus did not engage in child maltreatment indicated a majority of small to medium effects ranging from $d = 0.26$ to $d = 0.45$ (see Cicchetti et al., 2006; Cipriano-Essel et al., 2013; Haskett et al., 2008; Hurlburt et al., 2013; Sabourin-Ward & Haskett, 2008). The results of three studies indicated large effect sizes of up to $d = 1.55$ (see Ostler, 2010; Robinson et al., 2009; Skowron et al., 2011). The direction of these effects was in line with expectations throughout the studies. We shall now describe the studies' findings on the ability of the individual coding systems to detect such differences.

Cicchetti et al. (2006) aimed to evaluate the efficacy of different preventive interventions in mothers who had versus had not engaged in child maltreatment. Conducting a 3-hr home observation during the preintervention assessment, they coded maternal sensitivity with the MBQS. Observations revealed significantly lower scores for maternal sensitivity in the maltreatment group than in the nonmaltreatment group with a medium effect size ($d = -0.39$).

Cipriano-Essel et al. (2013) applied the SASB to detect differences in parent and child behavior by comparing mothers who had versus had not engaged in child maltreatment based on the observation of a 3- to 5-min teaching task. Results indicated significantly more strict/hostile control for mothers who had engaged in maltreatment than for mothers who had not engaged in maltreatment ($d = 0.38$). No significant group differences were found for maternal warm autonomy support and warm guidance.

Fagan and Dore (1993) examined differences in mother-child play by applying the P/CIS to compare mothers who had versus had not engaged in neglect when playing together with their children. Observations were conducted during a 30-min free play session at home. Results revealed significant group differences indicating that mothers who had engaged in neglect showed less positive responsiveness and less developmental appropriateness toward their children than mothers who had not

engaged in neglect. No significant group difference was found for maternal positive control.

Haskett et al. (2008) applied the Qualitative Ratings of Parent-Child Interactions in a sample of parent-child dyads that had versus had not engaged in/been affected by child maltreatment. Observations took 30 min and were conducted during three different tasks. Results showed significantly lower scores for parental sensitivity in the maltreatment group than in the nonmaltreatment group ($d = -0.36$).

Hurlburt et al. (2013) examined the efficacy of an intervention program comparing parent-child dyads with versus without a self-reported history of maltreatment. In the pretest assessment, the DPICS-II was applied during a 30-min home observation and the CII was used to gain the general coder impression of the parent-child interaction. Results on the DPICS-II revealed significantly more critical statements by parents who had engaged in maltreatment than by parents who had not engaged in maltreatment ($d = 0.36$). No significant group differences were found for the subscales "praise", "positive affect", "physical positive", and "total commands". The CII revealed significant group differences in all scales, indicating that parents who had engaged in maltreatment showed less nurturing/supportive behavior ($d = -0.39$), more harsh/critical behavior ($d = 0.37$), and less discipline competence than parents who had not engaged in maltreatment ($d = -0.26$).

Lau et al. (2006) compared abusive and nonabusive parent-child dyads in order to examine associations between parental reports of child behavior problems and observed parent-child interactions in parent-child dyads that had engaged in/been affected by child maltreatment. Observation took place for 13 min overall and included three short interaction tasks. After controlling for family income, age of child, sex, and parental psychopathology, abuse status related significantly to observed parent behavior: Parents who had engaged in abuse showed significantly more emotionally controlling behavior such as criticism, guilt induction, and intrusiveness; and they displayed less supportive behavior such as praise, encouragement, and displays of affection (because standard deviations were not reported in the original study, effect sizes were not calculated).

Ostler (2010) investigated a specific sample of mothers suffering from mental illness and reported that the CARE-Index scores correlated significantly with a risk of reoccurring maltreatment. Observations were based on a 3-min free play session. Mothers at high risk for reoccurring maltreatment showed significantly less sensitivity than mothers at low risk ($d = -1.55$), and mothers at moderate risk ($d = -0.57$). CARE-Index scores correlated significantly with other predictive variables indicating the

reoccurrence of child maltreatment such as the caregiving attitude and insight into mental illness.

Robinson et al. (2009) aimed to detect relations between emotion regulation, parenting, and psychopathology by comparing parent–child dyads that had versus had not engaged in/been affected by child maltreatment. They applied the Parent–Child Interaction Procedure that includes seven joint tasks varying in stress level and difficulty. Results revealed large effects, with significantly lower scores for parental positive affect intensity ($d = -1.11$) and significantly higher parental anger intensity ($d = 0.81$) for parents who had engaged in versus not engaged in child maltreatment.

Sabourin-Ward and Haskett (2008) aimed to explore and validate clusters of children who suffered from physical abuse by comparing parent–child dyads that had versus had not engaged in/been affected by physical abuse while they engaged in three different joint tasks. Observations were coded with the Qualitative Ratings of Parent–Child Interactions. They found significantly lower scores on parental sensitivity for parents who had engaged in versus not engaged in physical abuse ($d = -0.45$).

Skowron et al. (2011) investigated differences in mother’s and child’s physiological regulation (respiratory sinus arrhythmia [RSA]) and their association with observed parent–child interaction. Observations were conducted during a mother–child joint teaching task in samples of mother–child dyads who had engaged in/been affected by neglect, engaged in/been affected by physical abuse, and had not engaged in child maltreatment. Results showed moderate to large effects, indicating that mothers who had engaged in physical abuse ($d = 0.80$) and neglect ($d = 0.70$) were more likely to react with strict control to children’s positive bids for autonomy than mothers who had not engaged in child maltreatment. Mothers who had engaged in physical abuse were significantly less likely to affirm autonomy to their children than mothers who had not engaged in abuse ($d = -0.85$). No significant group differences were found for the subscales “nurture/protect” and “hostile control.”

