Exploring the use of Infant CARE-Index to assess parent-infant interactions (0-15 months): Measurement properties and acceptability in practice

Samiha Mirza

MSc by Research

University of York

Health Sciences

July 2020

Abstract

Background: The quality of parent-infant interactions is known to have a significant impact on the social and emotional development of infants. With the increasing mental health problems in children in the later years of life, there is a growing need of introducing an observational measure to assess parent-infant interactions during the first year of life as emphasised by NICE. Infant CARE-Index is a multi-dimensional observational play-based measure for assessing the quality of parent–infant interaction from birth up to 15 months. It can assess the risk to relationships, screen for developmental risk, as well as guide interventions and assess the outcomes of interventions. With these factors in mind, this research sought to answer the question: *How can the parent-infant interactions (birth-15 months) be assessed with the help of Infant CARE-Index in association with its measurement properties and acceptability in practice?*

Methods: A comprehensive search of the literature to date utilising Infant CARE-Index for the age range birth up to 15 months was performed encompassing key databases such as APA PsycINFO, MIDIRS, Embase, Ovid MEDLINE, CINAHL, Proquest etc. This was followed by a qualitative analysis (comprising of focus groups) on the acceptability of Infant CARE-Index from the perspective of health professionals within UK settings.

Results: Forty-one articles meeting the inclusion criteria were subjected to data extraction. COSMIN checklist was used for assessing reliability and validity evidence. Evidence on inter-rater reliability, hypothesis testing for construct validity and responsiveness was extracted and presented. While, the qualitative analysis revealed that there was a definite need of a measure for the age range (0-12 months), however, the training was rigorous and expensive, and a holistic approach was needed to improve the assessment procedures for assessing parent-infant interactions.

Conclusions: Exploring the use of Infant CARE-Index within large scale randomised controlled trials of parenting support programmes will reveal its applicability within community settings. Validation studies assessing its measurement properties will add the much-needed reliability and validity information still deficient in the literature. Adequate resources and support for training the health professionals by the commissioners is imperative in rolling out standardised assessments. Furthermore, perspectives from a wider range of users of Infant CARE-Index needs to be gathered to provide a broader overview on the acceptability of this measure.

Table of Contents

Abstract *2*

List of tables *6*

List of figures *6*

Acknowledgements *7*

Declaration *8*

Chapter 1: Introduction & background *9*

1.1 Significance of healthy parent-infant interactions **9**

1.2 Theoretical underpinnings of assessing parent-infant interactions **11**

1.3 Infant CARE-Index **12**

1.4 Infant CARE-Index and how it differs from other assessment measures **14**

1.5 Rationale for the current research **15**

1.6 Aims & Objectives **16**

Chapter 2: A review on the measurement properties of Infant CARE-Index ... *18*

2.1 Introduction **18**

2.2 Research questions **18**

2.3 Procedure **19**

2.3.1 Data sources and searches 19

2.3.2 Study selection 21

2.3.3 Data extraction and quality assessment 22

2.3.4 Data synthesis and analysis 25

2.4 Results **26**

2.4.1 Study characteristics 27

2.4.2 Measurement properties assessed and data synthesis 27

2.4.3 Methodological quality of the studies 29

2.4.4 Quality of the assessed measurement properties 30

2.4.5 Summarising the evidence 31

2.4.6 Grading the quality of evidence 32

2.4.7 Formulation of recommendations according to COSMIN 39

2.5 Discussion **40**

2.5.1 Strengths....41
2.5.2 Limitations....41
2.5.3 Recommendations....42
2.5.4 Conclusions....43

Chapter 3: Qualitative analysis on the acceptability of Infant CARE-Index in the UK....45

3.1 Introduction....45
3.1.1 Aims....45
3.1.2 Research questions....46

3.2 Methods....46
3.2.1 Eligibility criteria....46
3.2.2 Participant selection, sample size and data collection....46

3.3 Procedure....47
3.3.1 Focus group with Infant CARE-Index trainees....47
3.3.2 Focus group with health professionals....48

3.3 Analysis....48
3.3.1 Phases or steps of thematic analysis....49

3.4 Findings....50
3.5.1 Need of an assessment measure....51
3.5.2 Current practices....52
3.5.3 Need of structural changes....53
3.5.4 Relationship with the client family....53
3.5.5 Anticipated apprehension and ease of giving instructions for Infant CARE-Index video recordings....54
3.5.6 Coding methodology....55
3.5.7 Time constraints....56
3.5.8 Provision of training....57
3.5.9 Wider implications of assessments on parent-infant interactions....58

3.6 Discussion....59
3.6.1 Trustworthiness and rigor of the study....60
3.6.2 Validating the data and results....60
3.6.3 Strengths and limitations....60
3.6.4 Conclusions....61
3.6.5 Ethical concerns....61

Chapter 4: Overall Discussion *63*

4.1 Summary of results **63**

4.1.1 Applying the findings to a broader context 64

4.1.2 Parent-infant interactions as a measurement focus 65

4.2 Quality of evidence **66**

4.3 Strengths & limitations **66**

4.4 Recommendations for future research **67**

Appendix 1 ..... *69*

Definitions used for measurement properties **69**

Appendix 2 ..... *70*

Example search strategy **70**

Appendix 3 ..... *71*

Study summaries (Characteristics of included study populations) **71**

Appendix 4 ..... *80*

Results of studies on Reliability evidence **80**

Appendix 5 ..... *88*

Results of studies on Hypothesis testing **88**

Appendix 6 ..... *96*

Results of studies on Responsiveness **96**

Appendix 7 ..... *101*

Participant Information Sheet for focus groups with Infant CARE-Index trainees **101**

