

## Psychological tests for expectant parents and young children in the Nordic countries: A review of the evidence

Kirsi Peltonen, Marjo Kurki, Charlotte Reedtz, Sabine Kaiser, Lene-Mari P. Rasmussen, Marko Merikukka, Marte Rye, Taina Laajasalo, Henriette Kyrrestad, Piia Karjalainen, Susann Dahl Pettersen, Helene Eng, Kyrre Breivik & Monica Martinussen

To cite this article: Kirsi Peltonen, Marjo Kurki, Charlotte Reedtz, Sabine Kaiser, Lene-Mari P. Rasmussen, Marko Merikukka, Marte Rye, Taina Laajasalo, Henriette Kyrrestad, Piia Karjalainen, Susann Dahl Pettersen, Helene Eng, Kyrre Breivik & Monica Martinussen (2022): Psychological tests for expectant parents and young children in the Nordic countries: A review of the evidence, *European Journal of Developmental Psychology*, DOI: [10.1080/17405629.2022.2067141](https://doi.org/10.1080/17405629.2022.2067141)

To link to this article: <https://doi.org/10.1080/17405629.2022.2067141>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 16 May 2022.



[Submit your article to this journal](#)



Article views: 423








[View related articles](#)



[View Crossmark data](#)

## Psychological tests for expectant parents and young children in the Nordic countries: A review of the evidence

Kirsi Peltonen <sup>a</sup>, Marjo Kurki<sup>b</sup>, Charlotte Reedtz<sup>c</sup>, Sabine Kaiser <sup>c</sup>, Lene-Mari P. Rasmussen <sup>c</sup>, Marko Merikukka<sup>b</sup>, Marte Rye<sup>e</sup>, Taina Laajasalo<sup>d</sup>, Henriette Kyrrestad <sup>c</sup>, Piia Karjalainen<sup>b</sup>, Susann Dahl Pettersen<sup>c</sup>, Helene Eng<sup>c</sup>, Kyrre Breivik<sup>e</sup> and Monica Martinussen <sup>c</sup>

<sup>a</sup>Research Center for Child Psychiatry, University of Turku, Turku, Finland; <sup>b</sup>Department of Psychology, Itla Children's Foundation, Helsinki, Finland; <sup>c</sup>Regional Centre for Child and Youth Mental Health and Child Welfare, UIT The Arctic University of Norway, Tromsø, Norway; <sup>d</sup>University of Helsinki, Helsinki, Finland; <sup>e</sup>Regional Centre for Child and Youth Mental Health and Child Welfare, NORCE Norwegian Research Centre, Bergen, Norway

### ABSTRACT

The aim of this article is to review the psychometric properties of the psychological tests that are being used in the Nordic countries during pregnancy and the child's first two years of life. A systematic literature search was performed for 33 identified tests routinely used in health care, using specific review databases as well as PsycInfo, Embase and Medline. The trained authors conducted an overall assessment of quality for each test which was then verified by the editors. A total of 12% of the tests were rated at Level 1 'Low level of quality', 61% at Level 2 'Some evidence of Quality', 15% at level 3 'Good level of Quality' and 12% at Level 4 'High level of quality'. This indicates that the evidence for the psychometric properties is insufficient for many tests used in the Nordic countries for this purpose

**ARTICLE HISTORY** Received 23 June 2021; Accepted 7 April 2022

**KEYWORDS** Psychological tests; psychometrics; parents; babies; toddlers; pregnancy

Pregnancy and children's early development is closely monitored and screened in many countries in order to guarantee the well-being of the child and the parents (WHO, 2021). In the Nordic countries (i.e., in Norway, Sweden, Denmark, Finland, and Iceland) there has been a strong focus on early development, early interventions and increasingly also on the prevention of diseases and the promotion of health as the preferred strategies. Psychological tests are used to assess potential

**CONTACT** Kirsi Peltonen  [kirsi.peltonen@utu.fi](mailto:kirsi.peltonen@utu.fi)  Research Center for Child Psychiatry, University of Turku, Lemminkäisenkatu 3, Turku 20520, Finland

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.  
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

disturbances functioning in one or more areas of development. Tests are also needed in monitoring the changes in skills or functioning over time. Finally, tests can be used when offering recommendations or intervention programmes or when guiding parents and guardians around decisions concerning their children. In this study, we evaluate the psychometric properties of self-report measures, interviews and observations targeted at children 0–2 years old and their parents (including pregnancy).

Accurate measures are important in assessing risks and in identifying problems so that the correct interventions and services can be offered as early as possible and tailored efficiently. However, many measurement instruments are also routinely used without a real appreciation of their psychometric quality, potentially leading to biased observations and subsequently to poor practices and decisions regarding the parents and their child.

The aim of this article is to provide an overview of the psychological tests that are typically used in national health services in the Nordic countries during pregnancy and the child's first two years of life. The article research is based on a project called 'The First 1000 Days in the Nordic Countries' which is a three-year Nordic collaborative project initiated in 2019 as part of the Icelandic Presidency of the Nordic Council of Ministers (Danielsdottir & Ingudottir, 2020). The project focuses on the child's life, from the prenatal period until it is two years of age. Among other things, it aims to answer questions around how Nordic countries promote mental health and well-being during pregnancy and how they identify and respond to early risk factors in infants and toddlers and their families. The project is managed by the Directorate of Health in Iceland in partnership with The Danish Health Authority, the Public Health Agency of Sweden, The Norwegian Directorate of Health, The Regional Centre for Child and Youth Mental Health Øst and Sør in Norway and Itla Children's Foundation and the Finnish Institute for Health and Welfare in Finland.

We will use the generic term 'test' for all types of standardised devices, instruments, and procedures in which examinee behaviour is evaluated and scored using a standardised process (American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME), 2014; Federation of Psychologists' Associations guidelines (European Federation of Psychologists' Association (EFPA), 2013).

## Test quality: Reliability, validity, and norms

There is already an active debate ongoing in relation to insufficient test quality, especially regarding measures for children (Paap et al., 2018). To ensure that only high-quality tools are used in the health care system or other services for children and families, the system must define clear standards in terms of assessment quality.

American Educational Research Association AERA (2014) and European Federation of Psychologists' Association European Federation of Psychologists' Association (EFPA) (2013) guidelines emphasise the importance of reliability, validity and norms when evaluating the quality of tests. The nature and content of studies depend on the intended use of the test and the decisions that are intended to be made based on the test scores. Moreover, the language and cultural adaption of tests should be considered and examined locally, as translations and adaptations may impact the psychometric properties of tests.

A general definition of test reliability is the ratio of true variance over observed variance. The observed variance includes the true variance as well as random and/or systematic error. In Classical Test Theory the observed score is the sum of the true score and the error score and the reliability of a test is defined as the correlation between two parallel tests (Miller, 2010). Test reliability may be estimated in several ways, including test-retest reliability (test stability over time) internal consistency (split-half and Cronbach's alpha), or consistency across versions (parallel form) and finally inter-rater reliability (subjective effects on observations of behaviour). As the traditional methods attempt to measure each person's average response levels, modern test theory (Item Response Theory) reformulates the problem in terms of estimating the probability that a person's response to an item will lie in a particular category (Embretson & Reise, 2000).

Test validity refers to the degree to which the theory and evidence support the interpretation of test scores for the intended use of a test (Kane, 2013). For example, if a test is intended to be used for screening for healthy development among children in Western countries, evidence that supports this claim in age- and geographically specific ways, is needed. If a test is used for assessing a specific construct, such as developmental level or language skills, evidence comparing the test scores with other tests measuring the same construct, or examining the relationships between the test and other related constructs is needed. Construct validity signifies the extent to which a test measures the intended

construct (e.g., depression or attachment). Criterion-related validity signifies the extent to which a test can be used to predict the presence of a concurrent condition such as depression (criterion-related validity), or a future outcome (predictive validity). When the test is used for screening, it is important that it is accurate in its ability to identify whether a condition, for example depression, is likely to be present or absent. Based on a reference standard (e.g., a diagnostic classification or a defined outcome measure), *sensitivity* refers to the test's ability to correctly classify an individual as having the condition (true positive) while at the same time avoiding classifying individuals with a condition they do not have (false positive). *Specificity* refers to the ability of the test to correctly identify an individual as not having the condition (true negative).

Test norms are developed to compare a child or adult's performance with others. The raw score itself often has little meaning unless it is compared to something, such as the developmental level or a well-known comparison group. In order to do this, it is important to have a norm group that is relevant to the use of the test. Norms may differ between countries and cultures and should therefore be based either on local samples, or on evidence that confirms that the original norms may be used in the new settings.

Different kinds of studies can be used in assessing the aspects of test quality. Studies that are explicitly designed to examine the psychometric properties of the test are most important, but empirical studies where the test has been used may also be relevant as they provide information about the test's construct validity. Several studies, preferably meta-analyses, are thus required to study the extent to which psychometric properties generalise across different settings and groups (Schmidt, 2010).

In this study we aim to collect all of the existing evidence for the evaluation of the tests used during the first 1000 days of a child's life in the Nordic countries. We also aim to evaluate information on reliability, validity and norms in order to assess their quality. The rapid systematic reviews were conducted in relation to the 'First 1000 Days in the Nordic Countries' -project. The Regional Centre for Child and Youth Mental Health and Child Welfare (RKBU North) at The Arctic University of Norway and the Itla Children's Foundation were asked to assess the evidence base for the identified psychosocial programmes and psychological tests provided by each participating country.

## **Research aim**

The aim of this article is to evaluate the evidence base for the psychological tests that are being used in the Nordic countries among expectant parents and children up to the age of two years. This overall evaluation is based on rapid systematic reviews that have been conducted for each psychological test and published in a report [blinded for review].

## **Method**

### ***Inclusion criteria for tests***

Each participating country, Denmark, Finland, Iceland, Norway and Sweden, provided information on the relevant tests used in their respective country. The partners in the First 1000 Days project compiled a list of the tests routinely used in their countries. The project coordinator then performed a thorough review of this information, finding additional information about each intervention and test to ensure its suitability for further review.

Tests were included for further elaboration if they were: (1) tests, instruments, scales, observational methods, or diagnostic systems for assessing parental risk factors including mental health problems, drug and alcohol use, domestic violence, parenting styles, parenting and stress among pregnant women, parents and caregivers of young children (0–2 years of age); or (2) tests, instruments, scales, observational methods and diagnostic systems for assessing social, emotional, cognitive and motoric development, attachment, developmental problems, social withdrawal, or autism among young children (0–2 years); and (3) available in one or more Nordic languages. Tests were excluded if (1) they were unstandardised or without a written description (including information about items, administration and scoring); or (2) the target group did not include young children, pregnant women, parents, or caretakers of young children. We refer to the target group of the test as ‘adults’ if the test is generic and used routinely across populations but also to address parental risk factors. Assessment procedures that are not standardised, or that include a combination of many methods are considered to be outside the scope of this review.

The initial list included a total of 47 different tests suggested by the five participating countries. Only three of the proposed tests were used in more than one country. A total of 13 suggested tests were excluded due

to the inclusion or exclusion criteria mostly because they could not be defined as tests (lack of standardisation or a fixed set of questions). Some were more akin to methods or procedures and for a few, a written description was not publicly available. The remaining 33 were then included in the review.