Skowron et al. (2013) also examined the relationship between physiological regulation (respiratory sinus arrhythmia [RSA]) and parenting behavior in a sample of mothers who had versus had not engaged in maltreatment by observing a mother–child joint teaching task. Significantly lower rates of positive parenting and higher rates of parental strict/hostile control were found in mothers engaging in maltreatment. Furthermore, a greater variance of child physical regulation was observed in dyads not affected by maltreatment, as well as a greater consistency in parenting over time (because standard deviations were not reported in the original study, effect sizes were not calculated).

Classification of Parents as Engaging Versus Not Engaging in Maltreatment

Two out of 13 studies aimed to classify parents as engaging versus not engaging in maltreatment. Brassard et al. (1993) evaluated the PMRS with regard to its ability to discriminate between parents who had versus had not engaged in maltreatment with a focus on psychological maltreatment. The observational measure was used in a 15-min teaching task. They found that the PMRS correctly classified 81.63% of the mothers as engaging versus not engaging in psychological maltreatment. The PMRS showed a sensitivity of 0.92 (correctly identified parents as engaging in maltreatment) and a specificity of 0.71 (correctly identified parents as not engaging in maltreatment). Overall, classification of maltreatment status conducted with the PMRS was superior to classification with other maternal measures (e.g., social support, life stress, personal resources, IQ, depression symptoms).

Cerezo et al. (1996) examined interactive patterns in mother–child interactions by using the SOC-III to discriminate between mothers who had versus had not engaged in maltreatment. They were observed during several 60-min unstructured interactions at home. On average, each family was observed seven times. Coding for maternal behavior correctly identified 70.21% of mothers as engaging versus not engaging in maltreatment (sensitivity = 0.79; specificity = 0.61) and could explain 22% of the variance. Results of coding mother–child interactional behavior correctly identified 78.72% of the mothers as engaging versus not engaging in maltreatment (sensitivity = 0.88; specificity = 0.61) and explained 40% of the variance. A model including all subscales (maternal behavior, child behavior, and interactional behavior) correctly identified 82.98% of the mothers as engaging versus not engaging in maltreatment (sensitivity 0.88; specificity 0.78) and explained 36% of the total variance.

Characteristics of Observational Coding Systems

Table 3 presents an overview of characteristics of the observational coding systems identified through systematic literature search. To guide a specific and targeted selection in a risk-assessment context for each coding system, we present information on the measured constructs, dimensions, and subscales; the observation task; the targeted age range of the child; and the reliability as reported in the studies included in our literature review.

Across observational coding systems, results indicated that measured constructs and dimensions varied significantly due to differences in the underlying theories (e.g., measured constructs varied from maternal sensitivity [CARE-Index; MBQS] over parent–child interaction [CII, DPICS-II, SOC-III, UCLA Parent–Child Coding System],

Table 3 Characteristics of observational coding systems ($N = 11$; in alphabetical order)

Observational coding system	Author (year)	Age range	Measured construct	Dimensions	Subscales	Observation task	Reliability of coding system as reported in included studies
CARE-Index	Crittenden (2004/2006)	0–15 months (15–30 months; toddler version)	Maternal sensitivity Child's interactional behavior	Sensitivity Hostile control Unresponsiveness Cooperative Compulsive Difficult (threatening; toddler version) Passive (disarming; toddler version)	Facial expression Vocal expression Position and body contact Expression of affection Pacing of turns Developmental appropriateness of activity Facial expression Vocal expression Position and body contact Expression of affection Pacing of turns Developmental appropriateness of activity	3-min free play session in a clinical setting	Interrater: $\kappa = 0.91$ for maternal behavior; $\kappa = 0.87$ for child behavior (Ostler, 2010)
CII	Capaldi & Patterson (1989)	Not limited	Parent–child interaction	Parent behavior Child behavior	Parent warmth Appropriate discipline Harsh discipline Physical discipline Child warmth toward parent Child noncompliance toward parent	Coded for general impression after an interaction observation	Internal consistency: $\alpha = 0.84–0.91$ for parenting behavior; $\alpha = 0.67$ for child positive affect (Hurlburt et al., 2013)
DPICS-II	Eyberg et al. (1994)	3–7 years	Parent–child interaction Child deviant behavior	Amount of parental control Demand on child's compliance	27 subscales on parenting behavior 25 subscales on child behavior	In each case 5-min warm up, 15-min child-led play, parent-led play, and clean-up session (without warm up)	Interrater: ICC = 0.70–0.80 for parenting behavior; ICC = 0.79–0.87 for child behavior (Hurlburt et al., 2013)
MBQS	Pederson & Moran (1999)	6–42 months	Maternal sensitivity		90 items that assess features of maternal sensitivity	2–4 hr free home observation	Retest (4 months): $r = 0.71$ for sensitivity Interrater: ICC = 0.71–0.72 (Cicchetti et al., 2006)
P/CIS	Farran et al. (1986)	3–60 months	Parental involvement	Amount Quality Appropriateness	Physical involvement Verbal involvement Responsiveness Play interaction Teaching Control of child's behavior Positive statements Negative statements Directives Goal setting Facilitating relationships among activities	30-min free play in a natural setting	Interrater: $r = 0.88–0.94$ Internal consistency: $\alpha = 0.85–0.96$ (Fagan & Dore, 1993)
Parent–Child Interaction Procedure	Heller et al. (1998)	12–60 months	Parent–child relationship	Parental behavior Child behavior	Emotional responsiveness Behavioral responsiveness Positive affect Irritability Withdrawal/indifference Aggression towards the child Positive affect Withdrawal/Indifference Irritability/anger Noncompliance Aggression towards the parent	Crowell Procedure including seven different tasks (overall 30–45 min): free play, clean up, bubbles, and four teaching tasks	Interrater: $r = 0.71–0.84$ (Robinson et al., 2009)