Appendix 8 ..... *103*

Participant Consent Form for focus groups with Infant CARE-Index trainees **103**

Appendix 9 ..... *104*

Focus group guide for Infant CARE-Index trainees **104**

Appendix 10 ..... *105*

Participant Information Sheet for focus groups with Health professionals **105**

Appendix 11 ..... ***107***

Participant Consent Form for focus groups with Health professionals **107**

Appendix 12 ..... ***108***

Focus group guide for Health professionals **108**

Appendix 13 ..... ***109***

Consolidated criteria for reporting qualitative research (COREQ): A checklist for focus groups **109**

References ***112***

References of Included studies in the systematic review ***121***

List of tables

Table 1: Boxes of COSMIN risk of Bias checklist and order in which the measurement properties are assessed 26
Table 2: Generic hypotheses to evaluate construct validity and responsiveness 28
Table 3: Updated criteria for good measurement properties 30
Table 4: Definitions of quality levels 33
Table 5: Modified GRADE approach for grading the quality of evidence 33
Table 6: Guidelines and definitions on downgrading for Risk of Bias 34
Table 7: Summary of Findings table (Reliability) 38
Table8: Summary of findings table (Hypotheses testing) 39
Table 9: Summary of findings table (Responsiveness) 39

List of figures

Figure 1: Steps for conducting a systematic review on the measurement properties of an observational measure 23
Figure 2: Flow chart of article retrievals on Infant CARE-Index 24

Total word count: 19698 (Excluding Title, Abstract, Table of contents, list of tables & figures, appendices and references)

Acknowledgements

Thanks to my supervisors for their generous expertise, guidance and support over the course of this project.

Declaration

I declare that this thesis is a presentation of original work and I, am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Chapter 1: Introduction & background

1.1 Significance of healthy parent-infant interactions

The importance of the first few years of life in terms of the quality of the parent-infant interactions has been established to play a pivotal role for a range of outcomes in the child's later socioemotional development (Barlow, 2008, Stewart-Brown and Strader-McMillan, 2010). According to the World Health Organisation (WHO, 2018), worldwide 10-20% of children and adolescents are suffering from mental health disorders. In the UK alone, 10% of children and young people (aged 5-16 years) have a clinically diagnosable mental health difficulty (Green, 2005). It is reported that in 2017, the spending on mental health provision increased by £575m (NHS, 2017).

Sensitive parenting plays a major role in the early years of a child. In the context of parent-infant interactions, *sensitivity* of an adult is defined as 'any pattern of behaviour that pleases the infant and increases the infant's comfort and attentiveness and reduces its distress and disengagement' (Crittenden, 2010). Research has also established that *insensitive parenting* is found to be one of the major causes of abuse of children aged up to one year (Barnes, 2013). National statistics for England in 2018 revealed that out of the 53790 children (under 18 years) who were the subject of a child protection plan on 31st March 2018; approximately 5000 (9.2%) of these children were less than one year of age (DfE, 2018). Further breakdown of this data revealed that cases of **neglect** were the highest 53% (2710 out of a total 5280) for under one-year age group, followed by **physical abuse** 0.12% (510 out of 4120 cases), **emotional abuse** 0.07% (1410 out of a total 18860), and **sexual abuse** 0.06% (150 out of 2180 cases). This prevalence of '*children in need'* is increasing in all age groups (under 18 years) since the past few years. The national statistics for England documented the growing trend in terms of percentage of *children in need* on 31st March 2018 by initial category of abuse which revealed that emotional abuse (31.6% in 2012- 2013 to 35.1% in 2017-2018) and neglect (41.6% in 2012-2013 to 48% in 2017-2018) have proven to be the most detrimental factors (DfE, 2018).

Sensitive and responsible parenting is of utmost importance during the first critical year in a child's life, as the brain is undergoing rapid development and is sensitive to any negative effects. Inappropriate parent-infant interactions and traumatic experiences in the early period of a child's life leads to increased cortisol levels, which

later increases the risk of hyperactivity, anxiety and attachment difficulties (Gunnar & Donzella, 2002).

The growing mental health problems in school going children of the UK (NHS, 2017) has led to the development of a wide range of interventions, some of which are implemented in the form of parenting support programmes. Many of these programmes are aimed to be provided early in the development of an issue/problem (i.e. during pregnancy and the first three years of life). Although early intervention may not be able to address all the issues, however, it can substantially improve children's lives, if it is delivered to a high standard to the children or families who need it the most. Some of the programmes with good evidence of improving children's attachment security during the early years include *Infant Parent Psychotherapy*, *Child Parent Psychotherapy* and *Family Nurse Partnership* (EIF report, 2018). A review of reviews on effectiveness of parenting programs by Barlow & Coren, 2018, consistently pointed towards the effectiveness of parenting programmes in terms of improving child conduct problems, parental mental health, and parenting skills for parents of children aged 3–12 years in the short term. However, the evidence for effectiveness with parents of younger children, particularly less than 3 years of age is although present, but less effective (Barlow et al., 2012).