Tests fulfilling all of the inclusion criteria were included in this review. We used a Rapid Review Method, meaning that the timeframe for conducting the individual review was limited to  $\leq 5$  weeks and the sources used for the literature search were limited due to time constraints (Grant & Booth, 2009). The process did however use transparent and reproducible search methods. The Rapid Review approach is a justified method when the research question is relatively narrow (here, the quality of a certain test) and the selection of studies is based on inclusion/exclusion criteria (Khangura et al., 2012).

### *Literature search*

The literature search was conducted in three steps. First, the project partners who provided the names of the tests were asked to provide additional information about each test. The project coordinator then performed a thorough review of this information, finding additional information about each test to ensure its suitability for further review. Second, a search was conducted of those databases that summarise the evidence of tests. The relevant databases searched were PsykTestBarn.no in Norway and Metodguiden in Sweden. Denmark, Finland and Iceland did not have similar databases. Third, a systematic search was performed in the PsycInfo, Embase, and Medline databases. The search strategy included nine steps. In the first five steps, searches were limited to the Nordic countries (i.e., Norway, Sweden, Denmark, Finland or Iceland) and larger cities in those countries (e.g., 'Oslo or Bergen or Trondheim or NTNU or Tromso or Tromsø or Stavanger', in case the original article used the name of the city instead of the country when describing the data source). In step 6, these searches were combined. In step 7, a search was conducted for the name of the psychological test and/or its abbreviated name. In step 8, the search results from step 6 and step 7 were combined. In the final step, duplicates were removed. The total number of hits varied from test to test.

Studies revealed by the literature review were included in the final evaluation if they: (1) examined the psychometric properties (e.g., reliability, validity or norms) of tests (including scales, measurement

instruments, observational methods or diagnostic systems), (2) included children under 2 years old, or their parents as participants, (3) were conducted in the Nordic countries (in case no Nordic studies, international studies from Europe and North America were included) and (4) published in peer-reviewed journals, as part of PhD dissertations or as reviews in PsykTestBarn or Metodguiden.

### *Rating the quality of tests*

To rate the quality of psychological tests, criteria were based on the methods used by EFPA and the online journal PsykTestBarn. Tests were assessed on four levels ranging from 'Level 1. Test with no or a low level of quality' to 'Level 4. Tests with a high level of quality'. In general, a higher rating required more studies in addition to higher quality studies, as well as information supporting the quality of the psychometric properties (reliability, validity and norms) of the test for its intended use in the Nordic countries. An overview of the criteria for the four different levels is presented in [Table 1](#).

**Table 1.** Criteria for different levels of quality.

	Level 1. Tests with No or a Low Level of Quality	Level 2. Tests with Some Quality	Level 3. Tests with a Good Level of Quality	Level 4. Tests with a High Level of Quality
Number and quality of studies concerning the psychometric properties	No Nordic or international (European or North American) studies  OR	a. Only international studies OR b. Nordic studies  AND	One Nordic study  AND	At least one Nordic study  AND
Findings the studies from	No support of the test's psychometric properties	For a. at least adequate reliability, validity and norms  For b. inadequate reliability, validity and norms	at least adequate reliability, validity and norms  if used for screening, sensitivity and specificity at least adequate	good or excellent levels of reliability, validity and norms  if used for screening, sensitivity and specificity good or excellent



Based on this classification two authors wrote a review for each psychosocial test. The group of authors consisted of 14 researchers, all with high expertise in research on child development and clinical work among children and families. Nine authors were from Norway and five from Finland. In cases where the authors identified a lack of relevant studies, the systematic review was supplemented by a manual search. If the studies lacked the critical details of test features, the developers of the test were contacted to complete the evaluation. After both authors approved the draft, it was submitted to one of the two editors of the report [blinded for review], who reviewed it and sent it back to the authors who made the necessary changes and then resubmitted it. This procedure was repeated until a final version was accepted by the editors.

## Results

A total of 33 tests were evaluated. The detailed reviews of the tests are presented in 'The First 1000 Days in the Nordic Countries' – report [blinded for review].

### *Descriptive information*

The majority of tests were self-report measures (15 of 33 = 45%), with the remaining tests classified as interviews (7 of 33 = 21%), observations (8 of 33 = 24%), and a combination of observation and interview (1 of 33 = 3%). Two tests included equipment/stimulus material (6%). Children alone were the target group in nine of the tests (27%; [Table 2](#)), whereas children and caretakers were the target group in six tests and children and teachers in one test (21%; [Table 3](#)). Mothers, fathers, or parents/caregivers were the target group in 11 tests (33%). Six tests (18%) were targeted to the general adult population, addressing parental factors that may pose a risk for optimal child development and well-being and were therefore included in the review (target group categorised as 'Adults', [Table 4](#)). Five of them addressed mental health symptoms and one harmful or hazardous drinking habits. The majority of the tests used for both children and adults originated from outside the Nordic countries and the psychometric properties related to the Nordic version(s) were examined only to some degree [blinded for review]. A total of 20 tests (61%) were classified as screening instruments. All reviewed tests are presented in [Tables 2–4](#).

**Table 2.** Overview of tests for children and their quality ratings.

Name of test	Purpose	Type of test	Used for screening	Quality rating	Evidence of reliability, validity, and norms (Reviewed articles)
Achenbach System of Empirically Based Assessment (ASEBA) Preschool	To assess social skills, emotional and behavioral difficulties, in addition to language delays.	Questionnaire	No	2	Evidence of construct validity is lacking for the age group 0–2 years. No studies on the language sub-test (LDS) were found. (Kristensen et al., 2010)
Ages and Stages Questionnaires (ASQ)	To screen for developmental level including communication, fine and gross motor skills, problem solving, and personal/social development.	Questionnaire	Yes	3	There is some documentation of the reliability of the test, especially the total score. There are also studies supporting the construct validity, and norms exist for some languages. Since the intended test use is screening, more evidence to support this use is needed. (Martinussen & Valla, 2013; Wang et al., 2012; Valla et al., 2017; Østergaard et al., 2012; Vedel et al., 2020)
Ages and Stages Questionnaire: Social and Emotional (ASQ:SE)	To screen for and assess parent-reported social and emotional difficulties in children.	Questionnaire	Yes	2	The reliability seems satisfactory, but there is a lack of Nordic up-to-date norming and validation studies. (Rasmussen & Martinussen, 2013; Vaezghasemi et al., 2020; Eurenius et al., 2019; Squires et al., 2015; Pontoppidan et al., 2017; Veilikonja et al., 2016)
Alarm Distress Baby Scale (ADBB)	To screen for infant social withdrawal behavior, to detect signs of congenital, attachment, and relational difficulties.	Observation	Yes	3	Overall good to excellent inter rater reliability. No Nordic articles report on internal consistency and there are no norm studies. Evaluations of the construct validity are based on associations with other instruments and differences between groups. Sample sizes of the included studies were relatively small, especially for some of the samples used for estimating inter-rater reliability. (Moe et al., 2016; Smith-Nielsen et al., 2019; Puura et al., 2007, Puura et al., 2019, Puura et al., 2013, Puura et al., 2010; Guedeney & Fermanian, 2001)

(Continued)



Table 2. (Continued).

Name of test	Purpose	Type of test	Used for screening	Quality rating	Evidence of reliability, validity, and norms (Reviewed articles)
Bayley Scales of Infant and Toddler Development (BSID, BSID-II, BSID-III)	To assess developmental level.	Test incl. equipment	No	2	There are no studies reporting reliability or norms based on Nordic samples, but there are some studies supporting the construct validity of the test. (Richter & Valla, 2013; Månsson & Stjernqvist, 2014; Krogh et al., 2012; Kahr Nilsson et al., 2019; Skovgaard et al., 2008; Sajaniemi et al., 2001)
BOEL (Blik Orienteret Efter Lyd; Glance Oriented After Sound)	To test the child's interaction, attention and reaction to visual and sound stimuli, in order to identify hearing and communication disorders.	Test incl. equipment	Yes	2	There are no studies reporting the inter-rater reliability. The sensitivity of the BOEL test seems to be rather low. (Rasting & Lindbaek, 1998; Jakobsen et al., 2007; Mortensen et al., 2003; Ravn & Bjerager, 2004; Barr et al., 1978; Barr & Stensland Junker, 1978; Huber et al., 1978)
Diagnostic Classification of Mental Health and Developmental Disorders of Infancy and Early Childhood (DC:0-5)	Diagnostic classification of mental health and developmental disorders in early childhood.	Interview	No	2	DC: 0-5 is not a traditional assessment tool but a diagnostic multi-axial classification framework that is based on previous versions of the system and work from task force groups combining clinical experience, empirical evidence, and theoretical understandings. (Moe & Mothander, 2009; Mothander & Moe, 2008, Mothander & Moe, 2010; Kahr Nilsson et al., 2019)
Modified Checklist for Autism in Toddlers (M-CHAT) and revised version (M-CHAT-R/F)	To identify early indicators of autism.	Checklist	Yes	3	The internal consistency of the measure is good, and the so is the specificity when predicting diagnoses. The sensitivity is more variable between studies. (Robins et al., 2001; Kleinman et al., 2008; Chlebowski et al., 2013; Nygren et al., 2012, Mimsalco et al., 2018; Robins et al., 2014)
Vane-psy [Vauvan psyykkinen ja neurologinen kehitys]	To monitor the development of infants and toddlers.	Observation/ interview	No	3	Sensitivity and specificity are shown to be adequate in one Nordic study, with adequate methodological quality. (Mustonen et al., 2006)

**Table 3.** Overview of tests for children and parents and their quality ratings.

Name of test	Purpose	Type of test	Used for screening	Quality rating	Evidence of Reliability, Validity and Norms (Related articles)
Ainsworth Strange Situation Procedure (SSP)	To measure mother-child attachment quality and to classify attachment security.	Observation	No	2	Both Nordic and international studies provide inconsistent findings for reliability and validity. No Nordic studies exist about the construct validity. (Heidi, 2015; Lamb et al., 1982; Smith-Nielsen et al., 2016; Hautamäki et al., 2008; Kouvo et al., 2003; Cadman et al., 2018; Van Dam and Van Ijzendoorn, 1988; Fonagy et al., 2016; Luyten et al., 2017; Carcamo et al., 2014)
CARE Index	To screen for and assess adult sensitivity in a dyadic context (child 0–24 months).	Observation	No	2	There are two Nordic studies supporting the inter-rater reliability of the scoring, and some support of the construct validity of the test. However, documentation supporting the use for screening purposes is needed. (Kristensen et al., 2010; Pajulo et al., 2011)
Classroom Assessment Scoring System (CLASS) Toddler	Observational instrument to assess emotional and behavioral support within the classroom.	Observation	No	2	Results revealed adequate measurement quality. However, one study supported a three-domain structure instead of the current two-domain structure in CLASS-T. More research is warranted, especially within the Nordic countries. (Drugli et al., 2018; Bandel et al., 2014; Castle et al., 2016; Thomason & La Paro, 2009; Slot et al., 2017)
Crowell Procedure	Observation of caregiver-child interaction.	Observation	No	2	There is a need for more research examining the psychometric properties of the Crowell Procedure both internationally and in the Nordic countries. (Heller et al., 1999; Loop et al., 2017; Sprang & Craig, 2014)
Lausanne Trilogue Play (LTP)	Observational instrument to assess co-parenting alliance and interactions with the child.	Observation	No	2	There is limited documentation of the different scoring systems used and small sample sizes. (Hedenbro & Rydélius, 2014; Hedenbro & Rydélius, 2019; Hedenbro & Lidén, 2002; Korja et al., 2015.; Korja et al., 2016; Favez et al., 2011)

*(Continued)*

Table 3. (Continued).