Table 3 (continued)

Observational coding system	Author (year)	Age range	Measured construct	Dimensions	Subscales	Observation task	Reliability of coding system as reported in included studies
PMRS	Brassard et al. (1993)	5–9 years	Psychological neglect and abuse	Maltreating behavior Positive parenting behavior	Persistence Enthusiasm for the task Spurning Terrorizing Exploiting/Corrupting Denying emotional responsiveness Quality of instruction Supportive presence Respect for child's autonomy Strategies for maintaining task involvement Emotional response to task Mental status Touching Interaction: Experience of mutual pleasure Body harmonics	15-min teaching task	Interrater: $r_w = 0.72-0.94$ Retest (2 weeks): Exact agreement between 46% (denying emotional response) and 100% (terrorizing); Brassard et al., 1993)
Qualitative Ratings of Parent-Child Interactions	Cox (1997)	Not limited	Parenting		Positive regard for the child Negative regard for the child Sensitivity Disengaged Intrusiveness Flatness of affect	30-min interaction: free play, clean up, teaching task	Internal consistency: $\alpha = 0.78$ Interrater: $\kappa = 0.80$ (Haskett et al., 2008; Sabourin Ward & Haskett, 2008)
SOC-III	Cerezo et al. (1986)	Not limited	Parental interaction with the child Child's behavior	Valence Frequency Duration Sequence	Interactional: Approach Instruction Compliance Opposition Noninteractional: No response Interactional: Approach Instruction Compliance Opposition Noninteractional: Play Work Time out Rule violation Nothing Compliant	60-min free interaction in a natural setting	Interrater: $\kappa = 0.62-0.80$ for maternal behavior; $\kappa = 0.48-0.79$ for child behavior (Cerezo et al., 1996)
Benjamin, 1974	Not limited	Transitive (parental behavior) Intransitive (child behavior) Introjection	Warmth Degree of interdependence Warmth Degree of interdependence Warmth Degree of interdependence	Emancipate versus control Affirm versus blame Active love versus attack Protect versus ignore Separate versus submit Disclose versus sulk Reactive love versus recoil Trust versus wall-off Self-emancipate versus self-control Self-affirm versus self-blame	Different kinds of joint and teaching tasks	Interrater: $\kappa = 0.53-1.0$ (Cipriano-Essel et al., 2013); Skowron et al., 2013 $\kappa = 0.64-0.84$ (Skowron et al., 2011)	

Table 3 (continued)

Observational coding system	Author (year)	Age range	Measured construct	Dimensions	Subscales	Observation task	Reliability of coding system as reported in included studies
UCLA Parent–Child Coding System	Valeri et al. (2001)	Not reported	Parent–child interaction	Active self-love versus self-attack Self-protect versus self-neglect Parental behavior Child behavior	Hostility Negative detachment Emotional control Support Positivity Hostility Positivity Oppositionality Demandingness Compliance	13-min interaction task: teaching, planning, conflict	Interrater: ICC = 0.62 for parent behavior; ICC = 0.62 for child behavior (Lau et al., 2006)

Notes. CARE-Index = Child–Adult Relationship Experimental Index; CII = Coder Impressions Inventory; DPICS-II = Dyadic Parent–Child Interaction Coding System-II; MBQS = Maternal Behavior Q-Sort; P/CIS = Parent/Caregiver Involvement Scale; PMRS = Psychological Maltreatment Rating Scales; SOC-III = Standardized observation codes; SASB = Structural Analysis of Social Behavior; UCLA-Parent–Child Coding System = University of California Los Angeles-Parent–Child Coding System

to psychological maltreatment [PMRS]). Reliability coefficients reported by the individual studies ranged from satisfactory to high for all coding systems.

Discussion

This systematic review is the first to give a detailed overview of the current state of research on studies analyzing the potential of specific observational coding systems to discriminate parents who have versus have not engaged in child maltreatment based on their behavior when interacting with their child. We aim to provide professionals who have to perform a child abuse/neglect risk evaluation with information about the utility of different observational coding systems for parent–child interactions.

Observational Assessment of Child Maltreatment Risk

Behavioral observation of parent–child interaction is one of the core methods in child maltreatment risk assessment along with file analyses, interviews with parents and children, psychometric testing, and analyses of third-party reports (American Psychological Association, 2013). Through our systematic literature search, we found 13 studies that detected behavioral differences between parents who had and parents who had not engaged in child maltreatment. The findings add to our understanding of the assessment tools relevant to child maltreatment risk, because they allow some conclusions on which parenting dimensions appear to best differentiate parents engaging in child maltreatment from those not engaging in child maltreatment.

The largest body of evidence supports the conclusion that significantly lower levels of “parental sensitivity/responsiveness” differentiate parents who have engaged in child maltreatment. This comprises lower levels of sensitivity to the signals of the child and lower levels of understanding, empathic, and comforting responses to the child’s emotional distress. This finding was replicated in five of the included studies with effect sizes ranging from small to large (Cicchetti et al., 2006; Fagan & Dore, 1993; Haskett et al., 2008; Ostler, 2010; Sabourin Ward & Haskett, 2008).

Findings on significantly higher levels of “strict/hostile control” or “critical and controlling behavior” (e.g., harsh commands) in parents who had engaged in maltreatment were replicated in four of the included studies with either small or large effect sizes (Cipriano-Essel et al., 2013; Hurlburt et al., 2013; Skowron et al., 2011; Skowron et al., 2013).

