Clin Child Fam Psychol Rev
DOI 10.1007/s10567-015-0180-z

Observational Tools for Measuring Parent–Infant Interaction: A Systematic Review

Annett Lotzin · Xiaoxing Lu · Levente Kriston · Julia Schiborr · Teresa Musal · Georg Romer · Brigitte Ramsauer

Abstract The quality of the parent–infant interaction is essential for the infant's development and is most objectively measured by observation. The existing observational tools for assessing parent–infant interaction were identified and described, and their psychometric soundness was evaluated. Twenty electronic databases from inception through June 2013 were searched. Validity was evaluated in five domains (test content, response process, internal structure, relations to other variables, and consequences). Of the 23,961 citations identified, 24 tools were included. Most tools demonstrated a valid rating procedure, reproducibility, and discriminant validity, based on studies with credible quality. The tools lacked factorial and predictive validity, and standardized norms. Further refinement of the existing tools is needed, particularly in the domains of content validity and consequential validity. The synthesized validity evidence and descriptions of the tools reported in this review might guide clinicians and researchers in the selection of an appropriate tool.

Keywords Mother–infant interaction · Father–infant interaction · Behavioral observation · Measurement · Psychometrics · Validation studies

Electronic supplementary material The online version of this article (doi:10.1007/s10567-015-0180-z) contains supplementary material, which is available to authorized users.

A. Lotzin (✉) · J. Schiborr · G. Romer · B. Ramsauer
Department of Child and Adolescent Psychiatry, Psychotherapy and Psychosomatics, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany
e-mail: A.Lotzin@uke.de

Present Address:
A. Lotzin
Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

X. Lu · T. Musal
Department of Psychology, University of Hamburg, Hamburg, Germany

L. Kriston
Department of Medical Psychology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Present Address:
G. Romer
Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital Muenster, Muenster, Germany

Introduction

Pediatric (Committee on Psychosocial Aspects of Child Family Health 1997; Hagan et al. 2008) and Early Head Start (Early Head Start National Resource Center 2013) initiatives call for the routine observation of the parent–infant interaction in clinical work, early intervention, and research, stating that the "child's relationship and interactions with his or her caregiver should form the cornerstone of the assessment" (Early Head Start National Resource Center 2013, p. 6). These recommendations have been informed by research, indicating that the quality of the parent–infant interaction is crucial for the child's development in early childhood, including attachment (Barnard et al. 1989; Biringen et al. 2005; De Wolff and van Ijzendoorn 1997; Evans and Porter 2009), social–emotional development (Hofheimer and Appelbaum 1992; Kochanska et al. 2005a), cognitive development (Barnard et al. 1989; Evans and Porter 2009; Feldman et al. 2004; Hofheimer

Published online: 03 April 2015

and Appelbaum 1992) and clinical outcome (Ramchandani et al. 2013; Wan et al. 2012). Furthermore, the observed quality of the parent–infant interaction predicts the child's social and cognitive development (Stams et al. 2002; Stright et al. 2008) and academic competence (Stright et al. 2008) in middle childhood. Consequently, observation of the interaction between parent and infant is crucial for identifying strengths and problems that may impact the child's developmental outcomes.

Observational tools of parent–infant interaction permit an objective assessment of the parent's and the infant's behavior, whereas self-reports might be biased by the parent's linguistic skills, thoughts, and feelings, or the tendency to respond in a socially desirable manner (Corcoran and Fischer 2013). Microanalytic observational tools enable to assess the fine-grained details of the parent–infant interaction that often occur without awareness: how the behavior between parent and infant unfolds over time, and how the parent's or the infant's behavior is influenced by the behavior of the interaction partner (Gardner 2000). But observational tools involve a higher investment in training, technical equipment, and time for administration than questionnaire assessments (Bagner et al. 2012). Parent-report questionnaires might be preferred when less complex information about the parent–infant interaction is required; they are often brief, easy to administer, and particularly helpful for identifying infants who need more extensive assessment (Halle et al. 2010). Parent-reports even outweigh observational assessments if the addressed construct is difficult to observe, such as the parent's feelings, thoughts, or perceived relationship quality between parent and infant (Gardner 2000).

Importantly, observational and self-report assessments are non-interchangeable: They correlate only weakly to moderately (Alderfer et al. 2008). Therefore, an evaluation of the parent–infant interaction should be guided by the particular purpose of the assessment (Snow et al. 2008), for example, to screen for potential problems for further evaluation, or to diagnose problems in the parent–infant interaction to determine the need for intervention. If a comprehensive evaluation of the parent–infant interaction is needed, this should ideally include both parent-reports and behavioral observations (Bagner et al. 2012; Halle et al. 2010; Shepard et al. 1998).

Observational tools of parent–infant interaction cover a broad range of constructs. The most prevalent constructs addressed by observational tools include aspects of responsive caregiving, such as sensitivity, contingent responsiveness, or emotional availability. These constructs have been shown to be related to infants' attachment and cognitive development: Infants that had been exposed to more sensitive or responsive parents during their first year of life more frequently developed a secure attachment (De Wolff and van Ijzendoorn 1997; NICHD Early Child Care Research Network 1997) and were more likely to develop advanced attentional (Gartstein et al. 2008), cognitive (Feldman et al. 2004) or language skills (Tamis-LeMonda et al. 2001), compared with infants exposed to less sensitive or responsive parents. Some observational tools address facets of learning support, such as scaffolding, structuring, or encouraging the infant to explore objects. Infants of parents that had fostered their cognitive growth during early infancy later developed better cognitive skills (Barnard et al. 1989; Oxford and Findlay 2012). Observational tools of parent–infant interaction may also cover parent behaviors that interfere with infant development, for example, intrusiveness, control, or hostility. Infants of parents that had been more intrusive (Beebe et al. 2010; Biringen et al. 2005) or hostile (Biringen et al. 2005) were less likely to develop a secure attachment and enhanced cognitive abilities (Bernstein et al. 1987) than infants of parents that had been less intrusive or hostile.

Infant behaviors measured by observational tools of parent–infant interaction often cover infant responsiveness, engagement, or involvement. Infants that had been more responsive during their first year of life were more likely to develop a secure attachment (Barnard et al. 1989; Biringen et al. 2005), or advanced cognitive skills (Barnard et al. 1989). Observational tools may also measure infant positive or negative affect. Infants that had expressed more positive vocal or facial affect during early infancy were more likely to develop a secure attachment (Braungart-Rieker et al. 2001; Cohn et al. 1991), whereas infants that had displayed higher vocal distress, or more combined expressions of vocal and facial distress, were more likely to develop a disorganized than a secure attachment (Beebe et al. 2010). Infant clarity of cues is covered by some of the observational tools, which has been shown to predict infant secure attachment (Barnard et al. 1989) and cognitive abilities (Oxford and Findlay 2012).

On the dyadic level, observational tools of parent–infant interaction measure the coordination of the behaviors between parent and infant by constructs such as synchrony, reciprocity, or mutuality. Both lowered and heightened levels of synchrony or reciprocity may impair infant development. For example, infants of parent–infant dyads that had been shown lowered or heightened parent–infant synchrony of their facial or vocal affect more frequently developed unfavorable attachment outcomes (Beebe et al. 2010; Jaffe et al. 2001), compared with infants of dyads with medium levels of synchrony. Deviant parent–infant synchrony was also related to infants' later cognitive development (Gartstein et al. 2008; Kochanska et al. 2005b).

Although these relations between specific parent–infant interaction constructs and infants' later development have been documented—particularly for infants' attachment and cognitive abilities—it should be noted that most parent–infant interaction studies predict the infants' later development by a tool's total score derived from summing up different parent–infant interaction constructs (Mahoney et al. 1996).

The selection of a particular observational tool of parent–infant interaction is challenging because it requires extensive knowledge about the constructs needed for this specific purpose (Early Head Start National Resource Center 2013). Researchers or clinicians should also consider the availability of tools for measuring the behaviors of interest, and the limitations, as well as the strengths of these often complex tools (Early Head Start National Resource Center 2013; Margolin et al. 1998). This information may be difficult to obtain because most observational tools are not published, and no tool is accepted as the gold standard for measuring parent–infant interaction. But if an observational tool is not chosen carefully, the assessment will be costly in time and resources and may result in incorrect judgments that are useless or even harmful to the evaluated parents and infants (Mahoney et al. 1996).

Knowledge of research findings related to a tool's psychometric properties (i.e., validity evidence) is essential for the purposeful assessment of a tool. The "standards for educational and psychological testing" (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education 1999) outline five domains of possible validity evidence: evidence based on test content (content); the responses of the test administrators and assessed individuals (response process); the structure of item responses or subscales (internal structure); the relations between the test's scores and other variables relevant to the constructs (relations to other variables); and the consequences of testing (consequences). Each of these validity sources indicates possible evidence for or against the application of a test and should guide the selection of a tool with regard to the purpose for which the tool will be used.

For example, if researchers or clinicians plan to evaluate the effectiveness of an intervention to promote reciprocity between depressed parents and their infants, they might search for a tool that comprises evidence to measure the construct of reciprocity, has a rating procedure of reciprocity that is valid, and is sensitive to the change in reciprocity in response to the intervention. Even if all criteria are met, the assessment will be of limited use if an increase in reciprocity does not correspond with a benefit to the assessed family, such as the infant's developmental outcome. Researchers or clinicians might also consider practical arguments for or against the choice of a tool, such as the amount of time needed to conduct and evaluate the assessment, training availability and cost, and the popularity of a tool.

To obtain this critical information, a systematic review of the existing observational tools for measuring parent–infant interaction and an evaluation of the tools' psychometric properties might be particularly helpful. However, to our knowledge, a comprehensive systematic review of the psychometric properties of observational tools of parent–infant interaction has not yet been performed.

Bagner et al. (2012) reviewed assessment procedures for measuring behavioral and emotional problems in infants younger than 2 years. In addition to parent-report measures, the authors identified four observational tools of parent–infant interaction and concluded that the four observational tools provided adequate support for reliability and validity. However, the authors did not report on the eligibility criteria, search strategy, study selection process, or methods used to evaluate the observational tools. Consequently, it remains unclear why the authors identified only four out of the numerous existing observational tools of parent–infant interaction. Halle et al. (2011) conducted a comprehensive search for observational tools of parent–infant interaction, but narrowed their review to relations between the tools' constructs and infant outcomes; other sources of validity evidence were not considered. Finally, the systematic review of Munson and Odom (1996) considered indices of the reliability and validity of observational tools of parent–infant interaction, but is outdated. A more rigorous systematic review that focuses on the psychometric quality of current observational tools of parent–infant interaction is needed.

We systematically searched through the literature of the last hundred years to identify the available observer-rated tools for measuring parent–infant interaction. Our aim was to describe the existing tools and the evidence of their psychometric quality in order to support researchers and clinicians in making evidence-based choices.

Methods

Our systematic review report followed the guidelines published in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA, Moher et al. 2009). We pre-specified a review protocol and registered the systematic review in the PROSPERO database (http://www.crd.york.ac.uk/prospero; registration number CRD42012002435).

Eligibility Criteria for Tool Selection

We included tools that (1) were quantitative measures; (2) were designed for the observation of the interaction between one parent and one infant by an external observer;

and (3) measured both parent and infant behavior. We focused our review on tools that (4) could already be used before the infant's age of 12 months (i.e., tools applicable for infants aged 10–30 months were included). We focused on these tools because problems in the parent–infant interaction should be identified as early as possible before more severe difficulties have been developed (Early Head Start National Resource Center 2013).

We excluded tools that (1) primarily measured constructs other than the observed parent–infant interaction. This exclusion criterion was set because our pilot searches identified many tools that included one single rating of the parent–infant interaction, but focused on other constructs; for these tools, no or very limited evidence for the psychometric properties of the observational constructs existed. Tools were also excluded if (2) they were rated by the parent or (3) no full-text article was available that described or psychometrically evaluated the tool. To prevent the inclusion of minimally used tools, or tools that have been only used with older children, we excluded tools that had been used in (4) fewer than two peer-reviewed primary journal articles with infants aged between 0 and 12 months by different research teams since 1992.

For each tool, two types of articles were considered: (a) articles describing the development or application of the tool (e.g., a description of the tool's theoretical background, scale construction, or implementation) and (b) articles evaluating the tool's psychometric quality (e.g., validation studies of any study type that evaluated the tool's reliability or factorial validity).

Search Strategy

To identify potentially relevant tools and related articles, we conducted a systematic literature search using electronic databases and manual searches of other sources. The search strategy was developed with the assistance of a librarian. Pilot searches revealed that we could not identify most of the available psychometric studies by using methodological search terms. Therefore, we developed a more comprehensive search strategy. We first searched for articles that used or described observational tools for measuring parent–infant interaction. On the basis of these articles, we compiled a list of potentially relevant tools. We screened these tools against our inclusion and exclusion criteria (except exclusion criteria 3 and 4 for the tools; see section on eligibility criteria for tool selection). To locate further information on the relevant tools, we searched the electronic databases using the full titles and tool acronyms.

Databases were searched from each database's first allowable search date through June 2013. No limitations on language or publication year were used. The following databases were searched systematically: Child Care and Early Education Research Connections, Cumulative Index to Nursing and Allied Health Literature (CINAHL), Current Contents Connect (CCC), Cochrane Database of Systematic Reviews (CDSR), Cochrane Central Register of Controlled Trials (CENTRAL), Cochrane Methodology Register (CMR), Database of Abstracts of Reviews of Effects (DARE), Excerpta Medica dataBASE (EMBASE), Education Resources Information Center (ERIC), Educational Testing Service Test Collection (ETS), Medical Literature Analysis and Retrieval System Online (MEDLINE), Mental Measurements Yearbook with Tests in Print, PsycINFO, PSYNDEXplus Literature and Audiovisual Media, PSYNDEXplus Tests, Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI), SocINDEX, Health Technology Assessment Database (HTA), and ProQuest Dissertations & Theses. Google Scholar and Google were also searched. We used combinations of text words and subject headings (if available) related to "parent–infant interaction" and "behavioral observation." As an example, the search strategy for MEDLINE can be found in Appendix 1.

Our manual search covered the electronic directories of twelve test publishers (Behavior Science Systems, Brookes Publishing, Hogrefe, Pearson Assessment & Information, Mind Garden, NCAST Programs, PAR, PRO-ED, Riverside Publishing, Scholastic Testing Service, Slosson Educational Publications, and Western Psychological Services), as well as ten books that described assessments for parents and infants (Clark et al. 2004; DelCarmen-Wiggins and Carter 2004; Goldman et al. 2008; Groth-Marnat 2009; Keyser and Sweetland 1984; Maddox 1997; Shonkoff and Meisels 2000; Snow, Van Hemel, Committee on Developmental Outcomes and Assessments for Young Children, and National Research Council 2008; Zaslow 2011; Zeanah 2009). Additional test publishers and books identified during the search were also hand searched. Published and unpublished reviews of parent–infant tools, including their reference lists that were known by the authors or identified during the literature search, were also manually searched. An additional strategy involved contacting experts in the field to locate further tools.

Tool Selection

Two reviewers (AL and TM) independently performed the eligibility assessment of the tools. If there was disagreement between the two reviewers, a third reviewer was consulted, and the three reviewers reached a consensus for inclusion or exclusion.

Titles and abstracts of the articles that were identified through the database searches were screened first. If the name of the tool was not mentioned in the abstract, the full text of the article was examined. A database was created containing all potentially relevant tools and related sources that described or used the tool. If the identified information about a tool was insufficient to make a decision regarding eligibility

for inclusion, we contacted the authors or publishers ($n = 258$) to obtain further information. The response rate was 71 %. All authors or publishers were asked to provide information about the included tools. Articles in Japanese, Dutch, Italian, Portuguese, and Spanish were translated by professional translators into English or German.

Data Extraction

Data were extracted from articles that psychometrically evaluated or described the included tools. We followed a systematic approach by implementing a standard Excel data extraction sheet, pilot-tested it on twenty randomly selected eligible articles, and refined it accordingly. Two review authors (AL and XL) independently extracted the data from the articles. The reviewers resolved disagreements in data abstraction through consensus. If no agreement could be reached, a decision was made with the help of a third reviewer. Duplicate reports of any of the studies were eliminated from data extraction.