National Institute of Clinical Excellence (NICE) has provided guidelines for health and social care workers regarding children and young people with attachment difficulties (or at risk of attachment difficulties). According to the NICE guidelines, health and social care professionals are recommended to offer video feedback programme to the parents of preschool-age children on the edge of care to help them in improving their understanding of what their child's behaviour means and how to respond to it appropriately (NICE, 2015).

There are various measures developed to assess parent- infant interactions. Some of which are in the form of observational measures, self-reported questionnaires and interviews. However, the rapid progress which a child typically encounters across all developmental domains (for example intellectual, emotional, physical and locomotory) makes one measure unfit to measure parent-child interactions across all age ranges (Pontoppidan, 2017), hence different measures have been produced keeping in mind the developmental stage of the child.

With growing literature on the importance of accurate assessment procedures in the early years, many measures are being tested to establish their validity and reliability, while many new measures are also being developed to replace earlier measures

having inconsistent use. As, reliability and validity evidence of a measure constitutes one of the most important aspects of its overall quality; measures with insufficient psychometric properties are predisposed to become a major source of bias (Abedi A et al, 2019).

This has led to publication of several reviews aimed at identifying the most appropriate measures to assess parent-child interactions which are currently being used in randomised controlled trials on parenting programmes in the UK (Gridley et al. 2019). Such reviews are significant, as they provide evidence base on the measures which are being currently used, this may in turn aid in improving socioemotional wellbeing of a child, parent, family and the wider community on the long term by timely assessments and interventions (Barlow et al. 2008).

1.2 Theoretical underpinnings of assessing parent-infant interactions

The underlying framework of assessing parent-infant interactions rests upon the Attachment theory; which was the result of pioneering research by John Bowlby in the 1950s (Bowlby, 1953). Attachment is defined as 'a deep and enduring emotional tie that connects one person to another across time and space' (McLeod, 2017). In the field of parent-infant interactions, this phenomenon is typically exhibited between the mother and her infant especially when the infant feels vulnerable in times of stress (McLeod, 2017).

The Attachment theory was further advanced as an experimental and methodological assessment procedure by Mary Ainsworth in the 1960s into what is now known as the 'Strange Situation Procedure' (SSP) (Ainsworth, 1978). SSP is a laboratory based observational assessment of maternal-infant attachment (from 11 months to 24 months) during contact, separation, reunion and in the presence of a stranger. Since SSP is an observational measure and it adopts a systematic approach to detect and interpret behaviour, thus, to date the SSP is considered the gold standard in assessing attachment pattern in infancy and early childhood (Wright, 2015). The term infancy refers specifically to a child under one year of age, but it is often widely used to include children from birth up to two years of age.

With very young children; self-reported measures and interviews are not of a beneficial value, hence, it is here, where the observational measures come into play,

which have the potential of capturing the unfolding of behaviour over time. This, then helps in understanding the functionality in relation to the history and on-going synchrony between the parent and infant (Girard et al, 2016). The significance of carrying out observational assessments rather than parent/observer reported measures in assessing attachment is evident from a recent meta-analysis (Groh et al, 2017) on the role of early attachment on the socioemotional development. In this meta-analysis, the results showed, that the studies utilising parent reports (d=0.22) and teacher reports (d=0.30) produced smaller effect sizes than those studies that used direct observations (d=0.58) (Groh et al.2017).The explanation might be that in most instances, parents and teachers might find it difficult to report on internal symptoms due to the nature of such symptoms.

However, according to the researcher's knowledge, to date, there is no standardised observational measure present in the UK to assess parent-infant interactions from birth to one year. Furthermore, this is evident in the NICE guidelines on children's attachment; indicating non-mention of any measure to assess parent-infant interactions before 12 months of age (NICE guidance, 2015). The need for a standardised measure which can be implemented across a population in this particular age range is crucial, especially in the light of growing socioemotional difficulties in the later stages of children's lives (Groh et al. 2017). Ideally this measure should be age appropriate, easy to administer, capable of identifying *at risk* dyads and comprise of appropriate psychometric properties. Keeping these evidence gaps in mind, the researcher intends to explore an observational measure in this thesis.

1.3 Infant CARE-Index

In the field of assessing quality of parent-infant interactions from birth up to 15 months, numerous measures have been designed to study the various aspects of parent-infant interactions (Svanberg et al., 2013, Biringen, 2014). These measures have multiple uses in the form of screening, planning interventions and monitoring outcomes as a result of interventions. When used as a screener, they help in identifying parent-infant dyads at risk of developing difficulties in the child's later cognitive, social and emotional well-being. However, majority of these measures are generally based on measurement scales which assess *either* parenting behaviour e.g. Key to Interactive parenting scale (KIPS) (Comfort, 2011) and Parenting Interactions with Children Checklist of Observations Linked to Outcomes (PICCOLO) (Roggman,

2009) *or* infant behaviour e.g. Home Observation Measurement of the Environment (HOME) (Bradley, 1977), but not both (dyadic interaction). Measures which do assess dyadic interactions are designed for older children, e.g. Dyadic Parent-Child Interaction Coding System (DPICS) assesses parent-child dyads in the 3 to 6 years old age range (Eyberg, 2009).