Name of test	Purpose	Type of test	Used for screening	Quality rating	Evidence of Reliability, Validity and Norms (Related articles)
Marschak Interaction Method (MIM)	To assess parent–child interaction (positive vs. problematic).	Observation	No	2	MIM-P has not been validated for children aged 0–3 years. The existence of several rating systems complicates an overall evaluation of validity and reliability, but there seems to be a need for more research on all ratings systems for the age group 0–3. (Salo & Mäkelä, <i>in press</i> )
Parent–Child Early Relational Assessment (PCERA)	To measure the quality of the parent–child relationship by assessing the affective and behavioral characteristics of their interaction.	Observation	No	2	PCERA has frequently been used in Scandinavian studies (18 publications located in the literature search), but none of the Scandinavian studies were psychometric studies; rather, they were studies where the instrument has been used as a measure of parent–infant interaction for the purposes of the study. Nevertheless, these studies give some limited information of the reliability and validity of the test. No norm studies utilizing Scandinavian samples were found. (Anke et al., 2019; Clark, 1985, Clark, 1999; Haapsamo et al., 2013; Korja et al., 2008; Korja et al., 2010; Lotzin et al., 2015; Misund et al., 2016; Savonlahti et al., 2005.)

**Table 4.** Overview of tests for parents or adults in general population and their quality ratings.

Name of test	Purpose	Type of test	Used for screening	Quality rating	Evidence of Reliability, Validity and Norms (Related articles)
Adverse Childhood Experiences (ACE)	To measure for adverse experiences before the age of 18	Self-report	No	2	No test-retest reliability studies among our target group, and only one study of internal consistency. In addition to this, a lot of different versions of the test have been used, which indicates a lack of standardization. ACE has been found to correlate with many adverse outcomes among mothers pre- and post-natal, but that cannot be considered as conclusive evidence of construct validity. The majority of the studies were conducted with American samples. (Rohder et al., 2019; Chung et al., 2010; Felitti et al., 1998; Murphy et al., 2014; Steele et al., 2016; Sun et al., 2017).
Alcohol Use Disorders Identification Test (AUDIT)	To screen for harmful or hazardous drinking habits.	Self-report	Yes	2	Overall adequate to good internal consistency. However, results about validity are inconclusive and there is no adequate documentation about sensitivity and specificity. (Selin, 2003; Lund et al., 2019; Lehtikoinen et al., 2016; Magnusson et al., 2007; Stene-Larsen et al., 2013; Comasco et al., 2012; Göransson et al., 2003).
Clinical Outcomes in Routine Evaluation – Outcome Measures (CORE-OM)	To assess mental health symptoms and problems before and after treatment.	Self-report	No	4	Results demonstrated good to excellent reliability and validity for all the Nordic languages examined for both clinical and non-clinical samples. (Honkalmapi et al., 2017; Skre et al., 2013; Kristjánsdóttir et al., 2015; Héðinsson et al., 2013; Elfström et al., 2012).
Domestic Abuse, Stalking and Honor-Based Violence (DASH)	To help identify those at high risk of harm from domestic abuse, and who should be referred to a Multi-Agency Risk Assessment Conference (MARAC) meeting in order to manage the risk.	Interview	Yes	1	Overall, there is limited knowledge about the reliability of DASH and the inter-rater reliability seems to be modest. Low predictive validity. (Chalkey & Strang, 2017; Robinson et al., 2016; Sebire & Barling, 2016; Thornton, 2017; and Turner et al., 2019)
Domestic Violence Filter and Mapping Form	To identify intimate partner and domestic violence	Interview and Questionnaire	No	1	No data on the reliability, validity or norms.

(Continued)



Table 4. (Continued).

Name of test	Purpose	Type of test	Used for screening	Quality rating	Evidence of Reliability, Validity and Norms (Related articles)
Depression and Anxiety Stress Scale (DASS)	To screen for depression, anxiety and stress.	Self-report	Yes	2	Limited information regarding the translation of the instruments is provided. Internationally, the instrument has been translated into multiple languages and investigated in both clinical and non-clinical samples (e.g., Brown et al., 1997; Crawford & Henry, 2003), and sound psychometric properties are supported. However, more research within the Nordic countries is needed due to the limited documentation of the psychometrics properties, especially the validity of the measure.
Edinburgh Postnatal Depression Scale (EPDS) – fathers	To measure and screen for depression (postnatal and perinatal).	Self-report	Yes	4	(Jonsdottir et al., 2017; Dahlerup et al., 2018; Salari et al., 2014; Brown et al., 1997; Crawford & Henry, 2003) The included studies were of good methodological quality and findings indicate at least adequate internal consistency. Relationships with other instruments were in the expected direction, which supports that the construct validity and documentation about sensitivity and specificity were good. The translation procedures have not been described. (Fredriksen et al., 2019; Edhborg, 2008; Johansson et al., 2017; Kerstis et al., 2013; Kerstis, Nohler, et al., 2016; Kerstis, Aarts, et al., 2016; Massoudi et al., 2013; Mörelius et al., 2015; Seimyr et al., 2009; Karukivi et al., 2015).

(Continued)

**Table 4. (Continued).**

Name of test	Purpose	Type of test	Used for screening	Quality rating	Evidence of Reliability, Validity and Norms (Related articles)
Edinburgh Postnatal Depression Scale (EPDS) – mothers	To measure and screen for depression (postnatal and perinatal).	Self-report	Yes	4	The studies and reviews of EPDS indicated good internal consistency and test-retest reliability. Construct validity in terms of relationships with other relevant instruments was high and in the expected direction, and documentation about sensitivity and specificity was good but varied between studies depending on the cut-off score used, diagnostic method, and possibly sample (post- or perinatal). (Larun et al., 2013; Hanssen-Bauer & Welander-Yatn, 2012; Eberhard-Gran et al., 2001; Rubertsson et al., 2011; Smith-Nielsen et al., 2018; Lydsdottir, Howard, Olafsdottir, Thome, et al.)
Generalized Anxiety Disorder Scale 2-item (GAD-2)	To screen for generalized anxiety disorder (GAD), and other anxiety disorders.	Self-report	Yes	2	Results were acceptable, but variable, indicating that more research is warranted regarding the accuracy of the GAD-2. Overall, there is limited evidence of psychometric properties in the Nordic countries. Kujanpää et al., 2014; Berge et al., 2019; Plummer et al., 2016)
Generalized Anxiety Disorder Scale 7-item (GAD-7)	To screen for generalized anxiety disorder (GAD), and other anxiety disorders.	Self-report	Yes	4	There is adequate support for the psychometric properties of the test, especially the Finnish language version. There is still a lack of research based on pregnant and postpartum women. (Knapstad et al., 2020; Rozentel et al., 2018; Kujanpää et al., 2014; Tiirikainen et al., 2019; Malmberg-Ceder et al., 2019)

(Continued)





Table 4. (Continued).

Name of test	Purpose	Type of test	Used for screening	Quality rating	Evidence of Reliability, Validity and Norms (Related articles)
Parenting Stress Index (PSI)	To assess parental stress.	Self-report	Yes	3	The majority of Nordic studies are probably conducted on the third edition of PSI, and not the most recent fourth edition. There is considerable evidence supporting the reliability of the test, some studies supporting the construct validity, but norms seem to be lacking for the Nordic languages. (2011; Huhtala et al., 2012; Korja et al., 2014; Landsem et al., 2014; Östberg et al., 1997; Sarfi et al., 2013)
Parents' Evaluation of Development Status (PEDS®)	To assess children's development in language, motor skills, self-help, early academic skills, behavior and social-emotional/mental health.	Questionnaire	Yes	2	Overall, there was lack of documentation. No Nordic studies were identified, but according to the test's webpage, an evaluation of PEDS is ongoing in Iceland. A list of results from the manual implies that PEDS was re-standardized and re-validated in a representative sample in the US and Canada, and different reliability and validity assessments have been conducted; however, no documentation confirming this has been possible to retrieve. Glascoe et al. (2019) reported a sensitivity of 74–96% and specificity of 73–83% for PEDS, which was based on 44 published studies where PEDS was administered correctly, but no information regarding these studies is given, including where they are published. (Glascoe et al., 2019)
Resource form/barometer	To assess resources and strain factors of first-time mothers and fathers.	Self-report	Yes	1	Only the usefulness and necessity of the items were evaluated, not the psychometrics of the questionnaire as such. (Kajlunen et al., 2005)
TWEAK Alcohol Screening Test	To identify pregnant women with harmful drinking habits.	Interview	Yes	2	There were no norms or reliability estimates reported in any of the studies. This, combined with the relatively low sensitivity estimates, suggests insufficient documentation of the psychometric properties. (Præstegaard et al., 2018; Weile et al., 2020)

(Continued)



Table 4. (Continued).

Name of test	Purpose	Type of test	Used for screening	Quality rating	Evidence of Reliability, Validity and Norms (Related articles)
Vane-psy [Vauvan psykkinen ja neurologinen kehitys]	To monitor the development of infants and toddlers.	Observation/ interview	No	3	Sensitivity and specificity are shown to be adequate in one Nordic study, with adequate methodological quality. (Mustonen et al., 2006)
VAVU – Interview to support early parent–child interaction [Varhaista vuorovaikutusta tukeva haastattelu]	To help identify and address perceptions, concerns, and potential difficulties related to pregnancy, childbirth, and the baby.	Interview	Yes	1	There are no Nordic or international (European or North American) studies examining the psychometric properties
Whooley Questions	To screen for depression (pre- and postnatal).	Self-report	Yes	2	There is a lack of Nordic studies for our target group. No studies reported information regarding test reliability, and the diagnostic accuracy varied a lot between studies, Sulja et al., 2012; Bosanquet et al., 2015; Darwin et al., 2016; Howard et al., 2018; Littlewood et al., 2018)
Working Model of the Child Interview (WMCI)	To assess parents' internal representations (working models) of their relationship to a particular child.	Interview	No	2	There are several international studies providing information on psychometric properties of the WMCI, including a systematic review. However only two Scandinavian (Finnish) studies describing limited psychometric data of the WMCI were found. There were no norm studies based on Scandinavian samples. (Korja et al., 2010; Benoit, Parker et al., 1997; Theran et al., 2005; Vreeswijk et al., 2012; Ainsworth et al., 1978; Zeanah et al., 1994; Main et al., 1985; Zeanah et al., 1994; Korja et al., 2010).

### Quality of the tests

Of the 33 psychological tests reviewed, 4 (12%) were rated at level 1, 20 (61%) at level 2, 5 (15%) at level 3, and 4 (12%) at level 4. The mean level of quality was 2.27 ( $SD = 0.83$ ). There were no significant differences in the mean level of quality between the screening instruments and other instruments ( $M_{screen} = 2.3$ ,  $SD_{screen} = 0.98$  vs  $M_{not} = 2.2$ ,  $SD_{not} = 0.60$ ,  $t = -.23$ ,  $p > .05$ ). All tests that had the highest quality ratings were population level tests, meaning that they assessed risk factors among adults in the general population but were also routinely used to assess parents of young children. The Edinburgh Postnatal Depression Scale (EPDS) – for mothers and fathers was an exception here, being specifically targeted to pregnant families and scoring high (level 4) in the quality rating.