Our general description of the tools was primarily based on data from the tools' manuals. If a manual was not available or the desired information was not reported, we extracted the information from the authors' description of the tool in the identified articles. We extracted the name and purpose of the tool, the tool's target population, the measured constructs, the observational setting, and practical information for administration, such as tool and training availability. As an indicator of the popularity of the tools, we assessed the total number of published and unpublished empirical studies for each tool.

The following information from the articles that described or psychometrically evaluated the included tools was extracted: tool name, citation, study design, country, child age, sample characteristics, sample size, rater characteristics, location of observation, type of task, and type of observation. The unified validity theory of Messick (1995), endorsed by The Standards for Educational and Psychological Testing (American Educational Research Association et al. 1999), guided our extraction approach. In this framework, validity is conceptualized as an integrated evaluation of the degree to which the empirical evidence and theoretical rationales support the intended interpretations of test scores (Messick 1995). The degree of validity evidence is evaluated across multiple sources: test content, response process, internal structure of the test, relations of test scores to other variables, and the consequences of the assessment.

Data on a tool's validity argument were extracted in five validity domains:

(1) Content: (a) theory-driven item development; (b) literature review-based item development; and (c) item review by experts (expert consensus).
(2) Response process: responses of the raters and observed individuals; (a) behavioral response: validity of the observed behavior; (b) rater training; and (c) scoring procedure.
(3) Internal structure: (a) internal consistency: Cronbach's alpha, inter-item correlation, item-discrimination (item-total statistic); (b) reproducibility (objectivity): inter-rater reliability, intra-rater reliability, test–retest reliability, parallel-test reliability; (c) sensitivity to change; (d) dimensional structure: inter-scale correlation, factorial validity, measurement invariance.
(4) Relation to other variables: (a) convergent or divergent validity: correlation between a tool's construct and other measures of the same construct (convergent validity) or the absence of a correlation between a tool's construct and other measures of a different construct (divergent validity); (b) criterion validity: correlation between a tool's construct and the child's developmental outcomes (concurrent, predictive validity); c) discriminant validity: differences in the tool's constructs between relevant subgroups.
(5) Consequences of assessment: four outcome levels corresponding to Kirkpatrick's (1967) hierarchy: (a) reaction: parent's or rater's views on the assessment; (b) learning: parent's or rater's modification of knowledge, skills, or attitudes; (c) behavior: modification in parent's, infant's, or rater's behavior; (d) results: benefits of assessment to parents and infants, or change in health care practice or society.

For each sub-domain (e.g., discriminant validity) of the five validity domains, two reviewers (AL and XL) independently evaluated whether or not the extracted data supported evidence for validity (0 = no evidence, 1 = evidence). If the two reviewers disagreed, a third reviewer was consulted.

Study Quality Rating

The methodological quality of all included empirical studies was critically appraised by two independent review authors (AL and XL) using a standardized form for psychometric reviews (Fallon et al. 2008). Reviewers determined the adequacy of the study design, sample size, sample selection, representativeness of the sample for the tool's target population, tool description, data analysis methods, and blinding as "adequate," "not adequate," "unclear," or "not applicable." On the basis of these criteria, the credibility of each study was globally rated as "unequivocal" (evidence from a study that is of excellent methodological quality, directly relevant to the population of interest, and appropriately uses psychometric tests), "credible" (evidence from a study that has good methodological quality but contains potential flaws, such as a

small sample size, partial divergence from the population of interest, or questionable use of psychometric tests), or "unsupported" (evidence from a study with poor methodological quality, little direct relevance to the population of interest, or inappropriate use of psychometric tests). The inter-rater reliability of the credibility levels of the studies was $K = .87$. Differences in credibility levels between the two raters were reconciled by discussion. Studies with low quality were not excluded from the data synthesis.

Data Synthesis

We used the criteria and guidelines of Cicchetti (1994) and Cicchetti and Sparrow (1990) to evaluate psychological tests for appraising the validity evidence of the included tools. Discrepancies in validity evaluation were resolved by involving a third reviewer. Inter-rater reliability for the validity evidence of each data source ($n = 153$) in the five validity domains between the two reviewers was excellent (content validity $K = .93$, response process $K = .92$, internal structure $K = .90$, relations to other variables $K = .93$, consequences $K = .86$).

Results

Review Process

Figure 1 summarizes the results of the study identification process. After 10,008 duplicates were removed, 13,953 records were screened. We identified 906 potentially relevant tools based on the title and abstract or a full-text review. We excluded more than half of these potentially relevant tools because the tools were utilized in fewer than two peer-reviewed primary journal articles. Typically, these tools were developed for one study and were not subsequently used. After a detailed assessment, we included 24 tools with 104 articles describing or psychometrically evaluating the tools in our review. The full texts of five additional sources describing or evaluating the tools (two conference abstracts, two manuals, and one unpublished report; see Appendix 2) were irretrievable and were excluded from data extraction.

Characteristics of the Articles that Described or Psychometrically Evaluated Observational Tools for Measuring Parent–Infant Interaction

More than half of the included articles were peer-reviewed journal articles (see Table 1). Twenty percent of the articles were commercially published in formats other than peer-reviewed journal articles (book sections, conference abstracts, and tool manuals). Twenty-two percent of the articles were "grey literature" (Schoepfel et al. 2005), in other words non-commercially published or unpublished written articles (conference abstracts, tool manuals, and other unpublished reports).

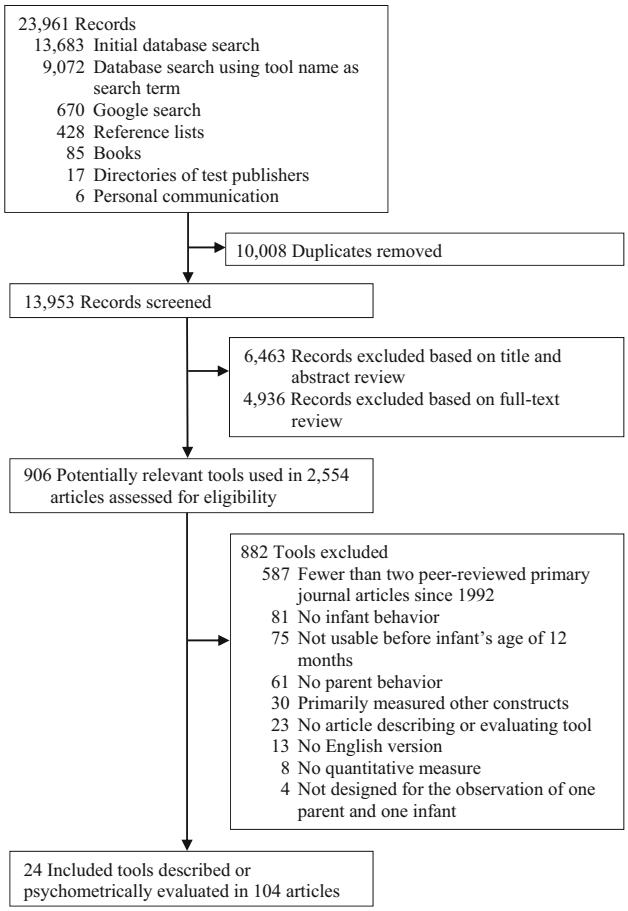

Fig. 1 Flowchart of the review process

The publication dates of the articles ranged from 1975 to 2012. The articles were from 19 countries, with most from the USA. Seventy-five percent of the articles were from North America, 19 % from Europe, 7 % from Asia, 2 % from Australia, and 2 % from South America.

Two-thirds (66 %, $n = 69$) of the included articles reported empirical research studies. In these 69 articles, 86 empirical studies were reported that involved heterogeneous populations, methods, and outcomes (see Supplement A in ESM 1). Two study types (Grimes and Schulz 2002, see Table 2) were most common: a cross-sectional study without a comparison group (descriptive cross-sectional study) and a prospective longitudinal design without a comparison group (descriptive longitudinal study). A few studies (8 %) used a prospective longitudinal design with a comparison group (cohort study). No study had a randomized or non-randomized controlled intervention trial design. Two percent of the studies conducted

Table 1 Characteristics of 104 articles describing or psychometrically evaluating direct observational tools for measuring parent–infant interaction

Characteristics	*n*	%
Type of article		
Peer-reviewed journal article	60	57.7
Conference abstract, poster, or paper	11	10.6
Book section	12	11.5
Tool manual	13	12.5
Master's thesis or dissertation	4	3.8
Other unpublished report	4	3.8
Publication (year)		
1975–1979	2	1.9
1980–1989	18	17.3
1990–1999	36	34.6
2000–2009	39	37.5
2010–2012	9	8.7
Country[a]		
Asia		
Israel	3	2.9
Japan	3	2.9
Taiwan	1	1.0
Australia	2	1.9
Europe		
Austria	1	1.0
Finland	2	1.9
France	1	1.0
Germany	3	2.9
Iceland	1	1.0
Italy	4	3.8
Netherlands	2	1.9
Portugal	1	1.0
Spain	1	1.0
Switzerland	1	1.0
United Kingdom	3	2.9
North America		
Canada	1	1.0
United States	77	74.0
South America		
Brazil	1	1.0
Argentina	1	1.0

[a] Percent add up to more than 100 as two studies included populations of more than one country

Table 2 Characteristics of 86 empirical studies evaluating direct observational tools for measuring parent–infant interaction

Characteristics	*n*	%
Study design		
Cohort	7	8.1
Cross-sectional	9	10.5
Descriptive cross-sectional	36	41.9
Descriptive longitudinal	32	37.2
Not reported[a]	2	2.3
Sample size (dyads)		
1–50	25	29.1
51–100	27	31.4
101–200	19	22.1
201–400	6	7.0
601–800	1	1.2
Database of studies[b]	4	4.7
Not reported	4	4.7
Study quality		
Unequivocal	5	5.8
Credible	29	33.7
Unsupported	52	60.5
Raters		
Student	8	9.3
Researcher	13	15.1
Health professionals	14	16.3
Not reported	55	64.0
Location[c]		
Center	6	7.0
Home	29	33.7
Hospital	14	16.3
Laboratory	21	24.4
Not reported	23	26.7

[a] Studies re-analyzed data from previous studies and the study design of the original studies were not reported

[b] Database of previously conducted studies. The sample size of the original studies were not reported

[c] Percent add up to more than 100 because eight studies used two locations of observation

secondary analyses of previously collected study data. Twenty-nine percent of the studies included sample sizes of 50 parent–infant dyads or fewer. Only 8 % of the studies used more than 200 dyads.

The majority of the studies (60.5 %, $n = 52$) had low methodological quality (see Supplement B and C in ESM 2 and 3). The study quality was most often rated as unsupported because the studies did not provide information on all of our study quality criteria (27.9 %, $n = 24$), reported small sample sizes (7.0 %, $n = 6$), or both (12.8 %, $n = 11$). The remaining studies with low quality (12.8 %, $n = 11$) were judged as unsupported because they showed considerable risk of bias in at least two of the following domains: The study was not well designed ($n = 9$), the sample was not representative of the target population of the tool ($n = 1$), the data analysis methods were not appropriate ($n = 8$), or the raters were not blinded to the tool's objective ($n = 10$).

Most studies did not report the characteristics of the raters of the parent–infant interactions. If reported, the raters were most often health care professionals (e.g., nurses, pediatricians, psychiatrists) or researchers. The observation of the parent–infant interaction was frequently conducted at the homes of the observed families. Observation in a laboratory was also common.

Description of Observational Tools for Measuring Parent–Infant Interaction

The 24 tools measured heterogeneous constructs (see Table 3), which ranged from direct behavior (e.g., gaze) to broad constructs (e.g., maternal sensitivity). Fifteen tools (63 %) assessed constructs of the parent–infant interaction at a dyadic level; four of these tools (17 %) exclusively measured dyadic constructs. The number of items of the tools ranged from 4 to 84. Eleven tools (46 %) contained less than 20 items and could be rated in a short time. Nineteen tools (79 %) used a global rating approach, four tools (17 %) used microanalytic coding, and one tool (4 %) used both microcoding and global ratings to score parent–infant interactions.

The authors often reported multiple purposes (Washington State Office of Superintendent of Public Instruction 2008) of the tools (see Table 4): Fourteen tools (58 %) were appropriate for screening potential problems to be further evaluated; six tools (25 %) aimed to diagnose problems in the parent–infant interaction to determine the need for intervention; twelve tools (50 %) were applicable for monitoring progress of parent–infant interaction over time; and seven tools (29 %) were appropriate to evaluate treatment outcome. For four tools (17 %), no purpose of the tools was stated.

Most authors designed the tools for both low- and high-risk populations ($n = 15$, 63 %). Three tools (13 %) were only appropriate for high-risk populations, and two tools (8 %) were only suitable for low-risk populations. For four tools (17 %), the authors did not specify whether the tool was designed for high-risk or low-risk populations. The entire age range of 0–12 months was covered by fourteen tools (58 %). Fourteen tools (58 %) were applicable for clinical work and research. Seven tools (29 %) were only appropriate for clinical use, and two additional tools (8 %) were only suitable for research use. The authors of one tool did not state the clinical or research use of the tool.

Most tools ($n = 17$, 71 %) could be used for home observations as well as for observations outside the home of the observed family. Six tools (25 %) were designed for settings outside the home, such as a laboratory, center, or hospital. One tool (4 %) was designed for home observation only. Twelve tools (50 %) were appropriate for different observational tasks or settings. The observational task most often used was play or face-to-face interaction ($n = 21$, 88 %). Eight tools (33 %) used feeding observations. Thirteen tools (54 %) required video recording of the parent–infant interaction, seven tools (29 %) could be used for either video or live rating, and four tools (17 %) were designed for live rating only. The recommended time of the behavioral observation ranged from 1 to 45 min. Although the authors frequently reported the tool's observation time, they rarely reported the time needed for scoring the observed behavior. If reported ($n = 8$, 33 %), scoring time ranged from 5 to 50 min.

A commercially published manual was available for six tools (25 %), and an unpublished manual was available for twelve tools (50 %). Eleven (46 %) of these published or unpublished manuals could be obtained from the tools' authors or a commercial test publisher, whereas seven manuals (29 %) were restricted to individuals who attended specific training. For six tools (25 %), no manual was available, but a tool description was published in a peer-reviewed journal article or book section. For most of the tools ($n = 21$, 88 %), authors stated that training was required for the use of the tool. The definition of training ranged from self-study of the manual to personal training by the tool's authors. A certificate of competence (achievement of inter-rater reliability) was needed for the use of five tools (21 %). Personal or distance training was unavailable for five tools (21 %). The number of citations of empirical studies for the tools ranged from 2 to 359 ($Md = 26.0$, $M = 69.5$, $SD = 98.7$). The median number of citations per year since the tool had been developed was 1.1 ($M = 2.7$, $SD = 3.6$, range 0.2–15.0). The EAS (full tool names are reported in Table 3) was the most popular tool, with a yearly citation rate of 15. A tool's popularity was not necessarily a good indicator of how often its psychometric properties were examined. The percentage of articles evaluating the tools out of all empirical studies of a tool ranged from 1.5 to 50 ($Md = 8.8$, $M = 12.8$, $SD = 13.2$).

Validity Evidence Based on Studies with Credible or Unequivocal Study Quality

The validity evidence for the observational tools in the five validity domains, synthesized on the basis of the studies with credible or unequivocal study quality, is reported in Table 5 (for more details, see Supplement B in ESM 2). None of the 24 tools provided evidence for all five domains of validity. Ten tools (42 %) demonstrated evidence for four validity domains (BMIS, EAS, GLOS-R, M-I/TFS, MICS, MRO, NCAFS, NCATS, PIOG, and PIPE).

Content

Evidence for content validity was provided for nine tools (EAS, IPSIC, M-I/TFS, MRO, NCAFS, NCATS, PCERA,

PIOG, and PIPE). For six tools (EAS, MRO, NCAFS, NCATS, PCERA, and PIOG), data indicated that the tools were theoretically based. For five tools (IPSIC, MRO, NCATS, PCERA, and PIPE), there was evidence that the tools were developed on the basis of a literature review. The IPSIC and the M-I/TFS demonstrated evidence that tool development included item reviews by experts.

Response Process

Eighteen tools (ADS, AMIS, BMIS, CARE-Index, DMC, EAS, GLOS-R, GRS, IPSIC, M-I/TFS, MICS, MRO, NCAFS, NCATS, PCERA, PIOG, PIPE, and RCS) demonstrated evidence for valid responses of the raters or observed individuals.