Three out of the 13 studies reported significantly lower levels of supportive and developmentally appropriate behavior (e.g., behavior consistent with the child's age, developmental stage, and abilities) in parents who had engaged in child maltreatment (Fagan & Dore, 1993; Hurlburt et al., 2013; Lau et al., 2006). Two studies revealed differences on emotional dimensions with parents who had engaged in maltreatment showing significantly higher levels of emotional control and parental anger and significantly lower levels of positive affect (Lau et al., 2006; Robinson et al., 2009).

Therefore, behavioral dimensions that may indicate differences in parental behavior focus not only on a lack of adaptive parental behavior but also on the presence of maladaptive parental behavior. Overall, the results of our systematic literature review support the conclusion that detecting a lack of adaptive parental behavior and the presence of maladaptive parental behavior by behavioral observation may be a valid approach with which to assess child maltreatment risk. However, our results indicate that this is limited to observations that focus systematically on specific dimensions such as “parental sensitivity and responsiveness,” “hostility and control,” “developmentally appropriate behavior,” “parental anger,” and “positive affect.”

The findings from our systematic review provide mixed results regarding which structured observational coding systems can measure child maltreatment risk in parenting behavior most reliably and validly. We identified 11 different observational coding systems that were applied by the individual studies, with the SASB being applied most frequently. Across observational coding systems, measured constructs and dimensions varied significantly, primarily because they were developed on the basis of different underlying theories. Reliability coefficients as reported in the original studies were generally satisfactory to high. This also applies to the high-risk samples, even though some of the coding systems were not originally developed for a risk assessment context.

Only two of the instruments identified (CARE-Index; PMRS) were specifically developed to detect maltreatment risks or psychological abuse. The psychometric properties and the specificity of the CARE-Index support its applicability in a child maltreatment risk evaluation context. Because observations conducted within the CARE-Index take only 3 to 5 min, and the free play session is neither invasive nor stressful, this may be beneficial in an evaluation context. The observation is not only economical to perform but can also be repeated multiple times and with several caregivers. Nonetheless, the CARE-Index is limited to a narrow age range and targets mainly younger children.

The correct classification of more than 80% of the mother–child dyads as psychologically maltreating versus

nonmaltreating might indicate that the PMRS is an appropriate measure to detect psychological maltreatment risk. However, the PMRS is limited in its ability to specifically detect psychological maltreatment and therefore does not cover the entire range of parental maltreating behavior. Furthermore, the specificity score was rather low and the reliability and the validity of the instrument remain open to further investigation.

Because the P/CIS assesses the construct “involvement” defined as the “amount, quality, and appropriateness of interactions between caregivers and children” (Fagan & Dore, 1993, p. 62), it may be a valid approach to specifically identify parents who had engaged in neglect, but possibly not the entire range of maltreating behavior. Neglect and involvement are commonly discussed as being highly correlated. Involvement seems to be a good predictor of neglect, but less for physical or psychological maltreatment (Wilson et al., 2008).

The DPICS-II appears to be a measure that is well-validated in different contexts. Observations conducted within the DPICS-II are efficient due to the short observation time, the fact that they can be repeated multiple times and with several caregivers, and the fact that they include a structured approach (5-min warm-up, 15-min child-led play, parent-led play, and clean-up session). However, the study assessing the DPICS-II included here found rather inconclusive results on whether parenting behavior indicated maltreatment risks as coded by DPICS-II scales.

The other observational coding systems show some individual strengths, but also some limitations regarding their applicability in a child maltreatment risk evaluation context: The SASB, the CII, the Parent–Child Interaction Procedure, and the Qualitative Ratings of Parent–Child Interaction are based on theoretical frameworks that have a low specific relation to child maltreatment. For the MBQS as well as the SOC-III, observation is conducted over several hours during natural home interactions. This requires enormous economic effort, is a high burden on the observed family, and leads to limitations regarding the replicability and comparability of the observations.

Note that across studies, the effect sizes calculated for the studies' outcomes indicate a majority of small effects. Sample sizes were rather small, with only 61.5% of the studies analyzing sample sizes larger than 100 dyads. As a small sample size does not, in itself, lead to small effect sizes, it however might have led to an increase of type-II errors due to low statistical power in some of the studies. On the other hand, despite small effect sizes, many effects were significant. Thus, small differences might indicate that the assessment approach is limited to clearly distinguishing between maltreating and nonmaltreating dyads. Rather, they can be used as one source of information about this distinction.

Our findings make it challenging to derive a sound recommendation on the applicability of specific observational coding systems in the context of child maltreatment risk evaluation. Our systematic review supports the conclusion that the selection of an observational measure for a risk assessment context cannot be generalized. Along with factors influencing the selection that lie within the individual case (such as the age of the child), the relative value of parent–child observations needs to take into account the discriminant validity, a potential focus on different types of child maltreatment (e.g., neglect, physical abuse, psychological abuse), as well as the position and value of parent–child observations in the entire risk evaluation.

This may partly explain that although there are several observation coding systems that discriminate behavioral differences between parents who have versus those who have not engaged in child maltreatment, none has gained widespread acceptance in risk assessment evaluation practice. Currently, when providing evidence-based assessments, mental health professionals may be encouraged to select a structured coding system when analyzing their behavioral data. Data should include the key behavioral dimensions that have been shown to discriminate the behavior of parents who have engaged in child maltreatment from parents who have not, and should acknowledge the age of the child in the individual case. Moreover, scoring with systematic observational coding systems generally requires extensive training.