### Discussion

The overall aim of this article was to assess the evidence base for the psychosocial tests that are being used in the Nordic countries during the first 1000 days of a child's life – that is, during the mother's pregnancy and the first two years after birth. The final list included 33 psychological tests. The mean level of quality ratings reflected the overall picture that there was 'some evidence of quality' (level 2) in the tests. Over half of the tests fell into that category, whereas a quarter of tests fell into the category 'good level of quality' (level 3) and less than a quarter in the categories 'no/low level of quality' (level 1) or 'high level of quality' (level 4).

The tests with the highest level of quality were those that are routinely used to assess adult mental health such as anxiety and depression in the general population, but that are also frequently used particularly in assessing parental mental health. The mental health problems depicted in these tests are contextualised as parental risk factors. On the other hand, there was a clear lack of studies in which the test scores were examined in relation to more objective criteria, for example a clinical diagnosis or an outcome. For other risk factors such as alcohol and drug use, as well as domestic violence, no sufficient documentation emerged in terms of the predictive validity of the tests. These tests were well validated for risk assessment purposes in the adult population in general, but the evidence was scarce in respect of parents of small children. As the level of knowledge about the

detrimental effects of poor parental mental health and substance abuse on child development is increasing (Goldman Fraser et al., 2010; Goodman et al., 2016; Parfitt et al., 2013) the tests assessing these factors should have high sensitivity and specificity in the given population.

The majority of the tests used for both children and adults originated from outside the Nordic countries and the psychometric properties related to the Nordic version(s) were examined only to some degree. In this review, in order to be rated as the highest levels of quality, the tests were required to have psychometric evidence based on Nordic samples. The documentation about the translation process and cultural adaptation was lacking for most of the included tests, while Nordic norms were also frequently missing. It may very well be that these tests have good psychometric properties, but Nordic evaluations should be conducted in future. Most tests are developed in one specific language and in a specific cultural context. When a test is translated into another language or used in a different culture, assessing the psychometric properties of the new version is important. Even a carefully conducted translation process may result in a focus on different psychometric properties due to various cultural factors.

There was significant variation in terms of professional requirements (formal education, training and supervision) for those administering and interpreting the tests. Some were mainly intended to be used in a psychological/medical examination, whereas others were intended to be used more freely in different settings. Many tests were also used in relation to specific interventions, as part of the inclusion/exclusion process. Therefore, the training of those who apply the test should cover both the technical administration as well as an understanding of the intervention aims and processes that are closely related to the test results. Most tests require some level of user qualification and training before they can be administered and interpreted correctly. The publishers of the tests often have specific requirements regarding formal education and competencies before a person is allowed to purchase a test, for example, that the prospective test user needs to be a psychologist. Other tests may be freely available, with no specific requirements for training or formal education. It is however always the user's responsibility to ensure that he or she has the necessary qualifications and knowledge in psychometric testing and that the test is used safely.

The studies included in the review varied in terms of the participation of the test developers. It is not unusual for the developer to conduct the first psychometric studies, but it is always desirable that further studies are conducted by independent researchers. Among other things related to the accumulation of evidence in quality assessments of the psychological tests, this is an important point of development for the future.

## Limitations

Rapid systematic reviews are somewhat more vulnerable to bias and error than other systematic reviews since the search for existing studies may be less comprehensive. In our case, the timeframe for conducting the individual reviews was limited to  $\leq 5$  weeks, and the sources were limited due to time constraints in relation to searching (Grant & Booth, 2009).

The search process in respect of the tests was carefully guided and monitored by the editors, but it is still possible that there were minor differences across those who rated the tests in terms of how they were searched and identified. If the authors of the evaluation identified a lack of relevant studies in the systematic search, they supplemented the search with a manual search based on their own professional experience. This may have led to variations in the reviews and possibly the omission of single tests. To increase consistency, we established criteria for the inclusion and exclusion of studies and for how the evidence should be rated. All of the authors participated in the same training to facilitate a joint understanding of the criteria and procedures. As an additional step in terms of quality assurance, all reviews were reviewed and approved by one of the editors.

## Conclusion and recommendations

Psychosocial tests with good psychometric properties are important as the test scores are used in making important decisions related to an individual's health. Our rapid systematic review showed that the evidence in respect of psychometric properties is insufficient for many tests currently available for use in pregnancy and the child's first two years in the Nordic countries. There is then a clear need for more high-quality studies evaluating the psychometric properties of tests

used during pregnancy and the child's first two years and this should be strongly prioritised in the future.

## Acknowledgments

The work is funded by The Nordic Council of Ministers, UiT The Arctic University of Norway and Itla Children's Foundation, Finland.

## Data availability statement

The data that support the findings of this study are openly available in <https://pub.norden.org/nord2021-037/>

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by the The Nordic Council of Ministers [-]; UiT The Arctic University of Norway [-]; Itla Children's Foundation [-].

## ORCID

Kirsi Peltonen  <http://orcid.org/0000-0002-5112-3823>

Sabine Kaiser  <http://orcid.org/0000-0002-2081-7734>

Lene-Mari P. Rasmussen  <http://orcid.org/0000-0002-7092-2495>

Henriette Kyrrestad  <http://orcid.org/0000-0003-1515-6502>

Monica Martinussen  <http://orcid.org/0000-0002-0937-1479>

## References

- Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the Strange Situation*. Erlbaum.
- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Anke, T., Slinning, K., Moe, V., Brunborg, C., Siqveland, T. S., & Skjelstad, D. V. (2019). Mothers with and without bipolar disorder and their infants: Group differences in mother–infant interaction patterns at three months postpartum. *BMC Psychiatry*, 19 (1), 292. <https://doi.org/10.1186/s12888-019-2275-4>

- Bandel, E., Aikens, N., Vogel, C. A., Boller, K., & Murphy, L. (2014). *Observed quality and - psychometric properties of the CLASS-T in the Early Head Start Family and Child Experiences Survey*. Econ Papers Home. <https://EconPapers.repec.org/RePEc:mpr:mprres:1017ea36a7044079911975a2123bc79e>
- Barr, B., Junker, K. S., & Svärd, M. (1978). Early discovery of hearing impairment: A critical evaluation of the BOEL test. *Audiology*, 17(1), 62–67. <https://doi.org/10.1080/00206097809080032>
- Barr, B., & Stensland Junker, K. (1978). BOEL screening for early discovery of deafness. *Scandinavian Audiology*, (Suppl.um(Suppl 8)), 9–12. <https://doi.org/10.1159/000399024>.
- Benoit, D., Parker, K. C., & Zeanah, C. H. (1997). . *Journal of Child Psychology and Psychiatry* Mothers' representations of their infants assessed prenatally: Stability and association with infants' attachment classifications. 38(3), 307–313.
- Berge, T., Bull-Hansen, B., Solberg, E. E., Heyerdahl, E. R., Jorgensen, K. N., Vinge, L. E., Aaronaes, M., Oie, E., & Hyldmo, I. (2019). Screening for symptoms of depression and anxiety in a cardiology department. *Tidsskrift for Den Norske*, 46(2), 153–159. <https://doi.org/10.1002/ppul.21334>
- Bosanquet, K., Bailey, D., Gilbody, S., Harden, M., Manea, L., Nutbrown, S., & McMillan, D. (2015). Diagnostic accuracy of the Whooley Questions for the identification of depression: A diagnostic meta-analysis. *BMJ Open*, 9(5), e008913. <https://doi.org/10.1136/bmjopen-2015-008913>.
- Brown, T. A., Chorpita, B. F., Korotitsch, W., & Barlow, D. H. (1997). Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. *Behavior Research and Therapy*, 35(1), 79–89. [https://doi.org/10.1016/s0005-7967\(96\)00068-x](https://doi.org/10.1016/s0005-7967(96)00068-x).
- Cadman, T., Belsky, J., & Fearon, R. M. (2018). The brief attachment scale (bas-16): A short measure of infant attachment. *Child: Care, Health and Development*, 44(5), 766–775. <https://doi.org/10.1111/cch.12599>.
- Carcamo, R. A., Ijzendoorn, M. H., Vermeer, H. J., & van der Veer, R. (2014). The validity of the Massie-Campbell attachment during stress scale (ADS). *Journal of Child and Family Studies*, 23(5), 767–775. <https://doi.org/10.1007/s10826-013-9728-z>
- Castle, S., Williamson, A. C., Young, E., Stubblefield, J., Laurin, D., & Pearce, N. (2016). Teacher–child interactions in early head start classrooms: Associations with teacher characteristics. *Early Education and Development*, 27(2), 259–274. <https://doi.org/10.1080/10409289.2016.1102017>
- Chalkley, R., & Strang, H. (2017). Predicting domestic homicide and serious violence in Dorset. *Cambridge Journal of Evidenced-Based Policing*, 1, 81–92.
- Chlebowski, C., Robins, D. L., Barton, M. L., & Fein, D. (2013). Large-scale use of the Modified Checklist for Autism in low-risk toddlers. *Pediatrics*, 131(4), e1121–e1127. <https://doi.org/10.1542/peds.2012-1525>
- Chung, E. K., Nurmohamed, L., Mathew, L., Elo, I. T., Coyne, J. C., & Culhane, J. F. (2010). Risky health behaviors among mothers-to-be: The impact of adverse childhood experiences. *Academic Pediatrics*, 10(4), 245–251. <https://doi.org/10.1016/j.acap.2010.04.003>
- Clark, R. (1985). *The Parent-Child Early Relational Assessment. Instrument and manual*. Department of Psychiatry, University of Wisconsin Medical School.