The EAS provided evidence for all three sub-domains of response process validity. Thirteen tools (ADS, AMIS, BMIS, CARE-Index, EAS, GLOS-R, GRS, IPSIC, NCAFS, NCATS, PCERA, PIPE, and RCS) demonstrated evidence for appropriate rater training. Sixteen tools (AMIS, CARE-Index, DMC, EAS, GLOS-R, GRS, IPSIC, M-I/TFS, MICS, MRO, NCAFS, NCATS, PCERA, PIOG, PIPE, and RCS) showed evidence for a valid rating procedure. The EAS and the M-I/TFS provided evidence that the behavioral responses of the observed individuals were valid. Evidence was established by the assessment of parent-reports, indicating that their behavior in the laboratory was representative of their typical interactions at home.

Internal Structure

Evidence of internal structure validity was demonstrated by sixteen tools (ADS, AMIS, Belsky/Unnamed, BMIS, CARE-Index, DMC, EAS, GLOS-R, GRS, MICS, MRO, NCAFS, NCATS, PCERA, PIPE, and RCS). The EAS and the MRO provided evidence for all four sub-domains of this validity domain.

For the sub-domain of internal consistency, ten tools (ADS, Belsky/Unnamed, BMIS, DMC, EAS, MICS, MRO, NCAFS, NCATS, and PCERA) demonstrated evidence, which most commonly was based on the Cronbach's α test statistic. Data on inter-item correlations additionally supported the internal consistency for the MRO.

Evidence in the sub-domain of reproducibility (objectivity) was provided for fifteen tools (ADS, AMIS, Belsky/Unnamed, BMIS, CARE-Index, DMC, EAS, GLOS-R, MICS, MRO, NCAFS, NCATS, PCERA, PIPE, and RCS). The most commonly reported test statistic was inter-rater reliability, for which the data of fourteen tools (ADS, AMIS, Belsky/Unnamed, BMIS, CARE-Index, DMC, EAS, MICS, MRO, NCAFS, NCATS, PCERA, PIPE, and RCS) established evidence. Only the NCATS demonstrated evidence of intra-rater reliability. For seven tools (Belsky/Unnamed, BMIS, CARE-Index, EAS, GLOS-R, MRO, and RCS), evidence of test–retest validity was demonstrated.

Evidence for the sub-domain of sensitivity to change in response to the infant's development or a parent–infant intervention was provided for five tools (Belsky/Unnamed, BMIS, EAS, MRO, and RCS).

In the sub-domain of dimensional structure, data of eight tools (ADS, CARE-Index, EAS, GLOS-R, GRS, MRO, PCERA, and RCS) established evidence. Data for six tools (ADS, CARE-Index, EAS, GRS, PCERA, and RCS) demonstrated within-tool inter-scale correlations in the expected directions (scales with similar constructs correlated higher than scales with dissimilar constructs). The GLOS-R, MRO, and PCERA provided evidence for factorial validity. Data supporting measurement invariance was available for the MRO.

Relations to Other Variables

The data of fifteen tools (Belsky/Unnamed, BMIS, CARE-Index, DMC, EAS, GLOS-R, GRS, M-I/TFS, MICS, MRO, NCAFS, NCATS, PIOG, PIPE, and RCS) demonstrated validity evidence. The BMIS and the PIOG provided evidence on all three sub-domains of the domain relations to other variables (convergent or divergent validity, discriminant validity, and criterion validity).

Evidence for convergent or divergent validity (association of the tools' constructs with similar or dissimilar constructs) was found for seven tools (BMIS, DMC, MICS, MRO, NCAFS, NCATS, and PIOG). Fourteen tools (Belsky/Unnamed, BMIS, CARE-Index, DMC, EAS, GLOS-R, GRS, M-I/TFS, MRO, NCAFS, NCATS, PIOG, PIPE, and RCS) discriminated between relevant groups (often high-risk vs. low-risk groups). Four tools (BMIS, EAS, GLOS-R, and PIOG) demonstrated evidence for concurrent criterion validity; four tools (BMIS, EAS, GLOS-R, and RCS) showed evidence for predicting a criterion (e.g., the infant's attachment security or development).

Consequences of Assessment

Five tools (BMIS, GLOS-R, M-I/TFS, MICS, and PIOG) provided validity evidence for the consequences of the assessment of a tool.

The BMIS and the MICS demonstrated evidence based on the reaction of raters in response to the tools' use, which was generally positive. For the BMIS and the PIOG, learning in terms of an increase in the raters' knowledge and skills in response to the assessment was reported. For none of the tools, data suggested evidence for modifications in rater behavior in response to the tool's use. Data from the BMIS, GLOS-R, and M-I/TFS presented evidence that the assessment had benefits for the evaluated parents or infants.

Evidence was mostly established by reporting norms or cut-off scores to guide treatment decisions. However, the samples on which norms were based could often not be considered fully representative of the target population.

Further Validity Evidence Based on Studies with Unsupported Study Quality, or Based on Articles Describing an Observational Tool

Additional validity evidence for the included observational tools of parent–infant interaction, based on studies with unsupported study quality, or based on articles describing a tool, is summarized in Table 6 (for details, see Supplement C and D in ESM 3 and 4).

Content

Evidence of content validity was provided by fifteen tools (ADS, AMIS, CARE-Index, CIB, DMC, FEAS, GLOS-R, MICS, MIPIS, Monadic Phases, NCAFS, NCATS, PIIS, PIOG, and RCS).

For eleven tools (ADS, CARE-Index, CIB, DMC, FEAS, GLOS-R, MICS, MIPIS, Monadic Phases, PIIS, and RCS), authors reported that the tools were theoretically based; for seven tools (ADS, AMIS, FEAS, MICS, NCAFS, PIIS, and PIOG), authors stated that the tools were developed on the basis of a literature review. For four tools (NCAFS, NCATS, PIOG, and RCS), experts reviewed the tools' items during the tools' development.

Response Process

Eleven tools (ADS, AMIS, CARE-Index, EAS, FEAS, M-I/TFS, Monadic Phases, NCAFS, NCATS, PIOG, and RCS) provided evidence for the validity of the response process.

Evidence for appropriate rater training was demonstrated by seven tools (CARE-Index, EAS, FEAS, M-I/TFS, NCAFS, NCATS, and PIOG). Evidence for a valid rating procedure was available for nine tools (ADS, AMIS, EAS, M-I/TFS, Monadic Phases, NCAFS, NCATS, PIOG, and RCS). The NCATS provided evidence that the behavioral responses of the observed individuals were valid.

Internal Structure

Based on studies with unsupported study quality, evidence of internal structure validity was found for fifteen tools (AMIS, BMIS, CIB, DMC, EAS, GLOS-R, M-I/TFS, MICS, MIPIS, Monadic Phases, NCAFS, NCATS, PCERA, PIIS, and RCS).

For the sub-domain of internal consistency, data of four tools (CIB, MIPIS, NCAFS, and NCATS) showed evidence. In the sub-domain of reproducibility (objectivity), evidence was demonstrated for eleven tools (BMIS, EAS, GLOS-R, M-I/TFS, MIPIS, Monadic Phases, NCAFS, NCATS, PCERA, PIIS, and RCS). Four tools (AMIS, DMC, MICS, and NCATS) provided evidence for the sub-domain of sensitivity to change in response to the infant's development or a parent–infant intervention. For the sub-domain of dimensional structure, the NCAFS and NCATS established evidence.

Relations to Other Variables

In the domain of relations to other variables, the data for nine tools (AMIS, BMIS, DMC, FEAS, MIPIS, Monadic Phases, NCAFS, NCATS, and PIIS) demonstrated validity evidence.

Evidence for convergent or divergent validity was found for four tools (Monadic Phases, NCAFS, NCATS, and PIIS). Five tools (BMIS, DMC, FEAS, NCAFS, and NCATS) discriminated between relevant groups. Evidence for criterion validity was demonstrated by four tools (AMIS, MIPIS, NCAFS, and NCATS).

Consequences of Assessment

Validity evidence for the consequences of the assessment of a tool was constituted by eight tools (ADS, BMIS, DMC, FEAS, M-I/TFS, NCAFS, NCATS, and PIOG).

Four tools (BMIS, NCAFS, NCATS, and PIOG) established evidence based on the reaction of raters in response to the tool's use. An increase in the raters' knowledge, skills, or attitudes in response to the assessment (learning) was reported for five tools (ADS, BMIS, NCAFS, NCATS, and PIOG). The data of four tools (BMIS, NCAFS, NCATS, and PIOG) suggested evidence for modifications in rater behavior in response to the tool's use. Evidence that the assessment had benefits for the evaluated parents or infants was reported for six tools (ADS, DMC, FEAS, M-I/TFS, NCAFS, and NCATS).

Discussion

We systematically reviewed the literature to identify observational tools for measuring parent–infant interaction, described the existing tools, and evaluated their psychometric soundness. We identified 24 observational tools for assessing parent–infant interaction that were described or psychometrically evaluated in 104 articles from 19 countries. Evidence in all five domains of validity (i.e., content, response process, internal structure, relations to other variables, and consequences) based on studies with credible study quality was identified for none of the tools. Ten observational tools provided evidence in four domains of validity; most of these tools have the potential for wider use with additional research on their consequential validity.

Aspects of Validity that Need Improvement

Content validity was most often demonstrated by a theory-driven or literature review-based development of item content. In contrast, content validity evidence based on item content reviews by experts was rarely reported and should be conducted more often in future research. Evidence on response process validity was often consolidated by the training of raters, whereas evidence indicating the validity of the observed behavior was minimally or not described. Because parent–infant interaction is often observed in non-naturalistic settings, the observed behavior might not represent the family's typical daily behavior at home (Gardner 2000). Further studies should attend more to the validity of the observed behavior of parents and infants.

In the validity domain of internal structure, authors typically reported and established adequate levels of internal consistency and inter-rater reliability. On the basis of the empirical studies with at least credible study quality, only three tools provided evidence for factorial validity. The theoretically assumed dimensionality of the remaining tools should be confirmed using factor analysis. Validity based on relations to other variables was often demonstrated by evidence that the tool's constructs correlated with similar constructs or discriminated between high- and low-risk groups. Whether the tool's scores prospectively predicted child developmental outcomes was less often confirmed. Rather than examining whether parent–infant interaction constructs discriminate between subgroups, further studies could evaluate whether these constructs are related to important child outcomes.

The consequences of the behavioral assessment were rarely reported. Few tools demonstrated that the assessment had benefits for the parent or the infant. In particular, more information is needed on the interpretability of scores by providing norms and cut-off scores, including the assessment of the specificity and sensitivity of a tool to identify infants who are at risk for unfavorable mental health outcomes. Without such information, a tool might be of limited value for diagnostic, prognostic, and treatment indications.

Steps to Improve the Quality of Parent–Infant Interaction Research

Our findings suggest further steps to improve the quality of research in the area of parent–infant interaction. Researchers and clinicians should use tools with evidence of validity. The existing research does not always follow this recommendation; we found more than 500 tools that were only used in one or no peer-reviewed journal article. The use and refinement of an established tool with evidence of validity might be a better investment of time and resources than the ongoing development of new assessment tools.

A more thorough understanding is needed of which of the wide range of heterogeneous constructs are causally related to child developmental outcomes. At present, the prediction of an infant's later developmental outcomes is often based on a tool's overall composite. At the level of the measured constructs, less support exists that constructs causally influence developmental outcomes (Mahoney et al. 1996). Constructs or items that are not related to the child's developmental outcome could be eliminated; this may be especially important for tools with an excessive amount of items (four tools included more than 60 items). To enable precise predictions of the child's later development, studies could specify which constructs relate to which domain of infant developmental outcomes (e.g., attachment, social–emotional development, or cognitive development).

Observational tools for measuring parent–infant interaction often lacked a user manual, and if available, the manuals often did not contain information on the tool's validity evidence. Tools would benefit from the development of user manuals with clear guidelines on scoring and interpretation. It would also be helpful if the purpose and target group of the tool were stated more clearly. Information about the tool's psychometric evidence, norms, and cut-off values would enable appropriate use. As is common practice in the field of questionnaire assessments, it would be helpful for tool administrators if tool manuals were available from commercial test publishers or authors before training is attended so that individuals interested in a measure could obtain critical information about a tool.

To date, none of the tools has been thoroughly validated with paternal samples. Thus, the tools cannot be recommended for use with fathers. Only one tool included a pure father sample in the initial validation study. The adaption and validation of observational tools for father–infant interaction remains an imperative avenue of research because fathers have become more involved in child care in Western cultures compared to past generations (Pleck and Pleck 1997). Further studies should clarify the domains in which father–infant interaction differs from mother–infant interaction. Current reports indicate conflicting results (Aksan et al. 2006; Harrison et al. 1999; Nakamura et al. 2000).

The cultural applicability of observational tools for assessing parent–infant interaction is another central field of future research. Most parent–infant interaction research included in this review was based on samples from North America and Western Europe. As culture may influence parent–infant interaction (Bornstein et al. 2012), further studies might clarify whether the items of the observational tools for measuring parent–infant interaction represent similar interactional constructs across cultures.

Research on parent–infant interaction needs improved standards in the design and report of studies. Most studies

Table 3 Scoring format and constructs of direct observational tools for measuring parent–infant interaction

Tool name (citation)	Scoring format	Number of scales and summary scores	Total number of items	Parent constructs (number of items)	Infant constructs (number of items)	Dyadic constructs (number of items)
Emotional Availability Scales (EAS, Biringen 2008)	3-point or 7-point global rating (1–3 or 1–7; higher values indicate more optimal behavior)	4 Caregiver scales 2 Infant scales 1 Total score 1 Global rating	42	Sensitivity (7) Structuring (7) Nonintrusiveness (7) Nonhostility (7)	Responsiveness to adult (7) Involvement to adult (7)	N/A
Coding Interactive Behavior (CIB, Feldman 1998)	5-point global rating (1 = low, 5 = maximum level of behavior)	3 Mother scales 2 Infant scales 2 Dyadic scales No total score	43	Sensitivity Intrusiveness Limit-setting (22 items in total)	Social involvement Negative emotionality (16 items in total)	Dyadic reciprocity Dyadic negative states (5 items in total) Additional codes and constructs for play and feeding
Child–Adult Relationship Experimental Index (CARE-Index, Crittenden 2006)	3-point global rating (0–2 or 0–14; higher values indicate more optimal behavior)	3 Caregiver scales 4 Infant scales 1 Dyadic synchrony rating	15	Sensitivity Control Unresponsiveness (7 items that can be assigned to each of the scales)	Cooperativeness Difficultness Compulsiveness Passivity (7 items that can be assigned to each of the scales)	Dyadic synchrony (1)
Nursing Child Assessment Feeding Scales (NCAFS, Sumner and Spietz 1994)	Dichotomous global rating (yes or no)	4 Parent scales 2 Infant scales 1 Parent total score 1 Infant total score 1 Total score 1 Infant contingency score 1 Parent contingency score 1 Contingency score	76	Sensitivity to cues (16) Response to the child's distress (11) Social–emotional growth fostering (14) Cognitive growth fostering (9)	Clarity of cues (15) Responsiveness to caregiver (11)	N/A
Nursing Child Assessment Teaching Scales (NCATS, Oxford and Findlay 2012)	Dichotomous global rating (yes or no)	4 Parent scales 2 Infant scales 1 Parent total score 1 Infant total score 1 Total score 1 Infant contingency score 1 Parent contingency score 1 Contingency score	73	Sensitivity to cues (11) Response to the child's distress (11) Social-emotional growth fostering (11) Cognitive growth fostering (17)	Clarity of cues (10) Responsiveness to caregiver (13)	N/A