Additionally, when interpreting findings from behavioral observation in a maltreatment risk assessment context, it is important to highlight the need to include cultural considerations. Some behavior and interactions might be misconstrued as pathological if not understood in the context of cultural norms, beliefs, and practices. Moreover, there is a danger that efforts to be culturally sensitive might also result in cultural relativism, perhaps by overlooking the risks or damage to the child as a result of culturally sanctioned practices (i.e., harsh discipline, or oppression based on gender). In 6 of the 13 studies we identified in this systematic review, the sample consisted mainly of persons belonging to ethnic minorities, whereas in six further studies, the majority of the sample was Caucasian. When interpreting findings from behavioral observation, it is imperative to take an understanding of the family's cultural background into consideration—including not only race and ethnicity, but also socioeconomic conditions, linguistic preferences, gender, sexual orientation, religious and spiritual practices, immigration status, and other forms of diversity (Thomas et al., 2019).

Clearly, the results of this systematic review must be embedded within the broader context of child maltreatment risk assessment in which behavioral observation is only one of several components and in which a lack of parenting

skills, parent–child interaction problems, or a maladaptive parent–child relationship are only one part of a larger set of risk factors that need to be assessed.

It is therefore undisputed that a child maltreatment risk assessment needs to take a multidimensional approach; and that an expert recommendation should never be based on one-dimensional findings. No single tool is meant to allow an identification of whether a caregiver is at risk for engaging in maltreatment. Therefore, next to caregiver–child observation, a full evaluation typically also includes file analyses, structured/semistructured caregiver interviews, structured/semistructured child interviews, analyses of third-party reports, standardized developmental assessments of children, and psychological assessments of parents, as for example described in the Guidelines for Psychological Evaluations in Child Protection Matters (American Psychological Association, 2013).

The few observational coding systems that examined the ability of the assessments to correctly classify maltreating versus nonmaltreating parents found that there were many false positives, and the effect sizes we identified in the individual studies indicated around one third to two fifths of one standard deviation difference between maltreating and nonmaltreating dyads. Therefore, findings based on parent–child observation need to be embedded in this multidimensional approach. Observational techniques are clearly limited when it comes to assessing other relevant risk factors in order to predict future child maltreatment such as parent-related risk factors (e.g., mental disorders, substance use problems) or family factors (e.g., socioeconomic stressors, lack of social support, intimate partner violence; see de Ruiter et al., 2020). Hence, any final conclusion on the risk for future child maltreatment in an individual case must be based on a comprehensive integration of multidimensional findings (see American Psychological Association, 2013).

Further classification guidelines, such as the Diagnostic Classification of Mental Health and Developmental Disorders of Infancy and Early Childhood (DC: 0–3 R; ZERO TO THREE, 2016) highlight the relevance of a relational lens to guide an assessment of disordered to dangerous parent–child relationships, and it acknowledges the complexity of incorporating cultural considerations into the diagnostic process. A diversity, inclusion, and fairness lens should be applied to all practices and services aimed at supporting infants, toddlers, and their families if they are to be of value in the given context (Thomas et al., 2019).

Limitations

When interpreting the findings from this systematic review, several aspects should be considered that indicate limits to their generalizability when answering our research question. Samples consisted mainly of mother–child dyads with

children being between 1 and 8 years of age. Only five studies involved other types of caregivers (such as fathers, step-fathers, or grandparents). In child maltreatment risk evaluation, it may also be the father or another caregiver who is alleged to be maltreating. Furthermore, only one study focused specifically on a sample of mothers suffering from mental illness (Ostler, 2010), although parental mental disorders commonly occur in child maltreatment risk evaluation practice.

We identified a number of behavioral dimensions that have been shown to discriminate the behavior of parents who did versus did not engage in child maltreatment. However, it cannot be ruled out that there are further behavioral dimensions with this potential that were not captured by the studies included in this systematic review.

Most studies classified maltreatment via having a substantiated history of maltreatment as recorded in child protection service reports. Not all studies differentiated between maltreatment forms such as psychological maltreatment, physical maltreatment, or neglect. However, specifics in parental behavior might be expected based on the specific maltreatment form.

All but one of the studies were conducted in the United States. Most studies identified did not consider confounding variables such as sociodemographic variables (parental age, parental education, gender, family ethnic background, family income), further risk variables (psychopathology, substance use, domestic violence), or methodological variables (regional differences between various local authorities investigating maltreatment, etc.) either when classifying parents or when determining differences in parental behavior. Replications of the findings when controlling for these variables, as well as replications in samples from other countries, are options for future research.

Furthermore, the small total number of included studies, the heterogeneous contexts, and the varying aims of the studies limit the number of parameters that could be chosen for this synthesis. With regard to the child abuse/neglect risk assessment context, our focus remained on studies discriminating the behavior of parents who have versus have not engaged in child maltreatment. This led to the exclusion of numerous studies (e.g., evaluations of intervention programs), because many analyzed samples consisting only of parents with a maltreatment history. For this reason, some studies applying well-established observation coding systems (e.g., the Home Observation Measurement of the Environment [HOME]; Caldwell & Badley, 1984) were excluded, even though they are frequently applied in high-risk samples.

In addition, a large number of the studies we excluded focused on deviant child behavior during the observation. This can be a consequence of maltreatment. Yet, even though maltreated children are at higher risk for deviant

behavior because of maltreatment, the presence of child deviant behavior itself is not an indicator of child maltreatment. Many children demonstrate deviant behavior without having a maltreatment history (Cerezo, 1997; Crittenden, 1988; Dadds et al., 2002; Herrenkohl et al., 1984; Lahey et al., 1984). This may have led to a further exclusion of observation coding systems that are frequently applied in high-risk samples.