- Clark, R. (1999). The Parent-Child Early Relational Assessment: A factorial validity study. *Educational and Psychological Measurement*, 59(5), 821–846. <https://doi.org/10.1177/00131649921970161>
- Comasco, E., Hallberg, G., Helander, A., Orelund, L., & Sundelin-Wahlsten, V. (2012). Alcohol consumption among pregnant women in a Swedish sample and its effects on the newborn outcomes. *Alcoholism, Clinical and Experimental Research*, 36(10), 1779–1786. <https://doi.org/10.1111/j.1530-0277.2012.01783.x>
- Crawford, J. R., & Henry, J. D. (2003). The Depression Anxiety Stress Scales (DASS): Normative data and latent structure in a large non-clinical sample. *British Journal of Clinical Psychology*, 42(2), 111–131. <https://doi.org/10.1348/014466503321903544>
- Dahlerup, B. R., Egsmose, E. L., Siersma, V., Mortensen, E. L., Hedegaard, M., Knudsen, L. E., Mathiesen, L., & Rosenfeld, C. S. (2018). Maternal stress and placental function, a study using questionnaires and biomarkers at birth. *PLoS One*, 13(11), e0207184. <https://doi.org/10.1371/journal.pone.0207184>
- Danielsdottir, S., & Ingudottir, J. (2020). *The First 1000 Days in the Nordic Countries: A Situation Analysis*. Nordic Council of Ministers. <https://norden.diva-portal.org/smash/get/diva2:1510727/FULLTEXT01.pdf>
- Darwin, Z., McGowan, L., & C, E. L. (2016). Identification of women at risk of depression in pregnancy: Using women’s accounts to understand the poor specificity of the Whooley and Arroll case finding questions in clinical practice. *Archives of Women’s Mental Health*, 19(1), 41–49. <https://doi.org/10.1007/s00737-015-0508-1>
- Drugli, M. B., Solheim, E., Lydersen, S., Moe, V., Smith, L., & Berg-Nielsen, T. . (2018). Elevated cortisol levels in Norwegian toddlers in childcare. *Early Child Development and Care*, 188(12), 1684–1695. <https://doi.org/10.1080/03004430.2016.1278368>
- Eberhard-Gran, M., Eskild, A., Tambs, K., Schei, B., & Opjordsmoen, S. (2001). The Edinburgh Postnatal Depression Scale: Validation in a Norwegian community sample. *Nordic Journal of Psychiatry*, 55(2), 113–117. <https://doi.org/10.1080/08039480151108525>
- Edhborg, M. (2008). Comparisons of different instruments to measure blues and to predict depressive symptoms 2 months postpartum: A study of new mothers and fathers. *Scandinavian Journal of Caring Sciences*, 22(2), 186–195. <https://doi.org/10.1111/j.1471-6712.2007.00512.x>
- Elfström, M. L., Evans, C., Lundgren, J., Johansson, B., Hakeberg, M., & Carlsson, S. G. (2012). Validation of the Swedish version of the Clinical Outcomes in Routine Evaluation Outcome Measure (CORE-OM). *Clinical Psychology & Psychotherapy*, 20(5), 447–455. <https://doi.org/10.1002/cpp.1788>
- Embretson, S. E., & Reise, S. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates.
- Eurenius, E., Sundberg, L. R., Vaezghasemi, M., Silfverdal, S.-A., Ivarsson, A., & Lindkvist, M. (2019). Social-emotional problems among three-year-olds differ based on the child’s gender and custody arrangement. *Acta Paediatrica*, 108(6), 1087–1095. <https://doi.org/10.1111/apa.14668>



- European Federation of Psychologists' Association (EFPA). (2013). EFPA Review Model for the description and evaluation of psychological tests: Tests review form and notes for reviewers. *EFPA Board of Assessment Document 110c*, vol. 4. <http://assessment.efpa.eu/documents/>
- Favez, N., Scaiola, C. L., Tissot, H., Darwiche, J., & Frascarolo, F. (2011). The Family Alliance Assessment Scales: Steps toward validity and reliability of an observational assessment tool for early family interactions. *Journal of Child and Family Studies*, 20(1), 23–37. <https://doi.org/10.1007/s10826-010-9374-7>
- Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., Koss, M. P., & Marks, J. S. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The Adverse Childhood Experiences (ACE) study. *American Journal of Preventive Medicine*, 14(4), 245–258. [https://doi.org/10.1016/S0749-3797\(98\)00017-8](https://doi.org/10.1016/S0749-3797(98)00017-8)
- Fonagy, P., Luyten, P., Moulton-Perkins, A., Lee, Y.-W., Warren, F., Howard, S., Ghinai, R., Fearon, P., Lowyck, B., & Laws, K. (2016). Development and validation of a self-report measure of mentalizing: The reflective functioning questionnaire. *PLoS ONE*, 11(7), ArtID e0158678. <https://doi.org/10.1371/journal.pone.0158678> .
- Fredriksen, E., von Soest, T., Smith, L., & Moe, V. (2019). Depressive symptom contagion in the transition to parenthood: Interparental processes and the role of partner-related attachment. *Journal of Abnormal Psychology*, 128(5), 397–403. <https://doi.org/10.1037/abn0000429> .
- Glascoe, F. P., Gellach, P., & Chen, V. (2019). When do clinicians decide to screen-children for mental health-behavioral-developmental delays/disorders: Is it time to218reconsider policy recommendations? *The Journal of Pediatrics*, 206, 248–255. <https://doi.org/10.1016/j.jpeds.2018.08.084>
- Goldman Fraser, J., Harris-Britt, A., Leone Thakkallapalli, E., Kurtz-Costes, B., & Martin, S. (2010). Emotional availability and psychosocial correlates among mothers in substance-abuse treatment and their young infants. *Infant Mental Health Journal*, 31(1), 1–15. <https://doi.org/10.1002/imhj.20239>
- Goodman, J. H., Watson, G. R., & Stubbs, B. (2016). Anxiety disorders in postpartum women: A systematic review and meta-analysis. *Journal of Affective Disorders*, 203, 292–331. <https://doi.org/10.1016/j.jad.2016.05.033>
- Göransson, M., Magnusson, A., Bergman, H., Rydberg, U., & Heilig, M. (2003). Fetus at risk: Prevalence of alcohol consumption during pregnancy estimated with a simple screening method in Swedish antenatal clinics. *Addiction*, 98(11), 1513–1520. <https://doi.org/10.1046/j.1360-0443.2003.00498.x> .
- Grant, M. J., & Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal*, 26(2), 91–108. <https://doi.org/10.1111/j.1471-1842.2009.00848.x>
- Guedeney, A., & Fermanian, J. (2001). A validity and reliability study of assessment and screening for sustained withdrawal reaction in infancy: The alarm distress baby scale. *Infant Mental Health Journal*, 22(5), 559–575. <https://doi.org/10.1002/imhj.1018>

- Haapsamo, H., Kuusikko-Gauffin, S., Ebeling, H., Larinen, K., Penninkilampi-Kerola, V., Soini, H., & Moilanen, I. (2013). Communication development and characteristics of influencing factors: A follow-up study from 8 to 36 months. *Early Child Development and Care*, 183(2), 321–334. <https://doi.org/10.1080/03004430.2012.674523>
- Hanssen-Bauer, K., & Welander-Vatn, A. (2012). Måleegenskaper ved den norske versjonen av Edinburgh Postnatal Depression Scale, (EPDS) [Psychometric properties of the Edinburgh Postnatal Depression Scale, (EPDS)]. *PsykTestBarn*, 2, 3. <https://doi.org/10.21337/0017>
- Hautamäki, A., Hautamäki, L., Maliniemi-Piispanen, S., & Neuvonen, L. (2008). Kiintymyssuhteen välittyminen kolmessa sukupolvessa - äidinaitien paluu? [Transmission of attachment across three generations: The return of grandnies?]. *Psykologia*, 43(6), 421–442. <https://doi.org/10.1080/17405620902983519>
- Hedenbro, M., & Lidén, A. (2002). CPICS: Child and Parents' Interaction Coding System in dyads and triads. *Acta Paediatrica*, 91(s440), 1–19. <https://doi.org/10.1111/j.1651-2227.2002.tb00459.x>
- Hedenbro, M., & Rydelius, P.-A. (2014). Early interaction between infants and their parents predicts social competence at the age of four. *Acta Paediatrica*, 103(3), 268–274. <https://doi.org/10.1111/apa.12512>
- Hedenbro, M., & Rydelius, P.-A. (2019). Children's abilities to communicate with both parents in infancy were related to their social competence at the age of 15. *Acta Paediatrica*, 108(1), 118–123. <https://doi.org/10.1111/apa.14430>
- Héðinsson, H., Kristjánsdóttir, H., Ólason, D. Þ., & Sigurðsson, J. F. (2013). A validation and replication study of the patient-generated measure PSYCHLOPS on an Icelandic clinical population. *European Journal of Psychological Assessment*, 29(2), 89–95. <https://doi.org/10.1027/1015-5759/a000136>
- Heidi, J. (2015). Bindung und sozio-emotionale Fähigkeiten von jungen Pflegekindern – eine norwegische prospektive Längsschnittstudie. [Young foster children's attachment and socio-emotional functioning – A Norwegian prospective and longitudinal study]. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 64(10), 752–758. <https://doi.org/10.13109/prkk.2015.64.10.752>
- Heller, S., Aoki, Y., & Sheffner, X. (1999). *Revision of the Crowell Parent-Child Relationship Scale*. Tulane University Medical Center.
- Honkalampi, K., Laitila, A., Juntunen, H., Lehmus, K., Piiparinen, A., & Törmänen, I. (2017). The Finnish Clinical Outcome in Routine Evaluation Outcome Measure: Psychometric exploration in clinical and non-clinical samples. *Nordic Journal of Psychiatry*, 71(8), 589–597. <https://doi.org/10.1080/08039488.2017.1365378>
- Howard, L. M., Ryan, E. G., Trevillion, K., Anderson, F., Bick, D., Bye, A., Byford, S., O'Connor, S., Sands, P., Demilew, J., Milgrom, J., & Pickles, A. (2018). Accuracy of the Whooley Questions and the Edinburgh Postnatal Depression Scale in identifying depression and other mental disorders in early pregnancy. *British Journal of Psychiatry*, 212(1), 50–56. <https://doi.org/10.1192/bjp.2017.9>
- Huber, C. J., Stangler, S. R., & Routh, D. K. (1978). The BOEL test as a screening device for otitis media in infants. *Nursing Research*, 27(3), 178–180. <https://doi.org/10.1097/00006199-197805000-00024>

- Huhtala, M., Korja, R., Lehtonen, L., Haataja, L., Lapinleimu, H., & Rautava, P., & PIPARI Study Group. (2012). Parental psychological well-being and behavioral outcome of very low birth weight infants at 3 years. *Pediatrics*, *129*(4), e937–44. <https://doi.org/10.1542/peds.2011-2411>
- Jakobsen, A. N., Skovgaard, A. M., Lichtenberg, A., & Jørgensen, T. (2007). Kan kontakt- og opmærksomhedsforstyrrelser hos smaborn opdages ved sundhedsplejerskernes boel-prove? Copenhagen County child cohort 2000 [Kan kontakt- og opmærksomhedsforstyrrelser hos småbørn opdages ved sundhedsplejerskernes BOEL-prøve? Copenhagen County Child Cohort 2000]. *Ugeskrift for Læger*, *169* (11), 1001–1005. <https://ugeskriftet.dk/videnskab/kan-kontakt-og-opmaerksomhedsforstyrrelser-hos-smaaboern-opdages-ved-sundhedsplejerskernes> .
- Johansson, M., Svensson, I., Stenström, U., & Massoudi, P. (2017). Depressive symptoms and parental stress in mothers and fathers 25 months after birth. *Journal of Child Health Care*, *21*(1), 65–73. <https://doi.org/10.1177/1367493516679015>
- Jonsdottir, S. S., Thome, M., Steingrimsdottir, T., Lydsdottir, L. B., Sigurdsson, J. F., Olafsdottir, H., & Swahnberg, K. (2017). Partner relationship, social support and perinatal distress among pregnant Icelandic women. *Women and Birth*, *30*(1), e46–e55. <https://doi.org/10.1016/j.wombi.2016.08.005>
- Kahr Nilsson, K., Landorph, S., Houmann, T., Olsen, E. M., & Skovgaard, A. M. (2019). Developmental and mental health characteristics of children exposed to psychosocial adversity and stressors at the age of 18-months: Findings from a population-based cohort study. *Infant Behavior & Development*, *57*(57), 101319. <https://doi.org/10.1016/j.infbeh.2019.04.001>
- Kaljunen, L., Pelkonen, M., & Hakulinen-Viitanen, T. (2005). Voimavaralomakkeen kehittäminen ensimmäistä lastaan odottavan perheen voimavaraisuuden tunnistamiseksi [ad English translation of title]. *Hoitotiede*, *18*(3), 132–139.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1–73. <https://doi.org/10.1111/jedm.12000>
- Karukivi, M., Tolvanen, M., Karlsson, H., & Karlsson, L. (2015). Alexithymia and postpartum anxiety and depression symptoms: A follow-up study in a pregnancy cohort. *Journal of Psychosomatic Obstetrics & Gynecology*, *36*(4), 142–147. <https://doi.org/10.3109/0167482X.2015.1089228>
- Kerstis, B., Aarts, C., Tillman, C., Persson, H., Engström, G., Edlund, B., Öhrvik, J., Sylvén, S., & Skalkidou, A. (2016). Association between parental depressive symptoms and impaired bonding with the infant. *Archives of Women's Mental Health*, *19* (1), 87–94. <https://doi.org/10.1007/s00737-015-0522-3>
- Kerstis, B., Engström, G., Edlund, B., & Aarts, C. (2013). Association between mothers' and fathers' depressive symptoms, sense of coherence and perception of their child's temperament in early parenthood in Sweden. *Scandinavian Journal of Public Health*, *41*(3), 233–239. <https://doi.org/10.1177/1403494812472006>
- Kerstis, B., Nohlert, E., Öhrvik, J., & Widarsson, M. (2016). Association between depressive symptoms and parental stress among mothers and fathers in early parenthood: A Swedish cohort study. *Uppsala Journal of Medical Sciences*, *121*(1), 60–64. <https://doi.org/10.3109/03009734.2016.1143540>