Table 3 continued

Tool name (citation)	Scoring format	Number of scales and summary scores	Total number of items	Parent constructs (number of items)	Infant constructs (number of items)	Dyadic constructs (number of items)
Parent–Infant Observation Guide (PIOG, Bernstein et al. 1992)	Dichotomous global rating (present or absent)	3–4 Parent scales 3–4 Infant scales No total score	27–36	Actions to meet child's needs and wants (5) Responding to child's activity and interests (5) Positive feelings shown to child (5) Helping child learn new skills and language (5, only for 4–12 months)	Expression of needs and wants (Clarity of signals) (4) Using parent's help (4) Involvement with parent (4) Positive feelings shown to parent (4, only for 4–12 months)	N/A
Parent–Infant Interaction Scale (PIIS, Clark and Seifer 1985)	5-point global rating (1 = poor, 5 = excellent)	7 Mother scales 2 Infant scales 1 Dyadic scale 3 Summary scores (Interaction style, Social reference, Interaction context)	10	Acknowledging (1) Imitating (1) Expanding/Elaborating (1) Parent direction of gaze (1) Parent affect (1) Forcing (1) Overriding (1)	Child social referencing (1) Child gaze aversion (1)	Dyadic reciprocity (1)
Dyadic Mutuality Code (DMC, Censullo 2004)	Dichotomous global rating (1 = absent, 2 = present)	2 Mother scales 1 Infant scale 3 Dyadic scales 1 Total score	6	Maternal sensitive responsiveness (1) Maternal pauses (1)	Infant clarity of cues (1)	Mutual attention (1) Positive affect (1) Turn-taking (1)
Parent–Child Early Relational Assessment (PCERA, Clark 1985)[a]	5-point global rating (1 = area of concern, 5 = area of strength)	3 Parent scales 3 Infant scales 2 Dyadic scales No total score	57	Positive affective involvement and verbalization (11) Negative affect and behavior (5) Intrusiveness, Insensitivity, Inconsistency (8)	Positive affect, communicative and social skills (8) Quality of play, interest, and attentional skills (10) Dysregulation and irritability (6)	Mutuality and reciprocity (4) Disorganization and tension (5)
Infant–Parent Social Interaction Code (IPSIC, Baird et al. 1992)	15-s time sampling Dichotomous rating (presence or absence) 3-point global rating (high, middle, low; higher values indicate higher frequency of behavior)	4 Mother scales 4 Infant scales 1 Dyadic scale No total score	9	Response contingency (1) Directiveness (1) Intrusiveness (1) Facilitation (1)	Initiation (1) Participation (1) Signal clarity (1) Intentional communicative acts (1)	Theme continuity (1)
Belsky/Unnamed (Belsky et al. 1984)	15-s time sampling Dichotomous rating (presence or absence)	2 Mother scales 2 Dyadic scales No total score	15	Non-involvement (2) Basic care (2)	N/A	Reciprocal interaction (9) Distress (2)

Table 3 continued

Tool name (citation)	Scoring format	Number of scales and summary scores	Total number of items	Parent constructs (number of items)	Infant constructs (number of items)	Dyadic constructs (number of items)
Assessment of Mother–Infant Sensitivity (AMIS, Price 1983)	5-point global rating (1–5; higher values indicate greater sensitivity)	15 Mother scales 7 Infant scales 3 Dyadic scales 3 Summary scales (Holding/handling, Social/affective, Feeding/caregiving) 1 Total score	25	Spatial distance (1) Holding Style (1) Predominant maternal mood/affect (1) Verbalization (tone) (1) Verbalization (content) (1) Visual interaction behavior (1) Modulation of distress episodes (1) Caregiving style (1) Stimulation of infant (1) Response to changing levels of infant activity (1) Burping style (1) Stimulation to feed (1) Manner of stimulation to feed (1) Frequency of stimulation to feed (1) Response to infant satiation (1)	Predominant infant state (1) Predominant infant mood/affect (1) Vocalizations (1) Distress (1) Visual behavior (1) Posture (1) Response to stimulation to feed at satiation (1)	Synchrony in response to pleasurable affect (1) Regulation of feeding at initiation (1) Regulation of feeding at termination (1)
Mother–Infant Play Interaction Scale (MIPIS, Walker and Thompson 1982)	5-point global rating (1–5; higher values indicate higher responsivity)	1 Mother scale 1 Infant scale 1 Dyadic scale 1 Total score	16	Maternal scale (10) (Holding style, expression of affect, expression of affect—quality of contingency to infant, caregiving style, visual interaction, style of play, vocalization style—general, vocalization style—quantity of contingency, attempts at smile elicitation, kinesthetic quality of interaction)	Infant scale (3) (predominant infant wakeful response level, predominant infant mood/affect, visual interaction)	Dyadic scale (3) (Overall dyadic quality of interaction, synchrony of affect, termination of interaction)
Mother–Infant Communication Screening (MICS, Raack 1989)	5-point global rating (1–5; higher values indicate more optimal interaction)	4 Mother scales 1 Dyadic scale 1 Total score	24	Distress (3) Feeding (5) Play/neutral state (4) Rest (4)	N/A	Language and synchrony (8)
Global Rating Scales of Mother–Infant Interaction (GRS, Fiori-Cowley et al. 2000)	5-point global rating (1 = poor, 5 = good)	3–4 Mother scales 2 Infant scales 1 Dyadic scale 1 Total score	25–28	Good to poor (5) Intrusive to remote (4) Signs of depression (4) Additional items for infants aged 4–5 months (3)	Good to poor (3) Inert to fretful (4)	Good to poor (5)

Table 3 continued

Tool name (citation)	Scoring format	Number of scales and summary scores	Total number of items	Parent constructs (number of items)	Infant constructs (number of items)	Dyadic constructs (number of items)
Greenspan–Lieberman Observational Scale-Revised (GLOS-R, Greenspan and Lieberman 1989)	15-s time sampling (event sampling for newborn version) Dichotomous rating (presence or absence)	No scales No total score (but factor analytically derived subscales by Hofheimer and Appelbaum 1992)	78	Somatic patterning/homeostasis (7) (Direct physical punishment, physical overstimulation, abrupt handling, rough-and-tumble, rhythmic behavior, affectionate physical behavior) Consoling, soothing Attachment (7) (Facilitation of interpersonal exchanges, pleasure, guarded "as if" pleasure, flat affect, non-participating but available, non-participating withdrawn, aversively intrusive) Somato-psychological differentiation (18) (Contingent response (6 sub-items), non-contingent behavior (6 sub-items), anticontingent behavior (6 sub-items)) Initiative and organization (15) (Chains of contingent exchanges (6 sub-items), chains of anticontingent exchanges (6 sub-items), developmentally facilitating behavior, facilitation of involvement, interfering or undermining)	Somatic patterning/homeostasis and attachment (12) (Resisting physical contact, physical overstimulation/distress, seeking proximity, seeking physical contact, affectionate physical behavior, avoidance of physical contact, initiation of social interaction, interrupts social interaction or physical contact, aggressive behavior, pleasure, distress, flat affect) Somato-psychological differentiation (15) (Contingent response (5 sub-items), non-contingent response (5 sub-items), anticontingent response (5 sub-items)) Initiative and organization (4) (Onlooking behavior, orientation to inanimate environment, exploratory manipulation, exploratory roaming)	N/A
Functional Emotional Assessment Scale (FEAS, Greenspan et al. 2001)	3-point global rating (0 = behavior not at all or very briefly observed, 2 = behavior consistently present, observed many times)	3–4 Parent scales 2–3 Infant scales 1 Total parent score 1 Total child score 1 Total score	27–47	Self-regulation (6; 5 clinical version) Attachment (5; 3–4 clinical version) Two-way communication (6; 5 clinical version) Behavioral organization (7; 3 clinical version; only for 10–12 months)	Self-regulation (9; 9 clinical version) Attachment (9; 5 clinical version) Two-way communication (5; 4 clinical version; only for 10–12 months)	N/A

Table 3 continued

Tool name (citation)	Scoring format	Number of scales and summary scores	Total number of items	Parent constructs (number of items)	Infant constructs (number of items)	Dyadic constructs (number of items)
Bethlem Mother–Infant Interaction Scale (BMIS, Kumar and Hipwell 1996)	5-point global rating (0 = appropriate, sensitive, and well organized, 4 = very severe disturbances resulting in physical separation of the child	6 Mother scales 1 Infant scale 1 Dialog score 1 Total score	7	Eye contact (1) Physical contact (1) Vocal contact (1) Mother´s Mood (1) General routine (1) Risk to the baby (1)	Baby´s contribution to interaction (1)	N/A
Massie-Campbell Scale of Mother–Infant Attachment Indicators During Stress (ADS, Massie and Campbell 1992)	5-point global rating (1 = avoid contact or are not responsive, 5 = overanxious behavior, intense attachment or unusually strong reaction to stress)	7 Mother scales 7 Infant scales No total score	14	Gazing (1) Vocalizing (1) Touching (1) Response to touch (1) Holding (1) Affect (1) Proximity (1)	Gazing (1) Vocalizing (1) Touching (1) Response to touch (1) Holding (1) Affect (1) Proximity (1)	N/A
Monadic Phases (Tronick et al. 1982)	1-s time sampling Dichotomous rating (presence or absence)	7 Mother scales 10 Infant scales	84	Avert (6) Social attend (neutral engagement) (3) Social elicit (4) Object attend (7) Play (positive engagement) (5) Object play (7) Talk (2)	Protest (9) Avert (4) Social attend (8) Social play (6) Object attend (8) Object play (6) Talk (5) Pick-me-up (2) Greet (1) Positive away (1)	N/A
Mother–Infant/Toddler Feeding Scale, also named Observational Scale for Mother–infant interaction during Feeding or Chatoor Feeding Scale (M-I/TFS, Chatoor et al. 1997)	4-point global rating (0–3; higher values indicate higher amounts of positive or negative behavior)	5 Dyadic scales No total score	46	N/A	N/A	Dyadic reciprocity (16) Dyadic conflict (12) Talk and distraction (4) Struggle for control (7) Maternal non-contingency (7)

Table 3 continued

Tool name (citation)	Scoring format	Number of scales and summary scores	Total number of items	Parent constructs (number of items)	Infant constructs (number of items)	Dyadic constructs (number of items)
Mutually Responsive Orientation (MRO, Aksan et al. 2006)	5-point global rating (1–5; higher values indicate higher MRO)	4 Dyadic scales 1 Total score	16–17	N/A	N/A	Coordinated routines (2) Harmonious communication (4) Mutual cooperation (4–5) Emotional ambiance (6)
Revised Relational Coding System (RCS, Fogel et al. 2003)	2-s (minimum) event sampling Dichotomous rating (presence or absence)	5 Dyadic scales No total score	9	N/A	N/A	Symmetrical (2) Asymmetrical (2) Unilateral (3) Disruption (1) Unengaged (1)
Pediatric Infant Parent Exam (PIPE, Fiese et al. 2001)	6-point or 7-point global rating (0–5 or 1–7; higher scores indicate increase of interaction problems)	3 Dyadic scales 1 Total score 1 Global rating	4	N/A	N/A	Easy engagement vs. inappropriate and bizarre engagement (1) Easy playfulness vs. inappropriately play (1) Gradual cool down vs. unable to stop game (1) Overall impression of interaction (adaptive vs. maladaptive, 1)

Tools are ordered according to the tools' similarity of the addressed constructs

A range of item or scale numbers (i.e., 0–5) indicate that different numbers of items or scales are used for different age ranges of the infant

N/A not applicable

[a] Scale description is based on Clark (1999) because the tool's manual was not available

Table 4 Characteristics of direct observational tools for measuring parent–infant interaction

Tool name	Purpose	Clinical or research use	Target population	Age of child (months)	Live or video	Location	Task	Time to administer (min)	Availability	Required qualifications	Training availability	Total number of studies	Number of studies per year
EAS	Screen Monitor Outcome	Clinical Research	LR and HR caregivers LR and HR children	0–14 years[a]	Video	Home Laboratory	Any task, stress suggested for short observations	Observation 20–30	Published manual[b] Published article	Training and certificate	Personal and distance training	359	15.0
CIB	Screen Monitor Outcome	Clinical Research	LR and HR caregivers LR and HR children	0–13 years[a]	Video	Home Laboratory	Free play Feeding Caregiving Problem-solving task Book reading	Not reported	Unpublished manual[b] Published article	Training and certificate	Personal training	65	4.3
CARE-Index	Screen Diagnose Monitor Outcome	Clinical Research	LR and HR caregivers LR and HR children	0–30[a]	Video	Center Home Hospital Laboratory	Free play	Observation 3–5 (at least 2) Scoring 10–15	Unpublished manual[b] Published article	Knowledge[c] Training and certificate	Personal training	137	4.4
NCAFS	Screen Monitor Outcome	Clinical Research	LR and HR caregivers LR and HR infants	0–12	Live Video	Home Laboratory	Feeding	Observation time = feeding time	Published manual[b] Book section	Practitioners Non-practitioners Training and certificate	Personal and distance training	182	5.4
NCATS	Screen Monitor Outcome	Clinical Research	LR and HR caregivers LR and HR children	0–36	Live Video	Home Laboratory	Teaching age appropriate task	Observation 1–5	Published manual[b] Book section	Practitioners Non-practitioners Training and certificate	Personal and distance training	321	9.4
PIOG	Screen Monitor	Clinical	HR parents Children	0–8 years[a]	Live	Home Laboratory	Feeding Play (free play, game play, toy play) Caregiving (diaper change, meal time, bathing)	Observation and scoring 10	Unpublished manual Published article	Practitioners Non-practitioners with knowledge[c] Training	Not available	20	0.8
PIIS	Diagnose	Clinical	Parents HR infants	0–18	Video	Center	Free play	Observation 6–8	Published article	Practitioners Non-practitioners Training	Not available	8	0.3
DMC	Not reported	Clinical Research	LR and HR mothers LR and HR infants	0–6	Live Video	Laboratory	Face-to-face interaction	Observation 5 Scoring 5	Unpublished manual Published article	Practitioners Researchers Training	Personal training	12	0.5

Table 4 continued

Tool name	Purpose	Clinical or research use	Target population	Age of child (months)	Live or video	Location	Task	Time to administer (min)	Availability	Required qualifications	Training availability	Total number of studies	Number of studies per year
PCERA	Diagnose Monitor Outcome	Clinical Research	LR and HR caregivers LR and HR children	0–60	Video	Center Home	Feeding Age appropriate task (toy play, book reading) Free play Separation/reunion	Observation 20	Unpublished manual[b] Published article	Practitioners Training	Personal training	193	6.9
IPSIC	Diagnose	Not reported	LR and HR mothers LR and HR children	0–36	Video	Home Laboratory	Free play	Observation 10 Scoring 15–50 (10–35 with routine)	Published article	Training	Distance training	5	0.3
Belsky/ Unnamed	Not reported	Clinical Research	Mothers Infants	0–9	Live	Home Laboratory	Naturalistic contexts	Observation 45 Scoring time = observation time	Published article	Training	Not available	21	0.8
AMIS	Screen	Clinical	Mothers Infants	0–3	Video	Home Laboratory	Feeding Play Diapering	Observation 15–30	Published article	Not reported	Not reported	22	0.6
MIPIS	Monitor	Clinical Research	LR mothers LR infants	1–1.5	Live Video	Home	Free play	Observation 5 Scoring 5–10	Book section	Training	Not available	11	0.4
MICS	Screen	Clinical	LR and HR mothers LR and HR infants	0–12	Live Video	Center Home Hospital	Health encounter Interview Caregiving (feeding) Rest Distress Free play Neutral state	Observation 10	Published manual Published article	Practitioners Training	Distance training (self-study of manual and guided instruction)	7	0.3
GRS	Screen	Clinical Research	LR and HR mothers Infants	2–5	Video	Home Laboratory	Face-to-face interaction	Observation 5 (at least 3)	Unpublished manual	Training	Personal training	56	3.5
GLOS-R	Screen	Clinical Research	HR mothers LR and HR children	0–48[a]	Video	Laboratory	Free play	Observation 8–12	Unpublished manual Book section	Training	Personal training	20	0.6
FEAS	Screen	Clinical Research	LR and HR caregivers LR and HR children	7–48[a]	Video Live	Home Hospital	Toy play (symbolic play, tactile play and movement play)	Administration 15–30 Observation 15	Published manual (book) Published article	Practitioners Training	Personal training	46	2.3