Infant CARE-Index was developed in the early 1980s as an observational measure to assess parent-infant interactions in a dyadic context in the age range from birth up to 15 months (Crittenden, 1981). It was developed by Crittenden, under the academic supervision of Mary Ainsworth; developer of SSP (Ainsworth, 1970). Infant CARE-Index is considered the oldest and versatile of the Dynamic-Maturational Model (DMM) methods for the assessment of attachment. The emphasis of DMM lies on protection and reproduction around which the behavioural organisation of a human develops (Crittenden, 2006). In this model, patterns of attachment are seen as self-protective mechanisms that display individual variations in how information relating to safety and threat is processed (Crittenden, 2008).

Seven items (facial expression, verbal expression, position, affection, turn-taking, control and choice of activity) are evaluated, which then contribute to the scores on three adult scales (sensitive, controlling or unresponsive) and four infant scales (cooperative, difficult, compulsive and passive). For each of the seven items, two points are assigned, either both on one scale (i.e. two points for difficulty in the facial item) or split between two scales (i.e. one point for difficulty and one point for cooperation in the facial item), with a total of 14 points, separately for the parent and the infant. The possible range for each scale (sensitive, controlling and unresponsive; cooperative, compulsive, difficult and passive) is from 0-14. These scores then contribute to an overall score of global/dyadic synchrony ranging from 0-14. The higher the parental sensitivity and infant co-operative scores, the better the parent-infant relationship. Using the global synchrony scores, parent-infant dyads are classified into one of four categories; highly sensitive (scores 11-14); adequately sensitive (scores 7-10); inept (scores 5-6) and at risk (scores 0-4). Scores ranging between 0 and 4 indicate the need for intervention. The Infant CARE-Index is unique in assessing the behaviour of both the participants of the dyad, as it simultaneously assesses parental sensitivity and also the infant's interactive patterns within a dyadic context. (Parfitt et al., 2013).

CARE-Index is available in two versions, namely Infant CARE-Index which is operational from birth up to 15 months and the Toddler CARE-Index , which is utilised

from 15 months up to 30 months. Trained and reliable coders are required to carry out the assessments. The training for Infant CARE-Index consists of an extensive 9-day course comprising of learning to code videos, coding them and receiving feedback, followed by additional practice and a reliability test. The reliability test consists of another 18 videos to be coded by the trainee coder, which are then send to Family Relations Institutes, USA for evaluation. Out of the various reliability levels (I, II, III+, III-, IV+, IV-) attained after attempting the reliability test, the coders must attain reliability level II to code independently.

1.4 Infant CARE-Index and how it differs from other assessment measures

Infant CARE-Index assesses the adult sensitivity in a dyadic context, which is characteristic of a specific relationship. This means that the same adult can exhibit different behaviours with different children (Crittenden, 2005). It emphasises on the 'interpersonal quality' rather than 'quantity' of behaviours. This leads to recognition of two opposite forms of insensitivity i.e. over and under-engagement. Another unique feature of Infant CARE-Index is its ability to assess the risk to relationships, which is more difficult to assess compared to demographic, physical or nutritional risk to individuals (Svanberg & Jennings, 2002). In fact, one of the most often missed diagnosis by other assessments concerning 'very high-risk dyads' can in fact be identified by the Infant CARE-Index. One such example is that of 'false positive affect' (that hides maternal hostility and infant displeasure) which can be identified by 'child's compulsive patterns' in the Infant CARE-Index observations. This added advantage makes it a useful screening measure for high risk dyads, which would be otherwise missed by other assessment measures.

In contrast to the SSP, where the child's self-protective strategies are elicited; the Infant CARE-Index is carried out in stress free and non-threatening conditions whether be it at home, clinic or research premises (Crittenden, 2010). Moreover, it takes only 3-5 minutes to administer, whereas, the SSP takes 20-25 minutes to administer. This short duration of administration complements to the ease of both administrator and parent-infant dyad, although coding the videos requires 20-25 minutes preceded by extensive training.

Infant CARE-Index is being used in many different countries with different cultures and practices (Leventhal, 2004). It can specifically discriminate between abusive and

non-abusive mothers and their infants through the interactional patterns observed by the coder (Cassidy et al., 1996; Crittenden, 1981). Crittenden defined child abuse as the application of parenting practices which were strict, punishing, controlling and rejecting in nature (Crittenden, 1992).

Moreover, Infant CARE-Index can be readily used for planning of appropriate interventions (Crittenden, 2005). It can help the practitioners in differentiating whether the distortions are in the 'affect' or 'cognitive' components of relations. It can help the practitioners in observing whether the parent and the infant overstress or diminish emotions. This helps in designing interventions specific to the nature of non-sensitive adult behaviour and a non-cooperative infant (Crittenden, 2005).

Furthermore, viewing the video recordings of parent-infant interaction (one's own or other mothers) can be used as an opportunity for analysis, reflection and development of observational skills by mothers. The developer of Infant CARE-Index stresses that if a mother is praised for carefully observing, verbalizing her thoughts and trying new skills with their infant, the intervention may pave way for positive reinforcement. This technique can be readily applied by practitioners who are delivering parenting programmes or home visits. Furthermore, Crittenden argues that, written plans, or teaching programs or manuals are all in conflict and it is essential that the mother needs a repertoire of interpersonal and observational skills, so she can adapt and react according to the specific needs of her child in a synchronous manner. Lastly, since, Infant CARE-Index is a non-stress inducing technique, it can be used numerous times, thus making it appropriate for programme evaluation. (Crittenden, 2005).