- Khangura, S., Konnyu, K., Cushman, R., Grimshaw, J., & Moher, D. (2012). Evidence summaries: the evolution of a rapid review approach [blinded for review]. *Systematic Reviews*, 1(1), 10. <https://doi.org/10.1186/2046-4053-1-10>
- Kleinman, J. M., Robins, D. L., Ventola, P. E., Pandey, J., Boorstein, H. C., Esser, E. L., Wilson, L. B., Rosenthal, M. A., Sutura, S., Verbalis, A. D., Barton, M., Hodgson, S., Green, J., Dumont-Mathieu, T., Volkmar, F., Chawarska, K., Klin, A., & Fein, D. (2008). The Modified Checklist for Autism in Toddlers: A follow-up study investigating the early detection of autism spectrum disorders. *Journal of Autism Developmental Disorders*, 38(5), 827–839. <https://doi.org/10.1007/s10803-007-0450-9>
- Knapstad, M., Lervik, L. V., Sæther, S., Aarø, L. E., & Smith, O. (2020). Effectiveness of Prompt Mental Health Care, the Norwegian version of Improving Access to Psychological Therapies: A randomized controlled trial. *Psychotherapy and Psychosomatics*, 89(2), 90–105. <https://doi.org/10.1159/000504453>
- Korja, R., Ahlqvist-Björkroth, S., Savonlahti, E., Stolt, S., Haataja, L., Lapinleimu, H., Piha, J., & Lehtonen, L. (2010). Relations between maternal attachment representations and the quality of mother–infant interaction in preterm and full-term infants. *Infant Behavior & Development*, 33(3), 330–336. <https://doi.org/10.1016/j.infbeh.2010.03.010>
- Korja, R., Huhtala, M., Maunu, J., Rautava, P., Haataja, L., Lapinleimu, H., & Lehtonen, L. (2014). Preterm infant's early crying associated with child's behavioral problems and parents' stress. *Pediatrics*, 133(2), e339–e345. <https://doi.org/10.1542/peds.2013-1204>
- Korja, R., Piha, J., Otava, R., Lavanchy Scaiola, C., Ahlqvist-Björkroth, S., Junntila, N., Aromaa, M., Räihä, H., & Study Group, S., & Steps Study Group. (2015). Parents' psychological well-being and parental self-efficacy in relation to the family's triadic interaction. *Infant Mental Health Journal*, 36(3), 298–307. <https://doi.org/10.1002/imhj.21512>
- Korja, R., Piha, J., Otava, R., Lavanchy-scaiola, C., Ahlqvist-Björkroth, S., Aromaa, M., & Räihä, H., & STEPS-study. (2016). Mother's marital satisfaction associated with the quality of mother–father–child triadic interaction. *Scandinavian Journal of Psychology*, 57(4), 305–312. <https://doi.org/10.1111/sjop.12294>
- Korja, R., Savonlahti, E., Ahlqvist-Björkroth, S., Stolt, S., Haataja, L., Lapinleimu, H., Piha, J., & Lehtonen, L. (2008). Maternal depression is associated with mother–infant interaction in preterm infants. *Acta Paediatrica*, 97(6), 724–730. <https://doi.org/10.1111/j.1651-2227.2008.00733.x>
- Kornør, H., & Martinussen, M. (2011). Måleegenskaper ved den norske versjonen av Parenting Stress index, 3. versjon (PSI) [Measurement properties of the Norwegian Version of the Parenting Index, third edition (PSI)]. PsykTestBarn. <https://psyktestbarn.r-bup.no/no/artikler/psi-parenting-stress-index>
- Kouvo, A., Virtanen, E., & Silvén, M. (2003). Kiintymys suomalaisissa perheissä: Äidin, isän ja lapsen kiintymyssuhteiden vastaavuus [Attachment in Finnish families]. *Psykologia*, 38(1), 34–40.
- Kristensen, S., Henriksen, T. B., & Bilenberg, N. (2010). The Child Behavior Checklist for ages 1.5–5 (CBCL/1½–5): Assessment and analysis of parent- and caregiver-reported problems in a population-based sample of Danish preschool children. *Nordic Journal of Psychiatry*, 64(3), 203–209. <https://doi.org/10.3109/08039480903456595>

- Kristjánsdóttir, H., Sigurðsson, B. H., Salkovskis, P., Ólason, D., Sigurdsson, E., Evans, C., & Sigurðsson, J. F. (2015). Evaluation of the psychometric properties of the Icelandic version of the Clinical Outcomes in Routine Evaluation-Outcome Measure, its transdiagnostic utility and cross-cultural validation. *Clinical Psychology and Psychotherapy*, 22(1), 64–74.
- Krogh, M. T., Væver, M. S., Harder, S., & Køppe, S. (2012). Cultural differences in infant development during the first year: A study of Danish infants assessed by the Bayley-III and compared to the American norms. *European Journal of Developmental Psychology*, 9(6), 730–736. <https://doi.org/10.1080/17405629.2012.688101>
- Kujanpää, T., Ylisaukko-Oja, T., Jokelainen, J., Hirsikangas, S., Kanste, O., Kyngäs, H., & Timonen, M. (2014). Prevalence of anxiety disorders among Finnish primary care high utilizers and validation of Finnish translation of GAD-7 and GAD-2 screening tools. *Scandinavian Journal of Primary Health Care*, 32(2), 78–83. <https://doi.org/10.3109/02813432.2014.920597>
- Lamb, M. E., Hwang, C.-P., Frodi, A. M., & Frodi, M. (1982). Security of mother– and father–infant attachment and its relation to sociability with strangers in traditional and nontraditional Swedish families. *Infant Behavior & Development*, 5(2), 355–367. [https://doi.org/10.1016/S0163-6383\(82\)80046-5](https://doi.org/10.1016/S0163-6383(82)80046-5)
- Landsem, I. P., Handegård, B. H., Tunby, J., Ulvund, S. E., & Rønning, J. (2014). Early intervention program reduces stress in parents of preterms during childhood, a randomized controlled trial. *Trials*, 15(1), 387. <https://doi.org/10.1186/1745-6215-15-387>
- Larun, L., Fønhus, M. S., Håvelsrud, K., Brurberg, K. G., & Reinart, L. M. (2013). Depresjonsscreening av gravide og barselkvinner [Screening for depression among pregnant and postpartum women]. Report from Kunnskapssenteret no. 1 –2013. Nasjonalt kunnskapssenter for helsetjenesten. Kunnskapssenteret. <https://www.fhi.no/publ/2013/depresjonsscreening-av-gravide-og-barselkvinner/>
- Lehikoinen, A., Ordén, M. R., Heinonen, S., & Voutilainen, R. (2016). Maternal drug or alcohol abuse is associated with decreased head size from mid-pregnancy to childhood. *Acta Paediatrica*, 105(7), 817–822. <https://doi.org/10.1111/apa.13416>
- Littlewood, E., Ali, S., Dyson, L., Keding, A., Ansell, P., Bailey, D., Bates, D., Baxter, C., Beresford-Dent, J., Clarke, A., Gascoyne, S., Gray, C., Hackney, L., Hewitt, C., Hutchinson, D., Jefferson, L., Mann, R., Marshall, D., McMillan, D., & North, A. (2018). Identifying perinatal depression with case-finding instruments: A mixed-methods study (BaBY PaNDA – Born and Bred in Yorkshire PeriNatal Depression Diagnostic Accuracy). *Health Services Delivery Research*, 6(6), 1–210. <https://doi.org/10.3310/hsdr06060>
- Loop, L., Mouton, B., Brassart, E., & Roskam, I. (2017). The observation of child behavior during parent–child interaction: The psychometric properties of the Crowell Procedure. *Journal of Child and Family Studies*, 26(4), 1040–1050. <https://doi.org/10.1007/s10826-016-0625-0>
- Lotzin, A., Lu, X., Kriston, L., Schiborr, J., Musal, T., Romer, G., & Ramsauer, B. (2015). Observational tools for measuring parent–infant interaction: A systematic review. *Clinical Child and Family Psychology Review*, 18(2), 99–132. <https://doi.org/10.1007/s10567-015-0180-z>

- Lund, I. O., Moen Eilertsen, E., Gjerde, L. C., Røysamb, E., Wood, M., Reichborn-Kjennerud, T., & Ystrom, E. (2019). Is the association between maternal alcohol consumption in pregnancy and pre-school child behavioural and emotional problems causal? Multiple approaches for controlling unmeasured confounding. *Addiction*, *114*(6), 1004–1014. <https://doi.org/10.1111/add.14573> .
- Luyten, P., Mayes, L. C., Nijssens, L., & Fonagy, P. (2017). The parental reflective functioning questionnaire: Development and preliminary validation. *PLoS ONE*, *12* (5), ArtID e0176218. <https://doi.org/http://dx.doi.org/10.1371/journal.pone.0176218>
- Lydsdottir, L. B., Howard, L. M., Olafsdottir, H., Einarsson, H., Steingrimsdottir, T., & Sigurdsson, J. F. (2019). Adverse life experiences and common mental health problems in pregnancy: A causal pathway analysis. *Archives of Women's Mental Health*, *22*(1), 75–83. <https://doi.org/10.1007/s00737-018-0881-7>
- Magnusson, Å., Göransson, M., & Heilig, M. (2007). Hazardous alcohol users during pregnancy: Psychiatric health and personality traits. *Drug and Alcohol Dependence*, *89*(2), 275–281. <https://doi.org/10.1016/j.drugalcdep.2007.01.015> .
- Main, M., Kaplan, N., & Cassidy, J. (1985). Security in infancy, childhood and adulthood: A move to the level of representation. In I. Bretherton & E. Waters (Eds.), *Growing points of attachment theory and research. Monographs of the Society for Research in Child Development* (pp. 66–104). University Of Chicago Press.
- Malmberg-Ceder, K., Haanpää, M., Korhonen, P. E., Kautiainen, H., Veromaa, V., & Soinila, S. (2019). The role of psychosocial risk factors in the burden of headache. *Journal of Pain Research*, *12*, 1733–1741. <https://doi.org/10.2147/jpr.s165263> .
- Månsson, J., & Stjernqvist, K. (2014). Children born extremely preterm show significant lower cognitive, language and motor function levels compared with children born at term, as measured by the Bayley-III at 2.5 years. *Acta Paediatrica*, *103*(5), 504–511. <https://doi.org/10.1111/apa.12585>
- Martinussen, M., & Valla, L. (2013). Måleegenskaper ved den norske versjonen av Ages and Stages Questionnaires (ASQ) [Measurement properties of the Norwegian version of Ages and Stages Questionnaire]. *PsykTestBarn*, *1*(6). <https://doi.org/10.21337/0023>
- Massoudi, P., Hwang, C. P., & Wickberg, B. (2013). How well does the Edinburgh Postnatal Depression Scale identify depression and anxiety in fathers? A validation study in a population based Swedish sample. *Journal of Affective Disorders*, *149*(1), 67–74. <https://doi.org/10.1016/j.jad.2013.01.005> .
- Miniscalco, C., Fernell, E., Thompson, L., Sandberg, E., Kadesjö, B., & Gillberg, C. (2018). Development problems were common five years after positive screening for language disorders and, or, autism at 2.5 years of age. *Acta Paediatrica*, *107*(10), 1739–1749. <https://doi.org/10.1111/apa.14358>
- Misund, A. R., Bråten, S., Nerdrum, P., Pripp, A. H., & Diseth, T. H. (2016). A Norwegian prospective study of preterm mother–infant interactions at 6 and 18 months and the impact of maternal mental health problems, pregnancy and birth complications. *BMJ Open*, *6*(5), e009699. <https://doi.org/10.1136/bmjopen-2015-009699> .