Table 4 continued

Tool name	Purpose	Clinical or research use	Target population	Age of child (months)	Live or video	Location	Task	Time to administer (min)	Availability	Required qualifications	Training availability	Total number of studies	Number of studies per year
BMIS	Diagnose Monitor	Clinical	HR mothers Infants	0–12	Live Video	Hospital	Clinical routine Feeding Free play	Not reported	Published article	Practitioners	Not available	12	0.5
ADS	Screen Monitor Outcome	Clinical	LR and HR parents LR and HR children	0–18	Live	Center Home Hospital	Any task (mild to moderate stress suggested)	Observation 6	Published manual Published article Website	Practitioners Training	Self-study of manual	30	1.0
Monadic Phases	Not reported	Research	Mothers Infants	0–6	Video	Laboratory	Face-to-face interaction	Observation 3–10 Scoring 'long'	Unpublished manual Published article	Training	Personal training	58	1.8
M-I/TFS	Diagnose Monitor	Clinical Research	Caregivers LR and HR children	1–36[a]	Video	Laboratory	Feeding	Observation 20	Unpublished manual[b] Published article	Training	Personal and distance training	35	1.3
MRO	Not reported	Clinical Research	LR parents LR infants	7–15	Video	Home Laboratory	Caregiving (meal time) Free play Toy play Opening a gift	Observation 45 (6 tasks)	Unpublished manual Published article	No training necessary	Self-study of manual	8	1.3
RCS	Screen Monitor	Research	Caregivers Children	Any age	Video	Home Laboratory	Free play Any task	Not reported	Unpublished manual Book section	Practitioners Training	Distance training	37	1.9
PIPE	Screen	Clinical	Caregivers LR and HR infants	4–12	Live	Center Home Hospital Laboratory	Brief game without toys (peekaboo)	Observation 1–5 Preparation 'short' Scoring 'short'	Unpublished manual Published article	Practitioners Non-practitioners with knowledge[c] Training	Personal training	2	0.2

Tools are ordered according to the tools' similarity of the addressed constructs

Number of studies = number of published and unpublished empirical studies which used the tool; screen = to identify potential problems which have to be further evaluated; diagnose = to identify problems to determine intervention; monitor = to monitor progress of the parent–infant interaction over time; outcome = to evaluate treatment outcomes on a group level; LR = low risk; HR = high risk; live or video = live or video observation; location = location of observation; published article = published journal article describing the tool; practitioners = practitioners in related professions, e.g., pediatricians, child psychologists, pediatric nurses

[a] Different versions for different age groups

[b] Manual restricted to persons who attended training

[c] Knowledge of parent–infant interaction or infant development

Table 5 Validity evidence of direct observational tools for measuring parent–infant interaction based on studies with credible or unequivocal study quality

Tool name	No. of validity domains	No. of validity sub-domains	Content			Response			Internal structure			
			Theoretical basis	Literature review	Expert consensus	Rater training	Rating procedure	Behavioral response	Internal consistency	Reproducibility	Sensitivity to change	Dimensional structure
ADS	2	4	0	0	0	1	0	0	1	1	0	1
AMIS	2	3	0	0	0	1	1	0	0	1	0	0
Belsky/ Unnamed	2	4	0	0	0	0	0	0	3	2	1	0
BMIS	4	10	0	0	0	1	0	0	3	2	2	0
CARE-Index	3	5	0	0	0	2	1	0	0	3	0	2
CIB	0	0	0	0	0	0	0	0	0	0	0	0
DMC	3	5	0	0	0	0	1	0	1	1	0	0
EAS	4	10	1	0	0	5	10	4	1	9	3	3
FEAS	0	0	0	0	0	0	0	0	0	0	0	0
GLOS-R	4	7	0	0	0	1	1	0	0	2	0	4
GRS	3	4	0	0	0	1	2	0	0	0	0	1
IPSIC	2	4	0	1	1	2	2	0	0	0	0	0
M-I/TFS	4	5	0	0	1	0	2	1	0	0	0	0
MICS	4	5	0	0	0	0	1	0	1	1	0	0
MIPIS	0	0	0	0	0	0	0	0	0	0	0	0
Monadic Phases	0	0	0	0	0	0	0	0	0	0	0	0
MRO	4	9	1	1	0	0	1	0	4	3	1	5
NCAFS	4	7	1	0	0	2	1	0	1	1	0	0
NCATS	4	8	1	1	0	4	2	0	2	3	0	0
PCERA	3	7	1	1	0	1	2	0	1	2	0	3
PIIS	0	0	0	0	0	0	0	0	0	0	0	0
PIOG	4	6	1	0	0	0	1	0	0	0	0	0
PIPE	4	5	0	1	0	1	1	0	0	1	0	0
RCS	3	7	0	0	0	4	2	0	0	3	1	4

Table 5 continued

Tool name	No. of validity domains	No. of validity sub-domains	Relations to other variables			Consequences			
			Convergent Divergent	Discriminant	Criterion	Reaction	Learning	Behavior	Results
ADS	2	4	0	0	0	0	0	0	0
AMIS	2	3	0	0	0	0	0	0	0
Belsky/ Unnamed	2	4	0	1	0	0	0	0	0
BMIS	4	10	1	2	2	2	1	0	1
CARE-Index	3	5	0	6	0	0	0	0	0
CIB	0	0	0	0	0	0	0	0	0
DMC	3	5	1	1	0	0	0	0	0
EAS	4	10	0	5	3	0	0	0	0
FEAS	0	0	0	0	0	0	0	0	0
GLOS-R	4	7	0	4	5	0	0	0	1
GRS	3	4	0	2	0	0	0	0	0
IPSIC	2	4	0	0	0	0	0	0	0
M-I/TFS	4	5	0	3	0	0	0	0	1
MICS	4	5	1	0	0	1	0	0	0
MIPIS	0	0	0	0	0	0	0	0	0
Monadic Phases	0	0	0	0	0	0	0	0	0
MRO	4	9	3	1	0	0	0	0	0
NCAFS	4	7	1	1	0	0	0	0	0
NCATS	4	8	1	3	0	0	0	0	0
PCERA	3	7	0	0	0	0	0	0	0
PIIS	0	0	0	0	0	0	0	0	0
PIOG	4	6	1	1	1	0	1	0	0
PIPE	4	5	0	3	0	0	0	0	0
RCS	3	7	0	15	3	0	0	0	0

Numbers reported in the five validity domains indicate the number of sources that supported evidence for validity. No. of validity domains = number of validity domains with validity evidence. No. of validity sub-domains = number of validity sub-domains with validity evidence

Table 6 Further validity evidence of direct observational tools for measuring parent–infant interaction based on studies with unsupported study quality and articles describing an observational tool

Tool name	No. of validity domains	No. of validity sub-domains	Content			Response			Internal structure			
			Theoretical basis	Literature review	Expert consensus	Rater training	Rating procedure	Behavioral response	Internal consistency	Reproducibility	Sensitivity to change	Dimensional structure
ADS	3	5	1	1	0	0	1	0	0	0	0	0
AMIS	4	4	0	1	0	0	1	0	0	0	1	0
Belsky/ Unnamed	0	0	0	0	0	0	0	0	0	0	0	0
BMIS	3	5	0	0	0	0	0	0	0	2	0	0
CARE-Index	2	2	1	0	0	1	0	0	0	0	0	0
CIB	2	2	1	0	0	0	0	0	1	0	0	0
DMC	4	4	1	0	0	0	0	0	0	0	1	0
EAS	2	3	0	0	0	2	1	0	0	2	0	0
FEAS	4	5	1	1	0	1	0	0	0	0	0	0
GLOS-R	2	2	1	0	0	0	0	0	0	1	0	0
GRS	0	0	0	0	0	0	0	0	0	0	0	0
IPSIC	0	0	0	0	0	0	0	0	0	0	0	0
M-I/TFS	3	4	0	0	0	1	1	0	0	2	0	0
MICS	2	3	1	1	0	0	0	0	0	0	1	0
MIPIS	3	4	1	0	0	0	0	0	1	1	0	0
Monadic Phases	4	4	1	0	0	0	1	0	0	2	0	0
MRO	0	0	0	0	0	0	0	0	0	0	0	0
NCAFS	5	14	0	1	1	6	4	0	2	5	0	1
NCATS	5	15	0	0	1	8	5	1	7	4	3	3
PCERA	1	1	0	0	0	0	0	0	0	1	0	0
PIIS	3	4	1	1	0	0	0	0	0	1	0	0
PIOG	3	7	0	1	1	1	1	0	0	0	0	0
PIPE	0	0	0	0	0	0	0	0	0	0	0	0
RCS	3	4	1	0	1	0	2	0	0	1	0	0

Table 6 continued

Tool name	No. of validity domains	No. of validity sub-domains	Relations to other variables			Consequences			
			Convergent Divergent	Discriminant	Criterion	Reaction	Learning	Behavior	Results
ADS	3	5	0	0	0	0	1	0	1
AMIS	4	4	0	0	1	0	0	0	0
Belsky/ Unnamed	0	0	0	0	0	0	0	0	0
BMIS	3	5	0	4	0	1	1	1	0
CARE-Index	2	2	0	0	0	0	0	0	0
CIB	2	2	0	0	0	0	0	0	0
DMC	4	4	0	2	0	0	0	0	1
EAS	2	3	0	0	0	0	0	0	0
FEAS	4	5	0	4	0	0	0	0	1
GLOS-R	2	2	0	0	0	0	0	0	0
GRS	0	0	0	0	0	0	0	0	0
IPSIC	0	0	0	0	0	0	0	0	0
M-I/TFS	3	4	0	0	0	0	0	0	2
MICS	2	3	0	0	0	0	0	0	0
MIPIS	3	4	0	0	1	0	0	0	0
Monadic Phases	4	4	1	0	0	0	0	0	0
MRO	0	0	0	0	0	0	0	0	0
NCAFS	5	14	1	17	3	1	1	1	4
NCATS	5	15	1	20	11	1	1	1	4
PCERA	1	1	0	0	0	0	0	0	0
PIIS	3	4	1	0	0	0	0	0	0
PIOG	3	7	0	0	0	2	1	1	0
PIPE	0	0	0	0	0	0	0	0	0
RCS	3	4	0	0	0	0	0	0	0

Numbers reported in the five validity domains indicate the number of sources that supported evidence for validity. No. of validity domains = number of validity domains with validity evidence. No. of validity sub-domains = number of validity sub-domains with validity evidence

included in this review suffered from unsupported study quality. For example, researchers recruited convenience samples, and 29 % of the studies used a sample of 50 or fewer participants. These study samples are too small to produce sufficiently precise reliability and validity coefficients (Charter 2003). If possible, future research should randomly select large representative clinical or general population samples. Multicenter studies could help to increase the sample size and to improve the generalizability of the findings. Guidelines that define and standardize the criteria by which the quality of a tool should be evaluated would also be helpful for an evidence-based selection of parent–infant interaction tools.

Strengths and Limitations

One major strength of this review is the comprehensiveness of our search. We searched twenty databases and identified and screened more than 13,000 unique sources, including grey literature, in all languages and publication years. This approach may have minimized the likelihood of missing relevant evidence. We used a highly sensitive approach by searching for observational tools in all identified articles that used or reported on an observational tool of parent–infant interaction. This strategy enabled the identification of a high number of tools. The tools and study characteristics chosen for data extraction and evaluation were determined a priori, and our search strategy was pilot-tested. The evaluation of the validity evidence of the tools was based on accepted standards in assessment (Cicchetti 1994; Cicchetti and Sparrow 1990) and showed sufficient inter-rater reliability.

In addition to these strengths, this review has limitations that should be considered in interpreting our results. No standardized guideline exists for psychometric reviews of observational tools. Therefore, the criteria we used to evaluate the validity evidence of the tools are debatable. Validity evidence was based on articles that psychometrically evaluated or described the tools. Other types of studies are likely to exist that could have added validity evidence. To examine a representative sample of the literature, we did not exclude studies with low methodological quality. Instead, we separately reported the validity evidence for studies with credible study quality and for studies with unsupported study quality. For the studies with unsupported study quality, study samples were often small, which may have resulted in inaccurate reliability and validity estimates (Charter 1999, 2003).

We focused our review on tools that measured both parent and infant behavior. Other observational tools with good psychometric properties might be available that assess only the parent's behavior or the infant's behavior. We also excluded tools that did not primarily focus on the parent–infant interaction, such as the Home Observation for Measurement of the Environment Inventory (Caldwell and Bradley 2001). Such tools were not in the scope of our review and should be evaluated in a separate review. The exclusion of tools that were only used in one or no peer-reviewed journal articles may have biased our analysis because these tools may differ systematically from the others. However, the exclusion of these tools prevented us from reporting a huge number of tools with little or no impact on the field. Finally, the extent to which the reviewed tools are valid across different cultures is questionable because most of the included studies were conducted in the USA.

Practical Recommendations for Selecting a Tool

Decisions about the selection of a tool should be guided by a clear idea of what behavior requires assessment, how it will be assessed, and the purpose for which it will be used (McCloskey 1990). After a particular purpose has been defined, researchers or clinicians should precisely define the constructs that are relevant to the assessment. According to these definitions of the constructs, an assessment approach with a corresponding level of specificity should be chosen, such as counts of single behaviors or global ratings of complex constructs (a description of the constructs of the tools can be found in Table 3). If multiple tools are identified as appropriate for a particular purpose, evidence for the validity of these tools should be reviewed. The tools that have established evidence in a high number of validity domains might be a good starting point. Researchers or clinicians should prioritize particular psychometric features of the tool with respect to their purpose of assessment (e.g., predictive validity if the measured aspects of parent–infant interaction should predict later child outcomes). This information can be obtained from Tables 5 and 6, and Supplement B and C in ESM 2 and 3. Practical constraints might also be important arguments for or against a tool, such as the accessibility of training, required resources for training, observation and rating, and the extent of the tool's use by others to allow comparisons with previous assessments (see Table 4). Clinicians should also consider the feasibility of tool implementation in clinical practice.

Conclusion

Our review provides a synthesis of the available validity evidence of observational tools for measuring parent–infant interaction. We identified 24 tools, of which most provided sufficient evidence on the rating procedure, the reproducibility (objectivity), and discriminant validity, based on studies with credible study quality. Most tools

suffered from a lack of credible empirical research in the domains of content validity and consequential validity. Several psychometric properties of the tools, particularly the behavioral response, sensitivity to change, factorial validity, predictive validity, norms, and cut-off scores, have been understudied and should be considered more carefully in future research.

This review adds to the current knowledge on the validity of observational tools for measuring parent–infant interaction. The synthesized validity evidence and descriptions of the tools can guide researchers and clinicians in the selection of an appropriate tool. Such an evidence-based selection may advance the quality of research on parent–infant interaction and family care in mental health and social services.

Acknowledgments We would like to thank the following individuals for their contributions to the review: Linda Mische and Sophie Kröger for their help in reviewing articles; Klaus-Dieter Papke for his support in the development of the search strategy; Justine König for her literature order; Ina Weitkamp for her review of the manuscript; and Jessica Terese Müller for her linguistic revision of the manuscript.

Conflict of interest The authors declare that they have no conflict of interest.

Appendix 1

MEDLINE (OVID) Search Strategy

1. mothers/ OR mother$.mp OR parents/ OR parent$.mp OR fathers/ OR father$.mp OR caregivers/ OR caregiv$.mp OR maternal behavior/ OR maternal$.mp OR paternal behavior/ OR paternal$.mp
2. infant/ OR infant behavior/ OR infant$.mp OR baby$.mp OR babie$.mp
3. 1 and 2
4. mother$-infant$.mp OR infant$-mother$.mp OR father$-infant$.mp OR infant$-father$.mp OR parent$-infant$.mp OR infant$-parent$.mp OR caregiver$-infant$.mp OR infant-caregiver$.mp OR maternal$-infant$.mp OR infant$-maternal$.mp OR paternal$-infant$.mp OR infant$-paternal$.mp OR mother$-bab$.mp OR bab$-mother$.mp OR father$-bab$.mp OR bab$-father$.mp OR parent$-bab$.mp OR bab$-parent$.mp OR caregiver$-bab$.mp OR bab$-caregiv$.mp OR maternal$-bab$.mp OR bab$-maternal$.mp OR paternal$-bab$.mp OR bab$-paternal$.mp
5. 3 or 4
6. interaction$.mp OR interactiv$.mp OR communication/ OR communicat$.mp OR nonverbal communication/ OR nonverbal communicat$.mp OR dyadic behavio$.mp OR interpersonal relations/ OR interpersonal relation$.mp OR mother–child relations/ OR mother–child relation$.mp OR father–child relations/ OR father–child relation$.mp
7. 5 and 6
8. observation/ OR observation$.mp OR behavio$ cod$.mp OR behavio$ assessment$.mp OR behavio$ measure$.mp OR rat$ scale$.mp OR cod$ system$.mp OR microanalys$.mp OR checklist/ OR checklist$.mp OR videotape recording/ OR videotap$.mp OR video$ record$.mp
9. 7 and 8

Appendix 2

References of Articles Describing or Psychometrically Evaluating Direct Observational Tools

Aksan, N., Kochanska, G., & Ortmann, M. R. (2006). Mutually Responsive Orientation between parents and their young children: Toward methodological advances in the science of relationships. *Developmental Psychology, 42*(5), 833–848. doi:10.1037/0012-1649.42.5.833.