1.5 Rationale for the current research

With recent advances in randomised controlled trials and cohort studies, there is considerable evidence available that difficulties in parent-infant interactions can be identified as early as 10 months in children from general population (Skovgaard et al. 2008). However, recent research has highlighted the lack of analysis of measurement properties of the observational measures, that are being used to assess parent-infant interactions (Groh et al.2017). Recently a systematic review by Gridley et al in 2019 has highlighted the psychometric properties of parent-child interactions (0-5yrs) outcome measures used in randomised controlled trials evaluating parent programmes. However, this review has captured the more commonly used validated measures, leaving out the less commonly used measures which might have the

potential to adequately assess parent-child interactions, but are in need of good quality validation studies and utilisation within randomised controlled trials. Although Infant CARE-Index has been assessed in various systematic reviews such as those by Mesman et al. 2013 and Tryphonopoulos et al, 2014 and Lotzin et al. 2015, however despite being included in these reviews, it has been combined in the results and discussions along with other measures, rather than being exclusively reviewed. In the review by Tryphonopoulos, the measurement properties were jointly presented for both versions of the CARE-Index, and the reader is not able to distinguish, whether the mentioned property is related to Infant CARE-Index or Toddler CARE-Index. Similarly, the reliability evidence in the above-mentioned review is presented in various forms, rather than one standardised form and no detail about the agreed level of adequate reliability is mentioned (Tryphonopoulos et al, 2014). Similarly, in the review by Lotzin et al, the reliability/validity evidence is collectively provided on CARE-Index (infant and toddler version) for response process, rater training, internal structure, reproducibility, inter-rater reliability, test-retest, inter-scale correlation and discriminant validity (Lotzin et al, 2015). Although measurement properties of Infant CARE-Index are present in these reviews, however, it is not provided in a clear and detailed account and does not follow any standardised guidelines. Furthermore, since the reviews are on all the observational measures assessing parent-child interactions e.g. from birth up to 15 years (Lotzin et al, 2015), there is very limited transfer of knowledge to the reader, due to wide scope of these studies. Nevertheless, to date no systematic review has been published which has explicitly reviewed all available studies on the measurement properties of Infant CARE-Index in a standardised method. Similarly, neither the acceptability of Infant CARE-Index (from the perspectives of practitioners) has been explored in UK or any other cultural context.

The parent-infant interactions have a link to the process of attachment development, yet, as a construct in its own right, it still lacks clarity and comprehension. This is combined with a recognised need for fast and effective measure of parent-infant interactions in the community healthcare settings, which can act as an indicator of maternal and infant emotional well-being (NICE, 2015).

1.6 Aims & Objectives

This thesis aims to explore how an observational measure might be used for accurate assessment of the parent-infant interactions during the first 15 months of an infant's life within community settings. With this in mind, the research question asks:

How can the parent-infant interactions (birth-15 months) be assessed with the help of an observational measure (Infant CARE-Index) in association with its measurement properties and acceptability in practice?

In order to answer this question, and to meet the broader research aims, the objectives for this research study are as follows:

-To perform a systematic review on the measurement properties of an observational measure (Infant CARE-Index).

-To conduct a qualitative analysis on the acceptability of an observational measure (Infant CARE-Index) from the perspective of professionals involved in providing health services and research within the UK context.

This research thesis comprises of a systematic review on the measurement properties of Infant CARE-Index, followed by a qualitative analysis on the acceptability of Infant CARE-Index within UK settings. Lastly, an overall discussion on the findings from the systematic review and qualitative analysis will be undertaken.

Chapter 2: A review on the measurement properties of Infant CARE-Index

2.1 Introduction

This chapter describes the design of a search strategy intended to gather the relevant research on the measurement properties of Infant CARE-Index in order to meet the specified aims and objectives of this research thesis. In a review of a single observational measure, such as this, an overview is given on available evidence of each measurement property of Infant CARE-Index to come to overall conclusion per measurement property. This review followed the standardised procedure developed by **Co**nsensus-based standards for the **s**election of health **M**easurement **In**struments (COSMIN) steering committee in the 'COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMS)-user manual' (Mokkink et al, 2018).

Protocol registration:

The protocol was registered with PROSPERO, an international database of prospectively registered systematic reviews in health and social care, welfare, public health, education, crime, justice and international development, where there is a health-related outcome. The protocol is accessible at: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42018104208

2.2 Research questions

The research questions for this review asks:

-What measurement properties are reported in the studies utilising Infant CARE-Index for assessing parent-infant interactions from birth up to 15 months?

-What is the strength of evidence for reliability and validity of the Infant CARE-Index, when combined across multiple studies?

2.3 Procedure

2.3.1 Data sources and searches

Inclusion criteria:

-Full text primary studies utilising the Infant CARE-Index and reporting on measurement properties.

-Validation/developmental studies specifically designed to measure the measurement properties of Infant CARE-Index.