- Moe, V., Braarud, H. C., Wentzel-Larsen, T., Slinning, K., Vannebo, U. T., Guedeney, A., Heimann, M., Rostad, A. M., & Smith, L. (2016). Precursors of social emotional functioning among full-term and preterm infants at 12 months: Early infant withdrawal behavior and symptoms of maternal depression. *Infant Behavior & Development*, *44*, 159–168. <https://doi.org/10.1016/j.infbeh.2016.06.012>
- Moe, R. G., & Mothander, P. R. (2009). Kartlegging av vansker hos sped- og småbarn [Infant mental health assessment]. *Tidsskrift for Norsk psykologforening*, *46*(8), 749–756. <https://doi.org/10.52734/p43yn32q>.
- Mörelius, E., Örténstrand, A., Theodorsson, E., & Frostell, A. (2015). A randomised trial of continuous skin-to-skin contact after preterm birth and the effects on salivary cortisol, parental stress, depression, and breastfeeding. *Early Human Development*, *91*(1), 63–70. <https://doi.org/10.1016/j.earlhumdev.2014.12.005>
- Mortensen, J. T., Olsen, J., Larsen, H., Bendsen, J., Obel, C., & Sørensen, H. T. (2003). Psychomotor development in children exposed in utero to benzodiazepines, antidepressants, neuroleptics, and anti-epileptics. *European Journal of Epidemiology*, *18* (8), 769–771. <https://doi.org/10.1023/a:1025306304635>.
- Mothander, P. R., & Moe, R. G. (2008). Infant Mental Health Assessment: The use of DC 0–3 in an outpatient child psychiatric clinic in Scandinavia. *Scandinavian Journal of Psychology*, *49*(3), 259–267. doi:<https://doi.org/10.1111/j.1467-9450.2008.00632.x>
- Mothander, P. R., & Moe, R. G. (2010). Self-reported depressive symptoms and parental stress in mothers and fathers who bring their infants to an infant mental health clinic. *Nordic Journal of Psychiatry*, *64*(5), 310–316. <https://doi.org/10.3109/08039480903528138>
- Murphy, A., Steele, M., Dube, S. R., Bate, J., Bonuck, K., Meissner, P., Goldman, H., & Steele, H. (2014). Adverse Childhood Experiences (ACEs) questionnaire and Adult Attachment Interview (AAI): Implications for parent child relationships. *Child Abuse and Neglect*, *38*(2), 224–233. <https://doi.org/10.1016/j.chiabu.2013.09.004>
- Mustonen, K., Hermanson, E., Koivu, M., Fellman, & von Wendt, V. L. (2006). Kuinka Vauvan neurologinen arviointi -menetelmä Vane pystyy ennustamaan pienten keskosten neurokognitiivista kehitystä? *Suomen Lääkärilehti*, *9*, 955–961.
- Nygren, G., Sandberg, E., Gillstedt, F., Ekeröth, G., Arvidsson, T., & Gillberg, C. (2012). A new screening programme for autism in a general population of Swedish toddlers. *Research in Developmental Disabilities*, *33*(4), 1200–1210. <https://doi.org/10.1016/j.ridd.2012.02.018>
- Östberg, M., Hagekull, B., & Wettergren, S. (1997). A measure of parental stress in mothers with small children: Dimensionality, stability and validity. *Scandinavian Journal of Psychology*, *38*(3), 199–208. <https://doi.org/10.1111/1467-9450.00028>
- Østergaard, K. K., Lando, A. V., Hansen, B. M., & Greisen, G. (2012). A Danish reference chart for assessment of psychomotor development based on the Ages & Stages Questionnaire. *Danish Medical Journal*, *59*(6), A4429.
- Paap, M. C. S., Martinussen, M., Jakobsen, R., Jozefiak, T., Axelsdottir, B., & Neumer, S. P. (2018, June 24th). Tester for barn og unge i Norge bør kvalitetssikres bedre [Tests for children and young people should be subjected to quality control]. *Psykologisk.no*. <https://psykologisk.no/2018/06/tester-for-barn-og-unge-i-norge-bor-kvalitetssikres-bedre/>



- Pajulo, M., Pyykkönen, N., Kalland, M., Sinkkonen, J., Helenius, H., & Punamäki, R.-L. (2011). Substance abusing mothers in residential treatment with their babies: Postnatal psychiatric symptomatology and its association with mother–child relationship and later need for child protection actions. *Nordic Journal of Psychiatry, 65*(1), 65–73. <https://doi.org/10.3109/08039488.2010.494310>
- Parfitt, Y., Pike, A., & Ayers, S. (2013). The impact of parents' mental health on parent–baby interaction: A prospective study. *Infant Behavior & Development, 36*(4), 599–608. <https://doi.org/10.1016/j.infbeh.2013.06.003>
- Plummer, F., Manea, L., Trepel, D., & McMillan, D. (2016). Screening for anxiety disorders with the GAD-7 and GAD-2: A systematic review and diagnostic meta-analysis. *General Hospital Psychiatry, 39*, 24–31.
- Pontoppiddan, M., Niss, N. K., Pejtersen, J. H., Julian, M. M., & Væver, M. S. (2017). Parent report measures of infant and toddler social–emotional development: A systematic review. *Family Practice, 34*(2), 127–137. <https://doi.org/10.1093/fampra/cmz003>
- Præstegaard, C., Kesmodel, P. S., & Kesmodel, U. S. (2018). Is TWEAK a valid screening questionnaire to identify alcohol risk drinkers among pregnant women in Denmark? *Acta Obstetrica et Gynecologica Scandinavica, 97*(4), 483–490. <https://doi.org/10.1111/aogs.13314>
- Puura, K., Guedeney, A., Mäntymaa, M., & Tamminen, T. (2007). Detecting infants in need: Are complicated measures really necessary? *Infant Mental Health Journal, 28*(4), 409–421. <https://doi.org/10.1002/imhj.20144>
- Puura, K., Leppänen, J., Salmelin, R., Mäntymaa, M., Luoma, I., Latva, R., Peltola, M., Lehtimäki, T., & Tamminen, T. (2019). Maternal and infant characteristics connected to shared pleasure in dyadic interaction. *Infant Mental Health Journal, 40*(4), 459–478. <https://doi.org/10.1002/imhj.21786>
- Puura, K., Mäntymaa, M., Leppänen, J., Peltola, M., Salmelin, R., Luoma, I., Latva, R., & Tamminen, T. (2013). Associations between maternal interaction behavior, maternal perception of infant temperament, and infant social withdrawal. *Infant Mental Health Journal, 34*(6), 586–593. <https://doi.org/10.1002/imhj.21417>
- Puura, K., Mäntymaa, M., Luoma, I., Kaukonen, P., Guedeney, A., Salmelin, R., & Tamminen, T. (2010). Infants' social withdrawal symptoms assessed with a direct infant observation method in primary health care. *Infant Behavior & Development, 33*(4), 579–588. <https://doi.org/10.1016/j.infbeh.2010.07.009>
- Rasmussen, L.-M. P., & Martinussen, M. (2013). Måleegenskaper ved den norske versjonen av Ages and Stages Questionnaire: Social and Emotional, ASQ:SE [Psychometric properties of the ASQ:SE]. *PsykTestBarn, 2:1*. <https://doi.org/10.21337/0024>
- Rasting, P., & Lindbaek, M. (1998). Tympanometri på helsestasjonen [Tympanometry at a health center]. *Tidsskrift for den Norske Laegeforening, 118*(22), 3426–3430.
- Ravn, S. H., & Bjerager, M. O. (2004). Boel-Testen: En falsk trygghed [The BOEL test: False confidence]. *Ugeskrift for Laeger, 166*(36), 3086–3088.
- Richter, J., & Valla, L. (2013). Måleegenskaper ved den norske versjonen av Bayley Scales of Infant and Toddler Development (BSID, BSID-II, BSID-III) [The psychometric properties of the Norwegian version of Bayley Scales of Infant and Toddler Development (BSID, BSID-II, BSID-III)]. *PsykTestBarn, 2, 3*. <https://doi.org/10.21337/0026>