Allen, D., & St Arnaud, S. (1994). NCAST: An effective nursing tool for assessing family interaction *AARN Newsletter, 50*, 14–15.

Als, H., Tronick, E., & Brazelton, T. B. (1979). Analysis of face-to-face interaction in infant–adult dyads. In M. E. Lamb, S. J. Suomi & G. Stephenson (Eds.), *Social interaction analysis: Methodological issues* (pp. 33–76). Madison, WI: University of Wisconsin Press.

Baird, S. M., Haas, L., McCormick, K., Carruth, C., & Turner, K. D. (1992). Approaching an objective system for observation and measurement: Infant–Parent Social Interaction Code. *Topics in Early Childhood Special Education, 12*(4), 544–571. doi:10.1177/027112149201200410.

Barnard, K., Eyres, S. J., Lobo, M. L., & Snyder, C. (1983). An ecological paradigm for assessment and intervention. In T. B. Brazelton & B. M. Lester (Eds.), *New approaches to developmental screening of infants* (pp. 199–218). New York, NY: Elsevier.

Barnard, K. E., Hammond, M. A., Booth, C. L., Bee, H. L., Mitchell, S. K., & Spieker, S. J. (1989). Measurement and meaning of parent–child interaction. In C. L. F. Morrison & D. Keating (Eds.), *Applied developmental psychology* (3 ed., pp. 39–80). New York, NY: Academic Press.

Barone, L., & Biringen, Z. (2007). Le scale per la valutazione della disponibilità emozionale [Emotional

Availability Scales - EAS]. In L. Barone (Ed.), *Emozioni e sviluppo. Percorsi tipici e atipici*. Rome, Italy: Carocci.

Belsky, J., Taylor, D. G., & Rovine, M. (1984). The Pennsylvania Infant and Family Development Project, II: The development of reciprocal interaction in the mother–infant dyad. *Child Development, 55*(3), 706–717. doi:10.1111/1467-8624.ep12422849.

Bernstein, V. J., Hans, S. L., & Percansky, C. (1991). Advocating for the young child in need through strengthening the parent–child relationship. *Journal of Clinical Child and Adolescent Psychology, 20*(1), 28–41. doi:10.1207/s15374424jccp2001_5.

Bernstein, V. J., Hans, S. L., & Percansky, C. (1992). *Parent–Infant Observation Guide for program planning around feeding and other daily routines*. Unpublished manual. Department of Psychiatry, University of Chicago. Chicago, IL.

Bernstein, V. J., Percansky, C., & Hans, S. L. (1987). *Screening for social-emotional impairment in infants born to teenage mothers*. Paper presented at the Biennial Meeting of the Society for Research in Child Development, Baltimore, MD.

Biringen, Z. (1998, April). "Apparent sensitivity": A methodological note and a proposed link with attachment. *Infant Behavior and Development, 21, Supplement*, 28. doi:10.1016/S0163-6383(98)91243-7.

Biringen, Z. (2000). Emotional availability: Conceptualization and research findings. *American Journal of Orthopsychiatry, 70*(1), 104–114. doi:10.1037/h0087711.

Biringen, Z. (2005). Training and reliability issues with the Emotional Availability Scales. *Infant Mental Health Journal, 26*(4), 404–405. doi:10.1002/imhj.20060.

Biringen, Z. (2008). *Emotional Availability (EA) Scales, 4th Edition. Infancy/early childhood version*. Boulder, CO: Emotionalavailability.com.

Biringen, Z., Batten, R., Neelan, P., Altenhofen, S., Swaim, R., Bruce, A.,… Zachary, V. (2010). Emotional availability (EA): The assessment of and intervention for global parent–child relational quality. *Journal of Experiential Psychotherapy, 49*, 3–9.

Biringen, Z., Damon, J., Grigg, W., Mone, J., Pipp-Siegel, S., Skillern, S., & Stratton, J. (2005). Emotional availability: Differential predictions to infant attachment and kindergarten adjustment based on observation time and context. *Infant Mental Health Journal, 26*(4), 295–308. doi:10.1002/imhj.20054.

Biringen, Z., & Easterbrooks, M. (2012). Emotional availability: Concept, research, and window on developmental psychopathology. *Development and Psychopathology, 24*(01), 1–8. doi:10.1017/S0954579411000617.

Biringen, Z., Emde, R. N., Brown, D., Lowe, L., Myers, S., & Nelson, D. (1999). Emotional availability and emotion communication in naturalistic mother–infant interactions: Evidence for gender relations. *Journal of Social Behavior and Personality, 14*(4), 463–478.

Biringen, Z., & Robinson, J. (1991). Emotional availability in mother–child interactions: A reconceptualization for research. *American Journal of Orthopsychiatry, 61*(2), 258–271. doi:10.1037/h0079238.

Boechler, V. L. (2000). *Predictors of father–child Interaction measured by the Nursing Child Assessment Teaching Scale.* Master's thesis, University of Alberta, Edmonton, Canada.

Boffman, J. L., Clark, N. J., & Helsel, D. (1997). Can NCAST and HOME assessment scales be used with Hmong refugees? *Pediatric Nursing, 23*(3), 235–244.

Bornstein, M. H., Gini, M., Suwalsky, J. T. D., Leach, D. B., & Haynes, O. M. (2006). Emotional availability in mother–child dyads: Short-term stability and continuity from variable-centered and person-centered perspectives. *Merrill-Palmer Quarterly, 52*(3), 547–571. doi:10.1353/mpq.2006.0024.

Bornstein, M. H., Suwalsky, J. T. D., Putnick, D. L., Gini, M., Venuti, P., de Falco, S.,… de Galperín, C. Z. (2010). Developmental continuity and stability of emotional availability in the family: Two ages and two genders in child–mother dyads from two regions in three countries. *International Journal of Behavioral Development, 34*(5), 385–397. doi:10.1177/0165025409339080.

Burns, K., Chethik, L., Burns, W. J., & Clark, R. (1991). Dyadic disturbances in cocaine-abusing mothers and their infants. *Journal of Clinical Psychology, 47*(2), 316–319. doi:10.1002/1097-4679(199103)47:2<316::AID-JCLP2270470220>3.0.CO;2-1.

Byrne, M. W., & Keefe, M. R. (2003). Comparison of two measures of parent–child interaction. *Nursing Research, 52*(1), 34–41. doi:10.1097/00006199-200301000-00005.

Campbell, B. K. (2009). *The use of the Massie Campbell Scale in therapeutic intervention with parents*. Paper presented at the Association for Child Psychoanalysis (ACP) Extension Division Meeting, Rio Rancho, NM.

Carter, A. S., Little, C. B., & Garrity-Rokous, F. (1998, April). Adapting the Emotional Availability Scales for 4-month old infant–mother dyads: Associations with molecular coding and maternal psychopathology. *Infant Behavior & Development, 21*(Suppl.), 25.

Cena, L., & Imbasciati, A. (2009). Strumenti in Psicologia Clinica Perinatale: Il CARE-Index per una valutazione della qualità della relazione primaria [Tools in perinatal clinical psychology: The CARE-Index for an assessment of the quality of the primary relationship].

Psychofenia: Ricerca ed Analisi Psicologica, 12(20), 153–172. doi:10.1285/i17201632vXIIn20p153.

Censullo, M. (1994). Strategy for promoting greater responsiveness in adolescent parent/infant relationships: Report of a pilot study. *Journal of Pediatric Nursing, 9*(5), 326–332.

Censullo, M. (2004). *Dyadic Mutuality Code manual*. Unpublished manual, School of Nursing and Health studies, Georgetown University. Washington, DC.

Censullo, M., Bowler, R., Lester, B., & Brazelton, T. B. (1987). An instrument for the measurement of infant–adult synchrony. *Nursing Research, 36*(4), 244–248. doi:10.1097/00006199-198707000-00015.

Chatoor, I., Getson, P., Menvielle, E., Brasseaux, C., O'Donnell, R., Rivera, Y., & Mrazek, D. A. (1997). A feeding scale for research and clinical practice to assess mother–infant interactions in the first three years of life. *Infant Mental Health Journal, 18*(1), 76–91. doi:10.1002/(SICI)1097-0355(199721)18:1<76::AID-IMHJ6>3.0.CO;2-Z.

*Chatoor, I., Loeffler, C., McGee, M., & Menvielle, E. (1998). *Observational Scale for Mother–Infant Interaction During Feeding*. Unpublished manual, Children's National Medical Center, Washington DC.

Chatoor, I., Schaefer, S., Dickson, L., Egan, J., Conners, C. K., & Leong, N. (1984). Pediatric assessment of non-organic failure to thrive. *Pediatric Annals, 13*(11), 844, 847–848, 850.

*Clark, R. (1985). *The Parent–Child Early Relational Assessment. Instrument and manual.* Unpublished manual, Department of Psychiatry, University of Wisconsin Medical School. Madison, WI.

Clark, G., & Seifer, R. (1985). Assessment of parent's interactions with their developmentally delayed infants. *Infant Mental Health Journal, 6*(4), 214–225. doi:10.1002/1097-0355(198524)6:4<214::AID-IMHJ2280060406>3.0.CO;2-3.

Clark, G. N., & Seifer, R. (1983). Facilitating mother–infant communication: A treatment model for high-risk and developmentally-delayed infants. *Infant Mental Health Journal, 4*(2), 67–82. doi:10.1002/1097-0355(198322)4:2<67::aid-imhj2280040202>3.0.co;2-i.

Clark, R. (1999). The parent–child early relational assessment: A factorial validity study. *Educational and Psychological Measurement, 59*(5), 821–846. doi:10.1177/00131649921970161.

Clark, R., Musick, J., Stott, F., Klehr, K., & Cohler, B. (1984). *Mother–child dyads at risk: Development of rating scales for early identification of disturbances in affect and behavior*. Paper presented at the Fourth International Conference on Infant Studies, New York, NY.

Cohn, J. F., & Tronick, E. Z. (1988). Discrete versus scaling approaches to the description of mother–infant face-to-face interaction: Convergent validity and divergent applications. *Developmental Psychology, 24*(3), 396–397. doi:10.1037/0012-1649.24.3.396.

Couturier, G. L. G. (2005). *Kunnen cirkels handvatten krijgen? De FEAS, een instrument om klinische bevindingen te ordenen [Can circles get handles? The FEAS, a tool to organize clinical findings]*. Diagnostiek van het jonge kind van nul tot vier. Unpublished report, Leiden: Boerhaave.

Crittenden, P. M. (2005). Der CARE-Index als Hilfsmittel für Früherkennung, Intervention und Forschung [Using the CARE Index for screening, intervention, and research]. *Frühförderung Interdisziplinär, 24*(3), 99–106.

Crittenden, P. M. (2006). *CARE-Index Infants (birth - 15 months). Coding manual*. Unpublished manual, Family Relations Institute, Miami, FL.

Davis, W. (1994). Use of the Nursing Child Assessment Teaching Scale in occupational therapy. *NCAST National News, 10,* 1–2, 7–8.

Eickhorst, A., Schweyer, D., Kohler, H., Jelen-Maubousssin, A., Kunz, E., Sidor, A., & Cierpka, M. (2010). Elterliche Feinfühligkeit bei Müttern und Vätern mit psychosozialen Belastungen. [Sensitivity of parents with psychosocial stress.]. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz, 53*(11), 1126–1133. doi:10.1007/s00103-010-1143-7.

Evans, C. A., & Porter, C. L. (2009). The emergence of mother–infant co-regulation during the first year: Links to infants' developmental status and attachment. *Infant Behavior and Development, 32*(2), 147–158. doi:10.1016/j.infbeh.2008.12.005.

Farel, A. M., Freeman, V. A., Keenan, N. L., & Huber, C. J. (1991). Interaction between high-risk infants and their mothers: The NCAST as an assessment tool. *Research in Nursing & Health, 14*(2), 109–118. doi:10.1002/nur.4770140205.

Feldman, R. (1998). *Coding Interactive Behavior (CIB)*. Unpublished manual, Department of Psychology, Bar-Ilan University, Ramat-Gan, Israel.

Feldman, R. (2012). Parenting behavior as the environment where children grow. In L. C. Mayes & M. Lewis (Eds.), *A developmental environment measurement handbook* (pp. 535–567). New York, NY: Cambridge University Press.

Fiese, B. H., Irwin, M., Gordon, M., & Poehlmann, J. (2001). *PIPE training manual 1st edition 2001*. Unpublished manual, Syracuse University, Syracuse, NY.

Fiese, B. H., Poehlmann, J., Irwin, M., Gordon, M., & Curry-Bleggi, E. (2001). A pediatric screening instrument to detect problematic infant–parent interactions: Initial reliability and validity in a sample of high- and

low-risk infants. *Infant Mental Health Journal, 22*(4), 463–478. doi:10.1002/imhj.1011.

Fogel, A. (2000). *Beyond the individuals: An historic-relational approach to the theory and research on communication*. Unpublished report, Department of Psychology, University of Utah, Salt Lake City, UT.

Fogel, A., De Koyer, I., Secrist, C., Sipherd, A., Hafen, T., & Fricke, M. (2003). *The Revised Relational Coding System*. Unpublished manual, Department of Psychology, University of Utah, Salt Lake City, UT.

Fujii, M. (1997). 新生児期の母子相互作用の検討: NCAFS の適用 (自然科学系) [Application of assessment tool (NCAFS) to investigate mother–neonate interaction]. *大阪府立看護大学医療技術短期大学部紀要 [Bulletin of Osaka Prefectural College of Health Sciences], 3*, 41–49.

Gaffney, K. F., Kodadek, M. P., Meuse, M. T., & Jones, G. B. (2001). Assessing infant health promotion: A cross-cultural comparison. *Clinical Nursing Research, 10*(2), 102–116. doi:10.1177/10547730122158815.

Galván-Bovaira, M. J., Gràcia, M., & del Río, M. J. (2002). Una propuesta de evaluación de las interacciones comunicativas madre-hijo: La prueba Mother–Infant Communication Screening (MICS) [A proposal for the assessment of the mother–child communication: The Mother–Infant Communication Screening (MICS)]. *Revista de Logopedia, Foniatría y Audiología, 22*(1), 15–23. doi:10.1016/S0214-4603(02)76217-7.

Göttler, C. (2010). *Fütterstörungen und peri-und postnatale Morbidität bei sehr kleinen Frühgeborenen mit einem Geburtsgewicht <= 1500 Gramm [Feeding disorders and peri-and postnatal morbidity in very small preterm infants with a birth weight <= 1500 grams]*. Doctoral dissertation, Ludwig Maximilian University of Munich, Munich, Germany.

Greenspan, S. I. (1992). The Functional Emotional Assessment Scale for infancy and early childhood. In S. I. Greenspan (Ed.), *Infancy and early childhood: The practice of clinical assessment and intervention with emotional and developmental challenges* (pp. 381–426). Madison, CT: International Universities Press.

Greenspan, S. I. (1996). Assessing the emotional and social functioning of infants and young children. In S. J. Meisels & E. Fenichel (Eds.), *New visions for the developmental assessment of infants and young children* (pp. 231–266). Washington, DC: Zero to Three.

Greenspan, S. I., DeGangi, G., & Wieder, S. (2001). *The Functional Emotional Assessment Scale (FEAS) for infancy and early childhood: Clinical and research applications*. Bethesda, MD: Interdisciplinary Council on Developmental & Learning Disorders.

Greenspan, S. I., & Lieberman, A. F. (1980). Infants, mothers, and their interactions: A quantitative clinical approach to developmental assessment. In S. I. Greenspan & G. H. Pollock (Eds.), *The course of life: Psychoanalytic contributions toward understanding personality development, Vol. I. Infancy and early childhood* (pp. 271–312). Madison, Wl: International Universities Press.