-Infants (Birth -15 months) and their parents.

Exclusion criteria:

-Prenatal/antenatal studies.

-Studies including children older than 15 months of age.

-No full text available, no measurement properties reported in paper.

-Articles published in a language other than English language.

-Conference abstracts, editorials, reviews, descriptive book chapters, user manuals, dissertations and protocols.

Literature Search

For the literature search, two types of articles were considered:

-Articles describing the development or application of the Infant CARE-Index e.g. description of the theoretical background, scale construction, or implementation and articles reporting primary research. Inclusion of such initial development papers and the content of the measure is considered crucial for extracting information on content validity of the measure.

-Articles evaluating the measurement properties (e.g., validation studies that evaluated the reliability and/or validity) of Infant CARE-Index.

Search strategy

The following electronic bibliographic databases were searched with the date ranges in brackets: APA PsycARTICLES Full Text, Embase (1980-2020 week 13), Maternity and Infant Care Database (MIDIRS 1971-Feb 2020), Social Policy & Practice (SOPP 202001), Ovid MEDLINE (R) ALL (1946-March 26, 2020), Econlit (1886-March 19, 2020), APA PsycINFO (1967-March week 4 2020), British Nursing Index (BNI),

ASSIA,CINAHL, Web of Science, Cochrane CENTRAL and ERIC. Searches were carried out iteratively during the course of the review, to capture any latest research studies that arose during the course of the review. The last searches were carried out in the first week of April 2020, hence the included studies and results are based upon the above-mentioned time point.

Key papers were recommended by supervisors early on and constituted the starting point for the handsearching of references and subsequent 'snowballing' of relevant citations and studies. Consultation with experienced librarian was sought for fine tuning of search strategy. In addition, citation tracking by screening of references of included full text articles, was conducted as a supplementary measure to identify any additional studies not captured by the database searching.

The updated list of Infant CARE-Index studies published in Dynamic Maturational Model (DMM) publications 2018, (DMM, 2018) provided list of the developmental studies. The DMM publications basically consist of a compilation of all kinds of publications associated with attachment and are published by 'International Association for the Study of Attachment' and these are updated regularly. Personal communication with the author of Infant CARE-Index was also helpful in getting access to the early studies. Upon literature review of these early studies, it was revealed that Infant CARE-Index was in fact originally denoted as 'maternal coding device' and 'infant coding device' (Crittenden, 1981; Crittenden et al, 1984; Crittenden, 1984). These measures were used separately, one for mother and one for infants (Crittenden, 1981). With further development and research, Crittenden used the term 'CARE-Index' in various studies, in which both the infant and maternal scales were jointly assessed (like in the present form). However, the age range of the infants varied widely from 1-29 months (Crittenden, 1987), 1-36 months (Crittenden. & Dilalla, 1988), birth to 48 months (Crittenden, 1988) and 1-48 months (Crittenden, 1992). Ward & Carlson mentioned CARE-Index as 'Crittenden's measure of sensitivity' in their study (Ward & Carlson, 1995). However, later in the early 2000s, CARE-Index was segregated into two versions, the infant version (birth-15 months) and the toddler version (15 -30 months), which are the present forms of CARE-Index. The present review only dealt with studies utilising 'Infant CARE-Index.'

Endnote was used for managing references. No language restrictions were applied during the searches; since the implication of potential language bias in systematic reviews is recognized. But due to time and resources constraint, articles in English

were included only. No restrictions were made as to the country or setting in which the research took place.

For those articles which were not easily accessible online, help was again sought from expert librarian at University of York. Those articles which were not easily accessible, requests were sent to the respective authors through Researchgate, which is a professional online network for researchers to share their research.

Search terms:

Search terms which incorporated the key concepts within the research question and aims were iteratively used in different combinations and modified according to the search functions of each individual database. Assistance in formulating search terms was sought from methodological PubMed search filter specifically designed for finding studies on measurement properties of measurement instruments developed by Terwee et al. in 2009. Wildcards were included to maximise the chances of all eligible papers being retrieved. An example of the search strategy can be found at the appendix (Appendix 1).

Following are the exact keywords that were used in all the databases:

1) parent OR mother OR father OR primary carer

2) Infant OR baby OR child OR toddler

3) CARE- Index

4) Interaction OR relationship OR attachment

2.3.2 Study selection

An initial screening of title and abstracts against the inclusion criteria was carried out, followed by full text screening and assessment of the identified relevant papers by the researcher. Though the gold standard for conducting systematic reviews is a minimum of two independent reviewers to be involved in the process, however, due to limited time and resources this was not possible (CRD, 2009; Higgins and Green, 2011). Discrepancies were resolved through discussion with the two supervisors of the researcher. The excluded studies were recorded and documented and the reasons for their exclusion were noted/mentioned in PRISMA diagram (Figure 2) which reports the flow of studies (Liberati et al., 2009).

Cochrane recommends reviewers to use various additional sources to find relevant studies for example dissertations, editorials, conference proceedings, and reports. However, COSMIN user manual states that the probability of finding any relevant

articles for systematic reviews of measurement tools in these additional sources is small (Mokkink et al, 2018). Hence, these were excluded at the title and abstract selection phase of articles.