Conclusions and Future Directions

In conclusion, the findings from this review indicate a remarkable discrepancy between the impact of behavior observations for maltreatment risk assessment as reported in several studies, and the limited empirical evidence on the applicability and psychometric quality of specific observational coding systems in this assessment context (Bennett et al., 2006; Budd, 2001; Garber, 2016; Wilson et al., 2008; Zumbach & Koglin, 2015). This apparent discrepancy leads to the need for increased effort to study constructs and measures indicating child maltreatment risk. This includes the further development of appropriate and specific coding systems for behavior observation. Determining the parenting capacity of a caregiver is often crucial for maltreatment risk assessment, and parenting behavior in parent–child interactions is a main indicator. Therefore, the development of refined observational measures and their validation remain a challenge for future research in order for such measures to be of use in the family assessments that family courts commission in order to inform their decisions.

Beyond this, it should still be mentioned that there may well be parenting behavior harming the child that is difficult to observe in a given setting. For example, one important form of child psychological abuse is the use of alienating behavior by one parent to cause the child to fear and avoid a relationship with the other parent. In cases of severe parental alienation, alienating behavior is often blatant and pervasive. Consequently, identifying specifics in parenting behavior indicating different forms of psychological maltreatment is a further challenge for future research.

The complexity and specificity of child maltreatment risk assessment indicates the need to develop multifaceted measures. Considering the high requirements, the aim is not to develop a general instrument, but rather to develop different specific and combinable observational measures for various age ranges and developmental levels of the child. Ultimately, the technique of behavioral observation remains one of several components within the body of assessment techniques for determining child maltreatment risks.

Funding Open Access funding enabled and organized by Projekt DEAL.

Compliance with Ethical Standards

Conflict of Interest The authors declare no competing interests.

Ethical Approval No ethical approval was obtained because data from previous published studies in which informed consent was obtained by primary investigators were retrieved and analyzed.

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ainsworth, M. D., & Wittig, B. A. (1969). Attachment and exploratory behavior of one-year-olds in a strange situation. In B. M. Foss (Ed.), *Determinants of infant behavior* (Vol. 4, pp. 113–136). Methuen.
- American Psychological Association. (2013). Guidelines for psychological evaluations in child protection matters. *American Psychologist*, 68, 20–31. <https://doi.org/10.1037/a0029891>.
- Assink, M., van der Put, C. E., Meeuwse, M. W. C. M., de Jong, N. M., Oort, F. J., Stams, G. J. J. M., & Hoeve, M. (2019). Risk factors for child sexual abuse victimization: A meta-analytic review. *Psychological Bulletin*, 145, 459–489. <https://doi.org/10.1037/bul0000188>.
- Barber, J. G., & Delfabbro, P. (2000). The assessment of parenting in child protection cases. *Research on Social Work Practice*, 10, 243–256. <https://doi.org/10.1177/104973150001000206>.
- Benjamin, L. S. (1974). Structural analysis of social behavior. *Psychological Review*, 81, 392–425. <https://doi.org/10.1037/h0037024>.
- Bennett, D. S., Sullivan, M. W., & Lewis, M. (2006). Relations of parental report and observation of parenting to maltreatment history. *Child Maltreatment*, 11, 63–75. <https://doi.org/10.1177/1077559505283589>.
- Brassard, M. R., Hart, S. N., & Hardy, D. B. (1993). The psychological maltreatment rating scales. *Child Abuse & Neglect*, 4, 715–729. [https://doi.org/10.1016/S0145-2134\(08\)80003-8](https://doi.org/10.1016/S0145-2134(08)80003-8).
- Budd, K. S. (2001). Assessing parenting competence in child protection cases: A clinical practice model. *Clinical Child and Family Psychology Review*, 4, 1–18. <https://doi.org/10.1016/j.childyouth.2004.11.008>.