Greenspan, S. I., & Lieberman, A. F. (1989). Infants, mothers, and their interaction: A quantitative clinical approach to developmental assessment. In S. I. Greenspan & G. H. Pollock (Eds.), *The course of life, Vol 1: Infancy* (pp. 503–560). Madison, CT: International Universities Press.

Gunning, M., Conroy, S., Valoriani, V., Figueiredo, B., Kammerer, M. H., Muzik, M.,… & Murray, L. (2004). Measurement of mother–infant interactions and the home environment in a European setting: Preliminary results from a cross-cultural study. *The British Journal of Psychiatry, 184*(Suppl 46), 38–44. doi:10.1192/03-336.

*Gunning, M., Murray, L., & Lawson, C. (2002). *Global Rating Scheme*. Paper presented at the Eighth Congress of the World Association for Infant Mental Health, Amsterdam, Netherland.

Hale, A. K., Holditch-Davis, D., D'Auria, J., & Miles, M. S. (1999). The usefulness of an assessment of emotional involvement of HIV-positive mothers and their infants. *Journal of Pediatric Health Care, 13*(5), 230–236. doi:10.1016/S0891-5245(99)90004-4.

Hans, S. L., & Bernstein, V. J. (1990). Planning programmes for high-risk infants: A facet analysis of parent–infant communication. *Applied Psychology: An International Review, 39*(4), 457–478. doi:10.1111/j.1464-0597.1990.tb01066.x.

Hans, S. L., Bernstein, V. J., & Percansky, C. (1991). Adolescent parenting programs: Assessing parent–infant interaction. *Evaluation and Program Planning, 14*(1), 87–95. doi:10.1016/0149-7189(91)90042-F.

Harrison, M. J., Magill-Evans, J., & Benzies, K. (1999). Fathers' scores on the Nursing Child Assessment Teaching Scale: Are they different from those of mothers? *Journal of Pediatric Nursing, 14*(4), 248–254. doi:10.1016/S0882-5963(99)80019-0.

Henning, A., & Aschersleben, G. (2008). *Mother–infant interaction at 3 and 6 months: A comparison of measures of responsiveness*. Paper presented at the XVIth International Conference on Infant Studies, Vancouver, Canada. Abstract retrieved from http://www.isisweb.org/SpringboardWebApp/userfiles/isis/file/ICIS08%20Abstract%20Volume.pdf#page=377.

Hipwell, A. E., & Kumar, R. (1996). Maternal psychopathology and prediction of outcome based on mother–infant interaction ratings (BMIS). *British Journal of Psychiatry, 169*(5), 655–661. doi:10.1192/bjp.169.5.655.

Hirose, T., Teramoto, T., Mikuni, K., Saitoh, S., Takahashi, I., Hiramatsu, M.,… Hashimoto, S. (2005). *An intervention study using NCATS for mothers in Japan*. Paper presented at the Society for Research in Child Development (SRCD) 2005 Biennial Meeting, Atlanta, GA.

Hirose, T., Teramoto, T., Saitoh, S., Takahashi, I., Hiramatsu, M., Okamitsu, M.,… Shirakawa, S. (2007). Preliminary early intervention study using Nursing Child Assessment Teaching Scale in Japan. *Pediatrics International, 49*(6), 950–958. doi:10.1111/j.1442-200X.2007.02498.x.

Hodges, E. A., Houck, G. M., & Kindermann, T. (2007). Reliability of the Nursing Child Assessment Feeding Scale during toddlerhood. *Issues in Comprehensive Pediatric Nursing, 30*(3), 109–130. doi:10.1080/01460860701525204.

Hodges, E. A., Houck, G. M., & Kindermann, T. (2009). Validity of the Nursing Child Assessment Feeding Scale during toddlerhood. *Western Journal of Nursing Research, 31*(5), 662–678. doi:10.1177/0193945909332265.

Hofheimer, J., & Appelbaum, M. (1992). *Risk detection using observations of interaction (Report No. 8MCH/CCS-92/06)*. Springfield, VA: National Technical Information Service.

Howes, C., & Obregon, N. B. (2009). Emotional availability in Mexican-heritage low-income mothers and children: Infancy through preschool. *Parenting: Science and Practice, 9*(3–4), 260–276. doi:10.1080/15295190902844589.

Huber, C. J. (1991). Documenting quality of parent–child interaction: Use of the NCAST scales. *Infants & Young Children, 4*(2), 63–75.

Kemppinen, K., Kumpulainen, K., Raita-Hasu, J., Moilanen, I., & Ebeling, H. (2006). The continuity of maternal sensitivity from infancy to toddler age. *Journal of Reproductive and Infant Psychology, 24*(3), 199–212. doi:10.1080/02646830600821249.

Kumar, R., & Hipwell, A. E. (1996). Development of a clinical rating scale to assess mother–infant interaction in a psychiatric mother and baby unit. *British Journal of Psychiatry, 169*(1), 18–26. doi:10.1192/bjp.169.1.18.

Massie, H. N., & Campbell, B. K. (1982). The AIDS scale - The Massie-Campbell Scale of Mother–Infant Attachment Indicators During Stress. In J. M. Stack (Ed.), *The special infant. An interdisciplinary approach to the optimal development of infants*. New York, NY: Human Sciences Press.

Massie, H. N., & Campbell, B. K. (1983). The Massie-Campbell Scale of Mother–Infant Attachment Indicators during Stress (AIDS scale). In J. Call, E. Galenson & R. Tyson (Eds.), *Frontiers of Infant Psychiatry* (pp. 394–412). New York, NY: Basic Books.

Massie, H. N., & Campbell, B. K. (1984). The Massie-Campbell Scale of Mother–Infant Attachment Indicators during Stress. In H. N. Massie & J. Rosenthal (Eds.), *Childhood psychosis in the first four years of life* (pp. 253–277). New York, NY: McGraw-Hill.

Massie, H. N., & Campbell, B. K. (1992). *The Massie-Campbell Scale of Mother–Infant Attachment Indicators during Stress: For use during the pediatric examination and other childcare situations*. Van Nuys, CA: Child Development Media.

Matias, R., Cohn, J. F., & Ross, S. (1989). A comparison of two systems to code infants' affective expression. *Developmental Psychology, 25*(4), 483–489. doi:10.1037/0012-1649.25.4.483.

Mischenko, J., Cheater, F., & Street, J. (2004). NCAST: Tools to assess caregiver–child interaction. *Community Practitioner, 77*(2), 57–60.

Nakamura, W. M., Stewart, K. B., & Tatarka, M. E. (2000). Assessing father–infant interactions using the NCAST teaching scale: A pilot study. *American Journal of Occupational Therapy, 54*(1), 44–51. doi:10.5014/ajot.54.1.44.

Oxford, M. L., & Findlay, D. M. (2012). *NCAST Caregiver/Parent–Child Interaction teaching manual*. Seattle, WA: NCAST Publication.

*Parks, P., Wendt, E., & Lenz, E. (1990, September). *Reliability and validity of the parent–child early relational assessment.* Paper presented at the Conference of the International Association for Infant Mental Health Ann Arbor, MI.

Pipp-Siegel, S., & Biringen, Z. (1998). Assessing the quality of relationships between parents and children: The Emotional Availability Scales. *The Volta Review, 100*(5), 237–249.

*Poisson, S. S., Hofheimer, J. A., Strauss, M. E., Lieberman, A. F., & Greenspan, S. I. (1983). *The assessment of infant–caregiver interaction: Interobserver reliability of a coding system*. Unpublished manuscript, NIMH Mental Health Study Center, Washington, DC.

Porter, C. L. (2003). Co-regulation in mother–infant dyads: Links to infants' cardiac vagal tone. *Psychological Reports, 92*(1), 307–319. doi:10.2466/pr0.92.1.307-319.

Porter, C. L., & Porter, A. E. (1999). *Co-regulated communication in mother–infant dyads at 3 and 6 months: Associations with maternal efficacy, infant temperament and cardiac vagal tone*. Paper presented at the Biennial Meetings of the Society for Research in Child Development (SRCD), Albuquerque, NM.

Porter, C. L., & Porter, A. E. (2000). *Co-regulation in mother–infant dyads: Links to infant temperament and exploration*. Paper presented at the International Conference for Infant Studies, Brighton, England.

Porter, C. L., Porter, A. E., Christensen, T. E., & Morgan, T. (1998). Co-regulated patterns of communication in mother–infant dyads: Contributions from mothers' feelings of efficacy and infant temperament and vagal tone. *Infant Behavior and Development, 21*(Suppl.), 626.

Porter, C. L., Sonntag, P., Porter, A. E., & Peterson, R. (2001). *Co-regulation in mother–infant dyads: Links to infant temperament and physiological regulation*. Paper presented at the Biennial Meetings of the Society for Research in Child Development (SRCD), Minneapolis, MN.

Price, G. M. (1975). *Influencing maternal care through discussion of videotapes of maternal–infant feeding interaction.* Doctoral dissertation, Boston University, Boston, MA.

Price, G. M. (1982). *The assessment of empathy in the mother–infant relationship*. Paper presented at the International Conference on Infant Studies, Austin, TX.

Price, G. M. (1983). Sensitivity in mother–infant interactions: The AMIS scale. *Infant Behavior and Development, 6*(3), 353–360. doi:10.1016/S0163-6383%2883%2980043-5.

Puura, K., Guedeney, A., Mantymaa, M., & Tamminen, T. (2007). Detecting infants in need: Are complicated measures really necessary? *Infant Mental Health Journal, 28*(4), 409–421. doi:10.1002/imhj.20144.

Raack, C. B. (1989). *Mother/Infant Communication Screening (MICS)*. Roselle, IL: Community Therapy Services.

Rankl, C. (1996). *Der interaktionelle Ansatz in der Säuglingsdiagnostik am Beispiel der Fütterungs- und nicht organischen Gedeihstörungen. Validierung der Mother–Infant-Feeding Scale von Chatoor [The interactional approach in neonatal diagnosis: Validation of Chatoor's Mother–Infant Feeding Scale].* Doctoral dissertation, University of Vienna, Austria.

Stocky, A. J. (1990). *A multicentre pilot study examining aspects of the reliability and validity of the Bethlem Mother–infant Interaction Scale.* Doctoral dissertation, Monash University, Melbourne, Australia.

Stocky, A. J., Tonge, B. J., & Nunn, R. J. (1996). The reliability and validity of the Bethlem Mother–Infant Interaction Scale. *British Journal of Psychiatry, 169*(1), 27–29. doi:10.1192/bjp.169.1.27.

Sumner, G., & Spietz, A. (1994). *NCAST Caregiver/Parent–Infant Interaction feeding manual*. Seattle, WA: NCAST Publications.

Thome, M., Mincke, L., Jonsdottir, H. K., & Sigurjonsdottir, K. (2007). Training health visitors and community nurses and midwives across the North Atlantic in assessment of carer–infant interaction during the first year of life using the NCAST feeding scale. *Journal of Reproductive and Infant Psychology, 25*(3), 249.

Tronick, E., Als, H., & Brazelton, T. B. (1980). Monadic Phases: A structural descriptive analysis of infant–mother face to face interaction. *Merrill Palmer Quarterly, 26*(1), 3–24.

Valoriani, V., Vaiani, S., Vanni, C., Serafini, E., Pallanti, S., Noci, I.,… Murray, L. (2012). Triplets, twins and singleton too: Comparing mother–baby early interaction by G.R.S. in a sample of art-babies versus a sample of "spontaneous" babies. The "art in florence project-team" longitudinal study. *Journal of Maternal-Fetal and Neonatal Medicine, 25*, 98–99. doi:10.3109/14767058.2012.679162.

Vliegen, N. (2005). Observatieschalen voor emotionele beschikbaarheid [Emotional Avalability Scales]. *Tijdschrift voor Psychotherapie, 31*(1), 3–16. doi:10.1007/BF03062119.

Walker, L. O., & Thompson, E. T. (1982). Mother–Infant Play Interaction Scale. In S. S. Humenick (Ed.), *Analysis of current assessment strategies in the health care of young children and childbearing families* (pp. 191–201). East Norwalk, CT: Appleton-Century-Crofts.

Wojnar, D. (2011). Application of the Barnard parent/caregiver–child interaction model to care of premature infants. In M. de Chesnay & B. A. Anderson (Eds.), *Caring for the vulnerable* (pp. 135–146). Burlington, MA: Jones & Bartlett Learning.

Wu, Y. T., Lin, U. C., Yu, Y. T., Hsieh, W. S., Hsu, C. H., Hsu, H. C.,… & Jeng, S. F.. (2010). Reliability of the Assessment of Mother–Infant Sensitivity - Chinese version for preterm and term Taiwanese mother–infant dyads. *Physiotherapy Canada, 62*(4), 397–403. doi:10.3138/physio.62.4.397.

Zavaschi, M. L., Costa, F., Maciel, A. L., Leutchuk, D. M., Neto, M., Zachia, M.,… Tramontina, S. (1994). Validaçao de um instrumento para avaliaçao da interaçao entre mae e recém-nascido [Reliability and validity of an instrument to assess early mother–infant interaction]. *Revista Brasileira de Psiquiatria, 16*(3), 113–117.

Ziv, Y., Aviezer, O., Gini, M., Sagi, A., & Koren-Karie, N. (2000). Emotional availability in the mother–infant dyad as related to the quality of infant–mother attachment relationship. *Attachment & Human Development, 2*(2), 149–169. doi:10.1080/14616730050085536.

References without asterisk add up to more than 104 articles because some of the articles were published more than once.

*Excluded from review because full text could not be retrieved.

References

Aksan, N., Kochanska, G., & Ortmann, M. R. (2006). Mutually responsive orientation between parents and their young children: Toward methodological advances in the science of relationships. *Developmental Psychology, 42*(5), 833–848. doi:10.1037/0012-1649.42.5.833.

Alderfer, M. A., Fiese, B. H., Gold, J. I., Cutuli, J., Holmbeck, G. N., Goldbeck, L., & Patterson, J. (2008). Evidence-based assessment in pediatric psychology: Family measures. *Journal of Pediatric Psychology, 33*(9), 1046–1061. doi:10.1093/jpepsy/jsm083.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Bagner, D. M., Rodríguez, G. M., Blake, C. A., Linares, D., & Carter, A. S. (2012). Assessment of behavioral and emotional problems in infancy: A systematic review. *Clinical Child and Family Psychology Review, 15*(2), 113–128. doi:10.1007/s10567-012-0110-2.

Baird, S. M., Haas, L., McCormick, K., Carruth, C., & Turner, K. D. (1992). Approaching an objective system for observation and measurement: Infant–Parent Social Interaction Code. *Topics in Early Childhood Special Education, 12*(4), 544–571. doi:10.1177/027112149201200410.

Barnard, K. E., Hammond, M. A., Booth, C. L., Bee, H. L., Mitchell, S. K., & Spieker, S. J. (1989). Measurement and meaning of parent–child interaction. In C. L. F. Morrison & D. Keating (Eds.), *Applied developmental psychology* (3rd ed., pp. 39–80). New York, NY: Academic Press.

Beebe, B., Jaffe, J., Markese, S., Buck, K., Chen, H., Cohen, P.,… Feldstein, S. (2010). The origins of 12-month attachment: A microanalysis of 4-month mother–infant interaction. *Attachment & Human Development*, 12(1–2), 1–135. doi:10.1080/14616730903338985.

Belsky, J., Taylor, D. G., & Rovine, M. (1984). The Pennsylvania Infant and Family Development Project, II: The development of reciprocal interaction in the mother–infant dyad. *Child Development, 55*(3), 706–717. doi:10.1111/1467-8624.ep12422849.

Bernstein, V. J., Hans, S. L., & Percansky, C. (1992). *Parent–Infant Observation Guide for program planning around feeding and other daily routines*. Unpublished manual, Department of Psychiatry, University of Chicago, Chicago, IL.

Bernstein, V. J., Percansky, C., & Hans, S. L. (1987). *Screening for social-emotional impairment in infants born to teenage mothers*. Paper presented at the Biennial Meeting of the Society for Research in Child Development, Baltimore, MD.

Biringen, Z. (2008). *Emotional Availability (EA) Scales, 4th edition. Infancy/early childhood version*. Boulder, CO: Emotionalav ailability.com.

Biringen, Z., Damon, J., Grigg, W., Mone, J., Pipp-Siegel, S., Skillern, S., & Stratton, J. (2005). Emotional availability: Differential predictions to infant attachment and kindergarten adjustment based on observation time and context. *Infant Mental Health Journal, 26*(4), 295–308. doi:10.1002/imhj.20054.