2.3.3 Data extraction and quality assessment

An initial 805 records were identified through database searching and 3 additional articles through DMM publications 2018. After duplicates removal, 526 articles underwent title and abstract screen, leaving 115 articles for full text screen. Of these 115 articles; 17 articles had no psychometrics of Infant CARE-Index, 17 articles were not in English language, 8 articles had child age range above 15 months, 1 article was on Toddler CARE-Index, 4 articles did not meet the eligibility criteria, 17 articles were (conference proceedings, dissertation abstracts, narrative accounts, presentations, e-poster). The remaining 38 articles were selected for data extraction. While 3 articles from DMM publications were considered eligible for inclusion. In the end, a total of 41 articles underwent data extraction (see Figure 2).

The data was extracted on the following variables from the included 41 articles:

-Characteristics of the included samples i.e. Author, year, country, study type, number of measurements for Infant CARE-Index, Construct measured within the Infant CARE-Index (Maternal, infant, dyadic synchrony), sample size, attrition, main carer, mean age of carer, mean age of child, percentage of parents/carers who were females, gender distribution of infants, pre-dominant ethnicity, recruitment settings, location of observations and recruitment method. (Appendix 3)

-Measurement properties identified across the included studies.

While conducting a systematic review, it is essential not only to consider the results of included studies but also their methodological quality. For this purpose, the updated '**CO**nsensus-based **S**tandards for the selection of health **M**easurement **IN**struments (COSMIN) methodology for systematic reviews of patient-reported outcome measures (PROMS)-user manual' (Mokkink et al, 2018) was utilised to review the methodological quality of the included studies. Although COSMIN was developed for PROMS, however, upon consultation with the developers of COSMIN, modifications were permitted to be made to suite the nature of the present review on observational measure. The reason being, that at the time of conducting this review, no standardised guidelines or checklist were present on conducting a review on the measurement properties of observational measures. Following correspondence with the developers of the COSMIN manual, the researcher was advised that except for reliability/measurement error and perhaps content validity, all boxes from the

COSMIN Risk of Bias checklist for PROMs need not to be adapted, as they start from the final scores of an instrument. The advice was taken into consideration for data synthesis of measurement properties of Infant CARE-Index. Figure 1 illustrates the steps taken to conduct this review.

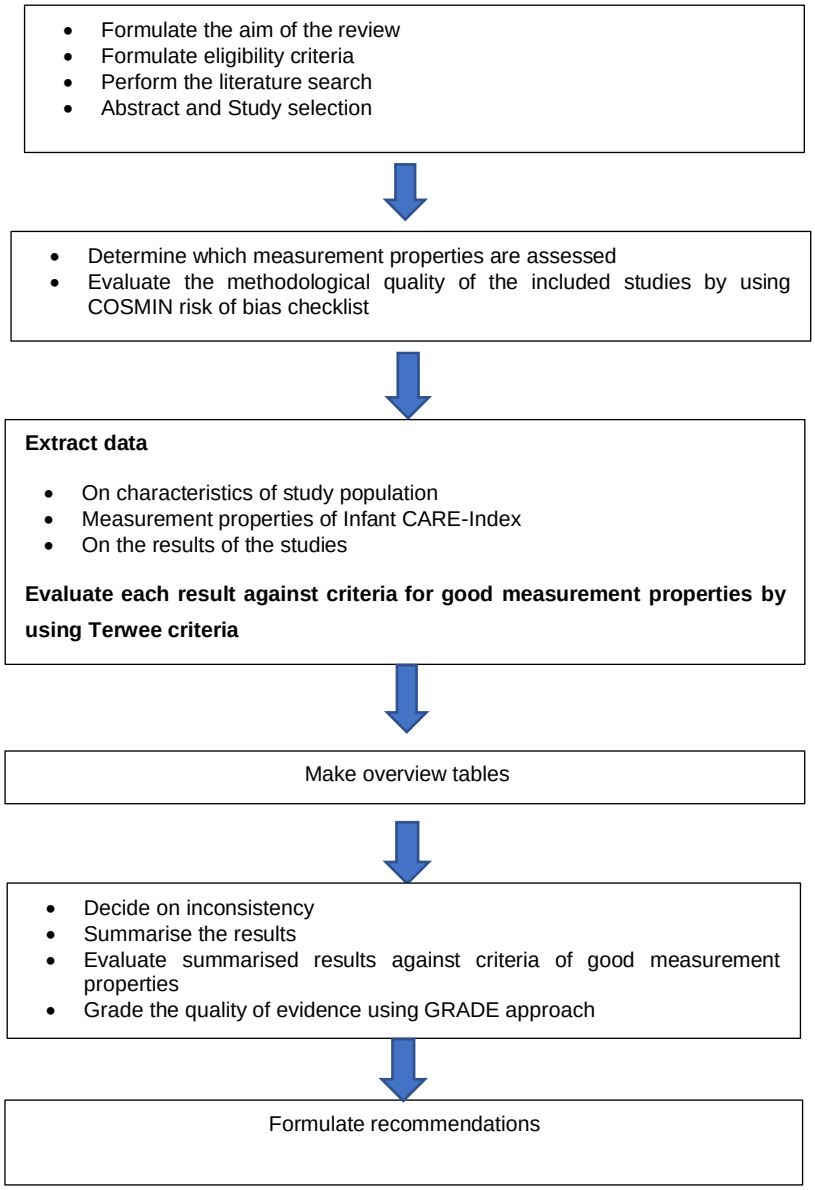


Figure 1: Steps for conducting a systematic review on the measurement properties of an observational measure