- Budd, K. S., & Holdsworth, M. J. (1996). Issues in clinical assessment of minimal parenting competence. *Journal of Clinical Child Psychology*, 25, 1–14. https://doi.org/10.1207/s15374424jccp2501_1.
- Caldwell, B., & Bradley, R. H. (1984). *Home observation for measurement of the environment*. University of Arkansas at Little Rock.
- Capaldi, D., & Patterson, G. R. (1989). *Psychometric properties of fourteen latent constructs from the Oregon Youth Study*. Springer.
- Cerezo, M. A. (1997). Abusive family interaction: A review. *Aggression and Violent Behavior*, 2, 215–240. [https://doi.org/10.1016/S1359-1789\(96\)00025-0](https://doi.org/10.1016/S1359-1789(96)00025-0).
- Cerezo, M. A., D'Ocon, A., & Dolz, L. (1996). Mother–child interactive patterns in abusive families versus nonabusive families: An observational study. *Child Abuse & Neglect*, 20, 573–587. [https://doi.org/10.1016/0145-2134\(96\)00045-2](https://doi.org/10.1016/0145-2134(96)00045-2).
- Cerezo, M. A., Keesler, T., Dunn, E. Y., & Wahler, R. G. (1986). *Standardized observation codes: SOC III* [Unpublished manuscript]. Child Behavior Institute. University of Tennessee, Knoxville, TN.
- Cicchetti, D., Rogosch, F. A., & Toth, S. L. (2006). Fostering secure attachment in infants in maltreating families through preventive interventions. *Development and Psychopathology*, 18, 623–649. <https://doi.org/10.1017/S0954579406060329>.
- Cipriano-Essel, E., Skowron, E. A., Stifter, C. A., & Teti, D. M. (2013). Heterogeneity in maltreated and non-maltreated preschool children's inhibitory control: The interplay between parenting quality and child temperament. *Infant and Child Development*, 22, 501–522. <https://doi.org/10.1002/icd.1801>.
- Cox, M. (1997). *Qualitative Ratings of Parent/Child Interaction* [Unpublished manual]. University of North Carolina at Chapel Hill.
- Crittenden, P. M. (1988). Relationship at risk. In J. Belsky, & T. Nezworski (Eds.), *Clinical implications of attachment* (pp. 136–167). Erlbaum.
- Crittenden, P. M. (2004). *CARE-Index infants (Birth–15 months). Coding manual*. Family Relations Institute.
- Crittenden, P. M. (2006). *CARE-Index Toddlers Coding Manual* [Unpublished manuscript]. Family Relations Institute.
- Cyr, C., & Alink, L. R. A. (2017). Child maltreatment: The central role of parenting capacities and attachment. *Current Opinion in Psychology*, 15, 81–85. <https://doi.org/10.1016/j.copsyc.2017.02.002>.
- Dadds, M. R., Mullins, M. J., McAllister, R. A., & Atkinson, E. (2002). Attributions, affect, and behavior in abuse-risk mothers: A laboratory study. *Child Abuse & Neglect*, 27, 21–45. [https://doi.org/10.1016/S0145-2134\(02\)00510-0](https://doi.org/10.1016/S0145-2134(02)00510-0).
- de Ruiter, C., Hildebrand, M., & van der Hoorn, S. (2020). The child abuse risk evaluation dutch version (CARE-NL): A retrospective validation study. *Journal of Family Trauma, Child Custody & Child Development*, 17(1), 37–57. <https://doi.org/10.1080/15379418.2019.1699488>.
- Eyberg, S. M., Bessmer, J. L., Newcomb, K., Edwards, D. L., & Robinson, E. A. (1994). *Dyadic Parent–Child Interaction System-II: A Manual*. Select Press.
- Fagan, J., & Dore, M. M. (1993). Mother–child play interaction in neglecting and non-neglecting mothers. *Early Child Development and Care*, 87, 59–68. <https://doi.org/10.1080/0300443930870106>.
- Farran, D. C., Kasari, C., Comfort, M., & Jay, S. (1986). *The Parent/Caregiver Involvement Scale: Manual*. Center for the Development of Early Education, Kamehameha Schools/Bishop Estates, Honolulu, Hawaii.
- Finkelhor, D. (1994). The international epidemiology of child sexual abuse. *Child Abuse & Neglect*, 18, 409–417. [https://doi.org/10.1016/0145-2134\(94\)90026-4](https://doi.org/10.1016/0145-2134(94)90026-4).
- Garber, B. D. (2016). Exploring a process-oriented forensic family observation protocol. *Family Court Review*, 54, 261–276. <https://doi.org/10.1111/fcre.12217>.
- Gilbert, R., Widom, C. S., Browne, K., Fergusson, D., Webb, E., & Janson, S. (2009). Burden and consequences of child maltreatment in high-income countries. *Lancet*, 373, 68–81. [https://doi.org/10.1016/S0140-6736\(08\)61706-7](https://doi.org/10.1016/S0140-6736(08)61706-7).