Bornstein, M. H., Cote, L. R., Haynes, O. M., Suwalsky, J. T., & Bakeman, R. (2012). Modalities of infant–mother interaction in Japanese, Japanese American immigrant, and European American dyads. *Child Development, 83*(6), 2073–2088. doi:10.1111/j.1467-8624.2012.01822.x.

Braungart-Rieker, J. M., Garwood, M. M., Powers, B. P., & Wang, X. (2001). Parental sensitivity, infant affect, and affect regulation: Predictors of later attachment. *Child Development, 72*(1), 252–270. doi:10.1111/1467-8624.00277.

Caldwell, B. M., & Bradley, R. H. (2001). *HOME inventory and administration manual*. Homewood, IL: University of Arkansas for Medical Sciences and University of Arkansas at Little Rock.

Censullo, M. (2004). *Dyadic Mutuality Code manual*. Unpublished manual, School of Nursing and Health studies, Georgetown University, Washington, DC.

Charter, R. A. (1999). Sample size requirements for precise estimates of reliability, generalizability, and validity coefficients. *Journal of Clinical and Experimental Neuropsychology, 21*(4), 559–566. doi:10.5555/jcen.21.4.559.889.

Charter, R. A. (2003). Study samples are too small to produce sufficiently precise reliability coefficients. *The Journal of General Psychology, 130*(2), 117–129. doi:10.1080/00221300309601280.

Chatoor, I., Getson, P., Menvielle, E., Brasseaux, C., O'Donnell, R., Rivera, Y., & Mrazek, D. A. (1997). A feeding scale for research and clinical practice to assess mother–infant interactions in the first 3 years of life. *Infant Mental Health Journal, 18*(1), 76–91. doi:10.1002/(SICI)1097-0355(199721)18:1<76:AID-IMHJ6>3.0.CO;2-Z.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284. doi:10.1037//1040-3590.6.4.284.

Cicchetti, D. V., & Sparrow, S. S. (1990). Assessment of adaptive behavior in young children. In J. J. Johnson & J. Goldman (Eds.), *Developmental assessment in clincial child psychology: A handbook* (pp. 173–196). New York, NY: Pergamon Press.

Clark, R. (1985). *The Parent–Child Early Relational Assessment. Instrument and manual*. Unpublished manual, Department of Psychiatry, University of Wisconsin Medical School, Madison, WI.

Clark, R. (1999). The Parent-Child Early Relational Assessment: A factorial validity study. *Educational and Psychological Measurement, 59*(5), 821–846. doi:10.1177/00131649921970161.

Clark, G., & Seifer, R. (1985). Assessment of parent's interactions with their developmentally delayed infants. *Infant Mental Health Journal, 6*(4), 214–225. doi:10.1002/1097-0355(198524)6:4<214:AID-IMHJ2280060406>3.0.CO;2-3.

Clark, R., Tluczek, A., & Gallagher, K. C. (2004). Assessment of parent–child early relational disturbances. In R. DelCarmen-Wiggins & A. S. Carter (Eds.), *Handbook of infant, toddler, and preschool mental health assessment*. Oxford; New York, NY: Oxford University Press.

Cohn, J. F., Campbell, S. B., & Ross, S. (1991). Infant response in the still-face paradigm at 6 months predicts avoidant and secure attachment at 12 months. *Development and Psychopathology, 3*(04), 367–376. doi:10.1017/S0954579400007574.

Committee on Psychosocial Aspects of Child Family Health. (1997). *Guidelines for health supervision 3*. Elk Grove Village, IL: American Academy of Pediatrics.

Corcoran, K., & Fischer, J. (2013). *Measures for clinical practice and research* (Vol. 2). New York, NY: Oxford University Press.

Crittenden, P. M. (2006). *CARE-Index Infants (birth–15 months). Coding manual*. Coding manual of the Child–Adult Relationship Experimental Index (CARE-Index). Unpublished manual, Family Relations Institute, Miami, FL.

De Wolff, M. S., & van Ijzendoorn, M. H. (1997). Sensitivity and attachment: A meta-analysis on parental antecedents of infant attachment. *Child Development, 68*(4), 571–591. doi:10.1111/j.1467-8624.1997.tb04218.x.

DelCarmen-Wiggins, R., & Carter, A. (2004). *Handbook of infant, toddler, and preschool mental health assessment*. New York, NY: Oxford University Press.

Early Head Start National Resource Center. (2013). *Developmental screening, assessment, and evaluation: Key elements for*

individualizing curricula in early head start programs: Technical assistance paper, issue 4. Administration for Children and Families Office of Head Start, U.S. Department of Health and Human Services, Washington, DC.

Evans, C. A., & Porter, C. L. (2009). The emergence of mother–infant co-regulation during the first year: Links to infants' developmental status and attachment. *Infant Behavior and Development, 32*(2), 147–158. doi:10.1016/j.infbeh.2008.12.005.

Fallon, A., Westaway, J., & Moloney, C. (2008). A systematic review of psychometric evidence and expert opinion regarding the assessment of faecal incontinence in older community-dwelling adults. *International Journal of Evidence-Based Healthcare, 6*(2), 225–259. doi:10.1111/j.1744-1609.2008.00088.x.

Feldman, R. (1998). *Coding Interactive Behavior (CIB)*. Unpublished manual, Department of Psychology, Bar-Ilan University, Ramat-Gan, Israel.

Feldman, R., Eidelman, A. I., & Rotenberg, N. (2004). Parenting stress, infant emotion regulation, maternal sensitivity, and the cognitive development of triplets: A model for parent and child influences in a unique ecology. *Child Development, 75*(6), 1774–1791. doi:10.1111/j.1467-8624.2004.00816.x.

Fiese, B. H., Irwin, M., Gordon, M., & Poehlmann, J. (2001). *PIPE training manual* (1st ed.). Unpublished manual, Syracuse University, Syracuse, NY.

Fiori-Cowley, A., Murray, L., & Gunning, M. (2000). *Global ratings of mother–infant interaction at 2 and 4 months* (2nd ed.). Unpublished manual, Winnicott Research Unit, Department of Psychology, University of Reading, Reading, England.

Fogel, A., De Koyer, I., Secrist, C., Sipherd, A., Hafen, T., & Fricke, M. (2003). *The Revised Relational Coding System*. Unpublished manual, Department of Psychology, University of Utah, Salt Lake City, UT.

Gardner, F. (2000). Methodological issues in the direct observation of parent–child interaction: Do observational findings reflect the natural behavior of participants? *Clinical Child and Family Psychology Review, 3*(3), 185–198. doi:10.1023/A:1009503409699.

Gartstein, M. A., Crawford, J., & Robertson, C. D. (2008). Early markers of language and attention: Mutual contributions and the impact of parent–infant interactions. *Child Psychiatry and Human Development, 39*(1), 9–26. doi:10.1007/s10578-007-0067-4.

Goldman, B. A., Egelson, P. E., & Mitchell, D. F. (2008). *Directory of unpublished experimental mental measures* (Vol. 9). Washington, D.C.: American Psychological Association.

Greenspan, S. I., DeGangi, G., & Wieder, S. (2001). *The Functional Emotional Assessment Scale (FEAS) for infancy and early childhood: Clinical and research applications*. Bethesda, MD: Interdisciplinary Council on Developmental & Learning Disorders.

Greenspan, S. I., & Lieberman, A. F. (1989). Infants, mothers, and their interaction: A quantitative clinical approach to developmental assessment. In S. I. Greenspan & G. H. Pollock (Eds.), *The course of life. Vol. 1: Infancy* (pp. 503–560). Madison, CT: International Universities Press.

Grimes, D. A., & Schulz, K. F. (2002). An overview of clinical research: The lay of the land. *The Lancet, 359*(9300), 57–61. doi:10.1016/S0140-6736(02)07283-5.

Groth-Marnat, G. (2009). *Handbook of psychological assessment* (5th ed.). Hoboken, NJ: Wiley.

Hagan, J., Shaw, J., & Duncan, P. (2008). *Bright futures: Guidelines for health supervision of infants, children, and adolescents* (3rd ed.). Elk Grove Village, IL: American Academy of Pediatrics.

Halle, T., Anderson, R., Blasberg, A., Chrisler, A., & Simkin, S. (2011). *Quality of caregiver–child interactions for infants and toddlers (QCCIIT): A review of the literature, OPRE 2011–25*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Halle, T., Vick Whittaker, J. E., & Anderson, R. (2010). *Quality in early childhood care and education settings: A compendium of measures*. Washington, DC: Child Trends. Prepared by Child Trends for the Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Harrison, M. J., Magill-Evans, J., & Benzies, K. (1999). Fathers' scores on the Nursing Child Assessment Teaching Scale: Are they different from those of mothers? *Journal of Pediatric Nursing, 14*(4), 248–254. doi:10.1016/S0882-5963(99)80019-0.

Hofheimer, J., & Appelbaum, M. (1992). *Risk detection using observations of interaction (Report No. 8MCH/CCS-92/06)*. Springfield, VA: National Technical Information Service.

Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., & Jasnow, M. D. (2001). Coordinated interpersonal timing (CIT) at age 4 months. *Monographs of the Society for Research in Child Development, 66*(2), 66–78. doi:10.1111/1540-5834.00141.

Keyser, D. J., & Sweetland, R. C. (1984). *Test critiques* (Vol. 1). Kansas City, MO: Test Corporation of America.

Kirkpatrick, D. L. (1967). Evaluation of training. In R. Craig & L. Bittel (Eds.), *Training and development handbook* (pp. 87–112). New York, NY: McGraw-Hill.

Kochanska, G., Aksan, N., & Carlson, J. J. (2005a). Temperament, relationships, and young children's receptive cooperation with their parents. *Developmental Psychology, 41*(4), 648–660. doi:10.1037/0012-1649.41.4.648.

Kochanska, G., Forman, D. R., Aksan, N., & Dunbar, S. B. (2005b). Pathways to conscience: Early mother–child mutually responsive orientation and children's moral emotion, conduct, and cognition. *Journal of Child Psychology and Psychiatry, 46*(1), 19–34. doi:10.1111/j.1469-7610.2004.00348.x.

Kumar, R., & Hipwell, A. E. (1996). Development of a clinical rating scale to assess mother–infant interaction in a psychiatric mother and baby unit. *British Journal of Psychiatry, 169*(1), 18–26. doi:10.1192/bjp.169.1.18.

Maddox, T. (1997). *Tests: A comprehensive reference for assessments in psychology, education, and business*. Austin, TX: Pro-Ed.

Mahoney, G., Spiker, D., & Boyce, G. (1996). Clinical assessments of parent–child interaction: Are professionals ready to implement this practice? *Topics in Early Childhood Special Education, 16*(1), 26–50. doi:10.1177/027112149601600105.

Margolin, G., Oliver, P. H., Gordis, E. B., O'Hearn, H. G., Medina, A. M., Ghosh, C. M., & Morland, L. (1998). The nuts and bolts of behavioral observation of marital and family interaction. *Clinical Child and Family Psychology Review, 1*(4), 195–213. doi:10.1023/A:1022608117322.

Massie, H., & Campbell, B. K. (1992). *The Massie–Campbell Scale of Mother–Infant Attachment Indicators during Stress: For use during the pediatric examination and other childcare situations*. Van Nuys, CA: Child Development Media.

McCloskey, G. (1990). Selecting and using early childhood rating scales. *Topics in Early Childhood Special Education, 10*(3), 39–64. doi:10.1177/027112149001000305.

Messick, S. (1995). Standards of validity and the validity of standards in performance asessment. *Educational Measurement: Issues and Practice, 14*(4), 5–8. doi:10.1111/j.1745-3992.1995.tb00881.x.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine, 151*(4), 264–269. doi:10.7326/0003-4819-151-4-200908180-00135.

Munson, L. J., & Odom, S. L. (1996). Review of rating scales that measure parent–infant interaction. *Topics in Early Childhood*

Special Education, 16(1), 1–25. doi:10.1177/027112149601600104.

Nakamura, W. M., Stewart, K. B., & Tatarka, M. E. (2000). Assessing father–infant interactions using the NCAST teaching scale: A pilot study. *American Journal of Occupational Therapy, 54*(1), 44–51. doi:10.5014/ajot.54.1.44.

NICHD Early Child Care Research Network. (1997). The effects of infant child care on infant–mother attachment security: Results of the NICHD study of early child care. *Child Development, 68*(5), 860–879. doi:10.2307/1132038.

Oxford, M. L., & Findlay, D. M. (2012). *NCAST Caregiver/Parent–Child Interaction teaching manual*. Seattle, WA: NCAST Publication.

Pleck, E. H., & Pleck, J. H. (1997). Fatherhood ideals in the United States: Historical dimensions. In I. M. E. Lamb (Ed.), *The role of the father in child development* (3rd ed., pp. 66–103). New York: Wiley.

Price, G. M. (1983). Sensitivity in mother–infant interactions: The AMIS scale. *Infant Behavior and Development, 6*(3), 353–360. doi:10.1016/S0163-6383%2883%2980043-5.

Raack, C. B. (1989). *Mother/Infant Communication Screening (MICS)*. Roselle, IL: Community Therapy Services.

Ramchandani, P. G., Domoney, J., Sethna, V., Psychogiou, L., Vlachos, H., & Murray, L. (2013). Do early father–infant interactions predict the onset of externalising behaviours in young children? Findings from a longitudinal cohort study. *Journal of Child Psychology and Psychiatry, 54*(1), 56–64. doi:10.1111/j.1469-7610.2012.02583.x.

Schoepfel, J., Stock, C., Farace, D. J., & Frantzen, J. (2005). Citation analysis and grey literature: Stakeholders in the grey circuit. *The Grey Journal, 1*(1), 31–40.

Shepard, L., Kagan, S. L., & Wurtz, E. (1998). *Principles and recommendations for early childhood assessments*. Washington, DC: National Education Goals Panel.

Shonkoff, J. P., & Meisels, S. J. (2000). *Handbook of early childhood intervention* (2nd ed.). New York, NY: Cambridge University Press.

Snow, C. E., Van Hemel, S. B., Committee on Developmental Outcomes and Assessments for Young Children, & National Research Council. (2008). *Early childhood assessment: Why, what, and how*. Washington, DC: National Academies Press.

Stams, G. J., Juffer, F., & van Ijzendoorn, M. H. (2002). Maternal sensitivity, infant attachment, and temperament in early childhood predict adjustment in middle childhood: The case of adopted children and their biologically unrelated parents. *Developmental Psychology, 38*(5), 806–821. doi:10.1037/0012-1649.38.5.806.

Stright, A. D., Gallagher, K. C., & Kelley, K. (2008). Infant temperament moderates relations between maternal parenting in early childhood and children's adjustment in first grade. *Child Development, 79*(1), 186–200. doi:10.1111/j.1467-8624.2007.01119.x.

Sumner, G., & Spietz, A. (1994). *NCAST Caregiver/Parent–Infant Interaction feeding manual*. Seattle, WA: NCAST Publications.

Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development, 72*(3), 748–767. doi:10.1111/1467-8624.00313.

Tronick, E., Krafchuk, E., Ricks, M., & Winn, S. (1982). *Monadic Phase manual*. Unpublished manual, Child Development Unit, University of Massachusetts, Amherst, MA.

Walker, L. O., & Thompson, E. T. (1982). Mother–Infant Play Interaction Scale. In S. S. Humenick (Ed.), *Analysis of current assessment strategies in the health care of young children and childbearing families* (pp. 191–201). East Norwalk, CT: Appleton-Century-Crofts.

Wan, M. W., Green, J., Elsabbagh, M., Johnson, M., Charman, T., & Plummer, F. (2012). Quality of interaction between at-risk infants and caregiver at 12-15 months is associated with 3-year autism outcome. *Journal of Child Psychology and Psychiatry,*. doi:10.1111/jcpp.12032.

Washington State Office of Superintendent of Public Instruction (2008). *A guide to assessment in early childhood: Infancy to age eight*. Olympia, WA. Retrieved from http://www.k12.wa.us/EarlyLearning/pubdocs/assessment_print.pdf.

Zaslow, M. J. (2011). *Quality measurement in early childhood settings*. Baltimore, MD: Paul H. Brookes Publishing.

Zeanah, C. H. (2009). *Handbook of infant mental health*. New York, NY: Guilford Press.