- Robins, D. L., Casagrande, K., Barton, M., Chen, C. M., Dumont-Mathieu, T., & Fein, D. (2014). Validation of the Modified Checklist for Autism in Toddlers, Revised with follow-up (M-CHAT-R/F). *Pediatrics*, *133*(1), 37–45. <https://doi.org/10.1542/peds.2013-1813>
- Robins, D. L., Fein, D., & Barton, M. (2018). *Modified Checklist for Autism in Toddlers, Revised, with follow-up (M-CHAT-R/F) TM*. [https://mchatscreen.com/wp-content/uploads/2015/09/M-CHAT-R\\_F\\_Rev\\_Aug2018.pdf](https://mchatscreen.com/wp-content/uploads/2015/09/M-CHAT-R_F_Rev_Aug2018.pdf)
- Robins, D. L., Fein, D., Barton, M. L., & Green, J. A. (2001). The Modified Checklist for Autism in Toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *31*(2), 131–144. <https://doi.org/10.1023/A:1010738829569>
- Robinson, A., Myhill, A., Roberts, J., & Tilley, N. (2016). Risk-led policing of domestic abuse and the DASH risk model. *What Works: Crime Reduction Research*. Cardiff & London: Cardiff University, College of Policing and UCL Department of Security and Crime Science.
- Rohder, K., Nystrom-Hansen, M., MacBeth, A., Davidsen, K. A., Gumley, A., Brennan, J., George, C., & Harder, S. (2019). Antenatal caregiving representations among expectant mothers with severe mental illness: A cross-sectional study. *Journal of Reproductive and Infant Psychology*, *37*(4), 370–383. <https://doi.org/10.1080/02646838.2019.1578868>
- Rozental, A., Shafran, R., Wade, T. D., Kothari, R., Egan, S. J., Ekberg, L., Wiss, M., Carlbring, P., & Andersson, G. (2018). Guided web-based cognitive behavior therapy for perfectionism: Results from two different randomized controlled trials. *Journal of Medical Internet Research*, *20*(4), e154. <https://doi.org/10.2196/jmir.9823>
- Rubertsson, C., Börjesson, K., Berglund, A., Josefsson, A., & Sydsjö, G. (2011). The Swedish validation of Edinburgh Postnatal Depression Scale (EPDS) during pregnancy. *Nordic Journal of Psychiatry*, *65*(6), 414–418. <https://doi.org/10.3109/08039488.2011.590606>
- Sajaniemi, N., Hakamies-Blomqvist, L., Katainen, S., & von Wendt, L. (2001). Early cognitive and behavioral predictors of later performance: A follow-up study of ELBW children from ages 2 to 4. *Early Childhood Research Quarterly*, *16*(3), 343–361. [https://doi.org/10.1016/S0885-2006\(01\)00107-7](https://doi.org/10.1016/S0885-2006(01)00107-7)
- Salari, R., Wells, M. B., & Sarkadi, A. (2014). Child behaviour problems, parenting behaviours and parental adjustment in mothers and fathers in Sweden. *Scandinavian Journal of Public Health*, *42*(7), 547–553. <https://doi.org/10.1177/1403494814541595>
- Salo, S., & Mäkelä, J. (in press). MIM Quantitative Rating Systems for Parent–Child. In S. Salo & P. Booth (Eds.), *Handbook of the MIM (Chapter 5)*. The Theraplay Institute.
- Sarfi, M., Sundet, J. M., & Waal, H. (2013). Maternal stress and behavioral adaptation in methadone- or buprenorphine-exposed toddlers. *Infant Behavior & Development*, *36*(4), 707–716. <https://doi.org/10.1016/j.infbeh.2013.08.006>
- Savonlahti, E., Pajulo, M., Ahlqvist, S., Helenius, H., Korvenranta, H., Tamminen, T., & Piha, J. (2005). Interactive skills of infants with their high-risk mothers. *Nordic Journal of Psychiatry*, *59*(2), 139–147. <https://doi.org/10.1080/08039480510022990>
- Schmidt, F. (2010). Detecting and Correcting the Lies That Data Tell. *Perspectives on Psychological Science*, *5*(3), 233–242. <http://www.jstor.org/stable/41613333>

- Sebire, J., & Barling, H. (2016). Assessing the assessors: An analysis of the consistency of risk grading by police when conducting domestic abuse investigations. *Policing, 10* (4), 351–360. <https://doi.org/10.1093/police/paw012>
- Seimyr, L., Sjögren, B., Welles-Nyström, B., & Nissen, E. (2009). Antenatal maternal depressive mood and parental-fetal attachment at the end of pregnancy. *Archives of Women's Mental Health, 12*(5), 269–279.
- Selin, K. H. (2003). Test-retest reliability of the alcohol use disorder identification test in a general population sample. *Alcoholism, Clinical and Experimental Research, 27*(9), 1428–1435.
- Skovgaard, A. M., Olsen, E. M., Christiansen, E., Houmann, T., Landorph, S. L., & Jørgensen, T., the CCC 2000 Study Group. (2008). Predictors (0–10 months) of psychopathology at age 1½ years – a general population study in The Copenhagen Child Cohort CCC 2000. *Journal of Child Psychology and Psychiatry, 49*(5), 553–562.
- Skre, I., Friborg, O., Elgarøy, S., Evans, C., Myklebust, L. H., Lillevoll, K., & Hansen, V. (2013). The factor structure and psychometric properties of the Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM) in Norwegian clinical and non-clinical samples. *BMC Psychiatry, 13*, 99. <https://doi.org/10.1186/1471-244x-13-99>
- Slot, P. L., Boom, J., Verhagen, J., & Leseman, P. P. M. (2017). Measurement properties of the CLASS Toddler in ECEC in the Netherlands. *Journal of Applied Developmental Psychology, 48*, 79–91. <https://doi.org/10.1016/j.appdev.2016.11.008>
- Smith-Nielsen, J., Lange, T., Wendelboe, K. I., von Wowern, R. K., & Væver, M. S. (2019). Associations between maternal postpartum depression, infant social behavior with a stranger, and infant cognitive development. *Infancy, 24*(4), 663–670. <https://doi.org/10.1111/infa.12287>
- Smith-Nielsen, J., Matthey, S., Lange, T., & Væver, M. S. (2018). Validation of the Edinburgh Postnatal Depression Scale against both DSM-5 and ICD-10 diagnostic criteria for depression. *BMC Psychiatry, 18*(1), 393. <https://doi.org/10.1186/s12888-018-1965-7>
- Smith-Nielsen, J., Tharner, A., Steele, H., Cordes, K., Mehlhase, H., & Væver, M. S. (2016). Postpartum depression and infant–mother attachment security at one year: The impact of co-morbid maternal personality disorders. *Infant Behavior & Development, 44*, 148–158.
- Sprang, G., & Craig, C. (2014). Crowell problem solving procedure: A psychometric analysis of a laboratory measure of the caregiver child relationship. *Child and Adolescent Mental Health, 20*(4), 202–209.
- Squires, J., Bricker, D., Twombly, E., Murphy, K., & Hoselton, R. (2015). *ASQ:SE-2 Technical Report*. Bookes Publishing.
- Steele, H., Bate, J., Steele, M., Dube, S. R., Danskin, K., Knafo, H., Nikitiades, A., Bonuck, K., Meissner, P., & Murphy, A. (2016). Adverse childhood experiences, poverty, and parenting stress. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences du Comportement, 48*(1), 32–38.
- Stene-Larsen, K., Torgersen, L., Strandberg-Larsen, K., Normann, P. T., & Vollrath, M. E. (2013). Impact of maternal negative affectivity on light alcohol use and binge drinking during pregnancy. *Acta Obstetrica et Gynecologica Scandinavica, 92*(12), 1388–1394.

- Suija, K., Rajala, U., Jokelainen, J., Liukkonen, T., Härkönen, P., Keinänen-Kiukaanniemi, S., & Timonen, M. (2012). Validation of the Whooley Questions and the Beck Depression Inventory in older adults. *Scandinavian Journal of Primary Health Care*, 30(4), 259–264. <https://doi.org/10.3109/02813432.2012.732473>
- Sun, J., Patel, F., Rose-Jacobs, R., Frank, D. A., Black, M. M., & Chilton, M. (2017). Mothers' adverse childhood experiences and their young children's development. *American Journal of Preventive Medicine*, 53(6), 882–891. <https://doi.org/10.1016/j.amepre.2017.07.015>
- Theran, S. A., Levendosky, A. A., Bogat, G. A., & Huth-Bocks, A. C. (2005). Stability and change in mothers' internal representations of their infants over time. *Attachment & Human Development*, 7(3), 253–268. <https://doi.org/10.1080/14616730500245609>
- Thomason, A. C., & La Paro, K. M. (2009). Measuring the quality of teacher–child interactions in toddler child care. *Early Education and Development*, 20(2), 285–304. <https://doi.org/10.1080/10409280902773351>
- Thornton, S. (2017). Police attempts to predict domestic murder and serious assaults: Is early warning possible yet? *Cambridge Journal of Evidence-Based Policing*, 1(2–3), 64–80. <https://doi.org/10.1007/s41887-017-0011-1>
- Tiirikainen, K., Haravuori, H., Ranta, K., Kaltiala-Heino, R., & Marttunen, M. (2019). Psychometric properties of the 7-item Generalized Anxiety Disorder Scale (GAD-7) in a large representative sample of Finnish adolescents. *Psychiatry Research*, 272, 30–35. <https://doi.org/10.1016/j.psychres.2018.12.004>
- Turner, E., Medina, J., & Brown, G. (2019). Dashing hopes? The predictive accuracy of domestic abuse risk assessment by police. *The British Journal of Criminology*, 59(5), 1013–1034. <https://doi.org/10.1093/bjc/azy074>
- Vaezghasemi, M., Eurenus, E., Ivarsson, A., Sundberg, L. R., Silferdal, S.-A., & Lindkvist, M. (2020). Social-emotional problems among Swedish three-year-olds: An Item Response Theory analysis of the Ages and Stages Questionnaires: Social-Emotional. *BMC Pediatrics*, 20(1), 149
- Valla, L., Birkeland, M. S., Hofoss, D., & Slinning, K. (2017). Developmental pathways in infants from 4 to 24 months. *Child: Care, Health and Development*, 43(4), 546–555. <https://doi.org/10.1111/cch.12467>
- Van Dam, M., & Van Ijzendoorn, M. H. (1988). Measuring attachment security: Concurrent and predictive validity of the parental attachment q-set. *The Journal of Genetic Psychology: Research and Theory on Human Development*, 149(4), 447–457.
- Vedel, C., Larsen, H., Holmskov, A., Andreasen, K. R., Uldbjerg, N., Ramb, J., Bødker, B., Skibsted, L., Sperling, L., Krebs, L., Zingenberg, H., Laursen, L., Christensen, J. T., Tabor, A., & Rode, L. (2020). Neonatal complications and neurophysiological development in twins – a long-term follow-up study. *The Journal of Maternal-fetal & Neonatal Medicine*, 35(2), 372–378. Advance online publication. <https://doi.org/10.1080/14767058.2020.1718647>
- Velikonja, T., Edbrooke-Childs, J., Calderon, A., Slead, M., Brown, A., & Deighton, J. (2016). The psychometric properties of the Ages & Stages Questionnaires for ages 2–2.5: A systematic review. *Child: Care, Health and Development*, 43(1), 1–17. <https://doi.org/10.1111/cch.12397>

- Vreeswijk, C. M. J. M., Maas, A. J. B. M., & van Bakel, H. J. A. (2012). Parental representations: A systematic review of the working model of the child interview. *Infant Mental Health Journal, 33*(3), 314–328. <https://doi.org/10.1002/imhj.20337>
- Wang, M. V., Lekhal, R., Aarø, L. E., & Schjølberg, S. (2012). Co-occurring development of early childhood communication and motor skills: Results from a population-based longitudinal study. *Child: Care, Health and Development, 40*(1), 77–84. <https://doi.org/10.1111/cch.12003>
- Weile, L. K. K., Wu, C., Hegaard, H. K., Kesmodel, U. S., Henriksen, T. B., Ibsen, I. O., & Nohr, E. A. (2020). Identification of alcohol risk drinking behaviour in pregnancy using a web-based questionnaire: Large-scale implementation in antenatal care. *Alcohol and Alcoholism, 55*(2), 225–232. <https://doi.org/10.1093/alcalc/agz100>
- WHO. (2021). Mental Health and Substance Use (who.int) Retrieved 28.5.2021. <https://www.who.int/westernpacific/about/how-we-work/programmes/mental-health-and-substance-abuse>
- Zeanah, C. H., Benoit, D., Hirshberg, L., Barton, M. L., & Regan, C. (1994). Mothers' representations of their infants are concordant with infant attachment classifications. *Developmental Issues in Psychiatry and Psychology, 1*, 9–18.