- Harnett, P. H. (2007). A procedure for assessing parents' capacity for change in child protection cases. *Children and Youth Services Review*, 29, 1179–1188. <https://doi.org/10.1016/j.childyouth.2007.04.005>.
- Haskett, M. E., Allaire, J. C., Kreig, S., & Hart, K. C. (2008). Protective and vulnerability factors for physically abused children: Effects of ethnicity and parenting context. *Child Abuse & Neglect*, 32, 567–576. <https://doi.org/10.1016/j.chiabu.2007.06.009>.
- Haynes S. N., & O'Brien W. H. (2000). Principles and practice of behavioral assessment. Springer. https://doi.org/10.1007/978-0-306-47469-9_12
- Heller, S., Aoki, Y., Crowell, J., Chase-Lansdale, L., Brooks-Gunn, J., Schoffner, K., & Zamsky, E. (1998). Parent child interaction procedure: Coding manual. *Unpublished manuscript*.
- Herrenkohl, E. C., Herrenkohl, R. C., Toedter, L., & Yanoushefski, A. M. (1984). Parent-child interactions in abusive and nonabusive families. *Journal of the American Academy of Child Psychiatry*, 23, 2641–2648. [https://doi.org/10.1016/S0002-7138\(09\)60530-6](https://doi.org/10.1016/S0002-7138(09)60530-6).
- Hurlburt, M. S., Nguyen, K., Reid, J., Webster-Stratton, C., & Zhang, J. (2013). Efficacy of the Incredible Years group parent program with families in Head Start who self-reported a history of child maltreatment. *Child Abuse & Neglect*, 37, 531–543. <https://doi.org/10.1016/j.chiabu.2012.10.008>.
- Kähkönen, P. (1999). The assessment of parenting in the child welfare practice. *Children and Youth Services Review*, 21, 581–603. [https://doi.org/10.1016/S0190-7409\(99\)00040-7](https://doi.org/10.1016/S0190-7409(99)00040-7).
- Lahey, B. B., Conger, R. D., Atkinson, B. M., & Treiber, E. A. (1984). Parenting behavior and emotional status of physically abusive mothers. *Journal of Consulting and Clinical Psychology*, 52, 1062–1071. <https://doi.org/10.1037/0022-006X.52.6.1062>.
- Lau, A. S., Valeri, S. M., McCarty, C. A. & Weisz, J. R. (2006). Abusive' parents reports of child behavior problems: relationship to observed parent-child interactions. *Child Abuse & Neglect*, 30, 639–655. <https://doi.org/10.1016/j.chiabu.2005.11.009>.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G., The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. *PLoS Medicine*, 6, e1000097 <https://doi.org/10.1371/journal.pmed.1000097>.
- Mulder, T. M., Kuiper, K. C., van der Put, C. E., Stams, G., & Assink, M. (2018). Risk factors for child neglect: A meta-analytic review. *Child Abuse & Neglect*, 77, 198–210. <https://doi.org/10.1016/j.chiabu.2018.01.006>.
- Ostler, T. (2010). Assessing parenting risk within the context of severe and persistent mental illness: Validating an observational measure for families with child protective service involvement. *Infant Mental Health Journal*, 31, 467–485. <https://doi.org/10.1002/imhj.20267>.
- Pederson, D. R., Moran, G., & Bento, S. (1999). *Maternal behavior Q-sort manual. Version 3.1*. Retrieved October 21, 2019 from www.ssc.uwo.ca/psychology/faculty/pedmor/mbqmanual.html.
- Robinson, L. R., Sheffield Morris, A., Scott Heller, S., Scheeringa, M. S., Boris, N. W., & Smyke, A. T. (2009). Relations between emotion regulation, parenting, and psychopathology in young maltreated children in out of home care. *Journal of Child & Family Studies*, 18, 421–434. <https://doi.org/10.1007/s10826-008-9246-6>.
- Sabourin Ward, C., & Haskett, M. E. (2008). Exploration and validation of clusters of physically abused children. *Child Abuse & Neglect*, 32, 577–588. <https://doi.org/10.1016/j.chiabu.2007.07.012>.
- Skowron, E. A., Cipriano-Essel, E., Benjamin, L. S., & Pincus, A. L. (2013). Cardiac vagal tone and quality of parenting show concurrent and time-ordered associations that diverge in abusive, neglectful, and non-maltreating mothers. *Couple and Family Psychology: Research and Practice*, 2, 95–115. <https://doi.org/10.1037/cfp0000005>.
- Skowron, E. A., Loken, E., Gatzke-Kopp, L. M., Cipriano-Essel, E. A., Woehle, P. L., Van Epps, J. J., Gowda, A., & Ammerman, R. T. (2011). Mapping cardiac physiology and parenting processes in maltreating mother-child dyads. *Journal of Family Psychology*, 25, 663–674. <https://doi.org/10.1037/a0024528>.
- Sledjeski, E. M., Dierker, L. C., Brigham, R., & Breslin, E. (2008). The use of risk assessment to predict recurrent maltreatment: A classification and regression tree analysis (CART). *Prevention Science*, 9, 28–37. <https://doi.org/10.1007/s11121-007-0079-0>.
- Smith Slep, A. M., Heyman, R. E., & Foran, H. M. (2015). Child maltreatment in DSM-5 and ICD-11. *Family Process*, 54, 17–32. <https://doi.org/10.1111/famp.12131>.
- Stith, S., Liu, T., Davies, L., Boykin, E. L., Alder, M. C., Harris, J. M., Som, A., Mcpherson, M., & Dees, J. (2009). Risk factors in child maltreatment: A meta-analytic review of the literature. *Aggression and Violent Behavior*, 14, 13–29. <https://doi.org/10.1016/J.AVB.2006.03.006>.
- Thomas, J., Brunton, J., & Graziosi, S. (2010). *EPPI-Reviewer 4: Software for research synthesis*. EPPI-Centre Software. Social Science Research Unit, UCL Institute of Education.
- Thomas, K., Noroña, C. R., St. John, M. S. (2019). Cross-sector allies together in the struggle for social justice: Diversity-informed tenets for work with infants, children, and families. *Zero to Three Journal*. Retrieved August 24, 2021 from <https://www.irvingharrisfdn.org/wp-content/uploads/2019/01/Thomas-Norona-St.-John-January-2019-Tenets-2nd-Edition.pdf>.
- Toth, S. L., & Cicchetti, D. (2013). A developmental psychopathology perspective on child maltreatment. *Child Maltreatment*, 18, 135–139. <https://doi.org/10.1177/1077559513500380>.
- Valeri, S. M., Lau, A. S., & McCarty, C. A. (2001). *Microanalytic coding interpretive manual: UCLA clinic study* [Unpublished manuscript]. University of California.
- Wilson, S. R., Rack, J. J., Shi, X., & Norris, A. M. (2008). Comparing physically abusive, neglectful, and non-maltreating parents during interactions with their children: A meta-analysis of observational studies. *Child Abuse & Neglect*, 32, 897–911. <https://doi.org/10.1016/j.chiabu.2008.01.003>.
- World Health Organization (2016). *Child maltreatment*. [online] Retrieved September 16, 2019 from <https://www.who.int/news-room/fact-sheets/detail/child-maltreatment>.
- ZERO TO THREE. (2016). *DC: 0–5 diagnostic classification of mental health and developmental disorders of infancy and early childhood*. ZERO TO THREE.
- Ziegenhain, U., Fegert, J. M., Ostler, T., & Buchheim, A. (2007). Risk assessment of neglect and maltreatment in infants and toddlers. Chances of early preventive intervention. *Praxis für Kinderpsychologie & Kinderpsychiatrie*, 56, 410–428. <https://doi.org/10.13109/prkk.2007.56.5.410>.
- Zumbach, J., & Koglin, U. (2015). Psychological evaluations in family law proceedings: A systematic review of the contemporary literature. *Professional Psychology Research and Practice*, 46, 221–234. <https://doi.org/10.1037/a0039329>.
- Zumbach, J., & Oster, A. (2021). Elterliche Erziehungsfähigkeit: Definitionen, Indikatoren und Erfassungsmöglichkeiten [Parenting capacity: Definitions, indicators, and assessment]. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 49, 37–50. <https://doi.org/10.1024/1422-4917/a000770>.