(Adapted from Prinsen et al, 2018 & Mokkink et al 2018)

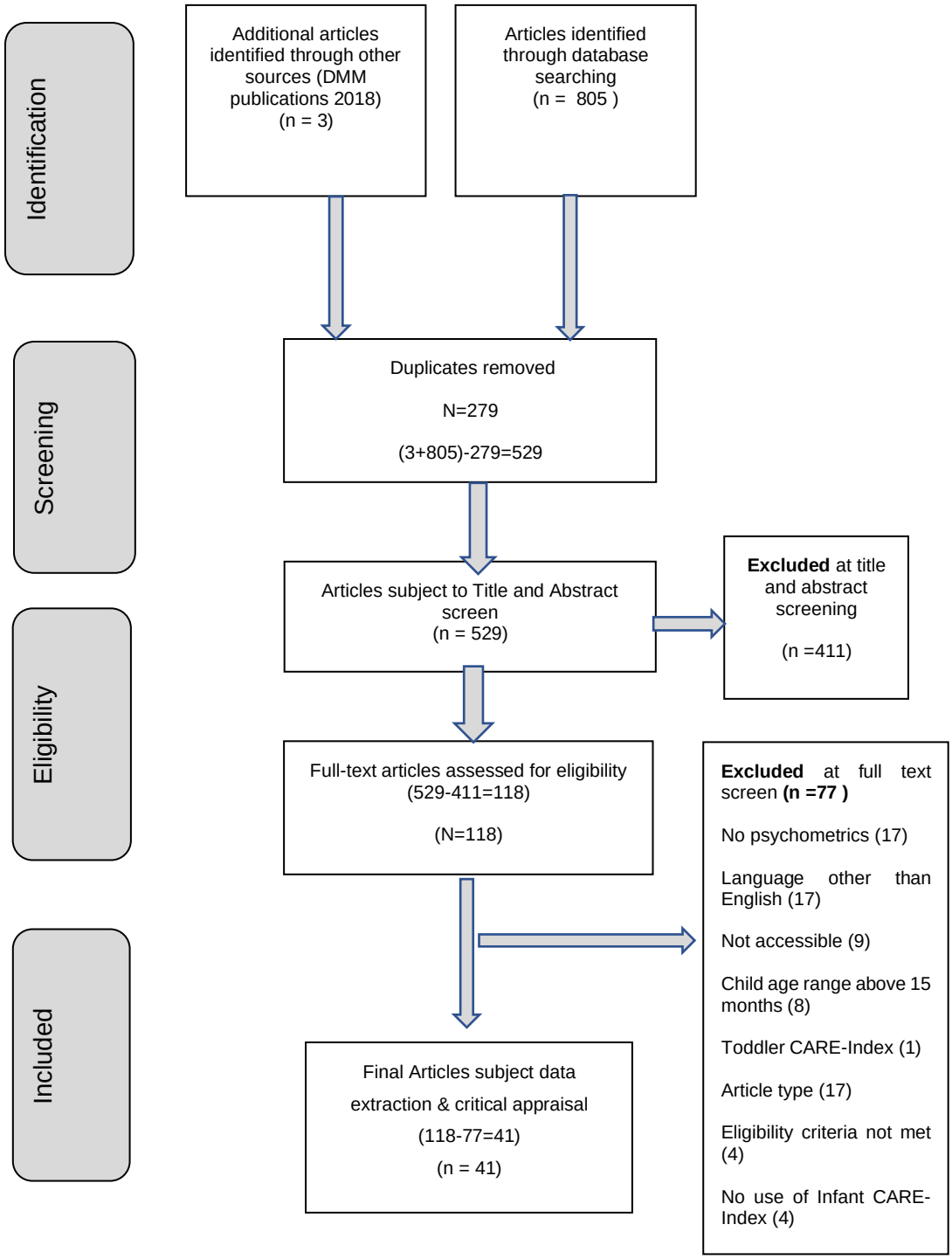


Figure 2: Flow chart of article retrievals on Infant CARE-Index

From: Liberati A, et al.(2009). The PRISMA Group.. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097).

* DMM: Dynamic maturational model publication

2.3.4 Data synthesis and analysis

The analysis of each measurement property included three sub steps:

Determining which measurement properties were assessed

There are 10 boxes in the checklist (see Table 1), which deal with instrument development, content validity, structural validity, internal consistency, cross-cultural validity/measurement invariance, reliability, measurement error, criterion validity, hypothesis testing for construct validity and responsiveness. Appendix 1 provides an overview of the definitions of the measurement properties. Utmost importance was given to consider how the design and analyses in the studies related to the COSMIN taxonomy (Mokkink et al, 2010). The rationale behind this was to ensure a consistent evaluation of the measurement properties across the included studies; even though the terminologies used by different authors varied considerably. The result of each single study reporting a measurement property was rated against the updated criteria for good measurement properties (Terwee et al, 2012) (see Table 3).

Evaluating the methodological quality of studies

The methodological quality of each single study reporting a measurement property was assessed using the COSMIN Risk of Bias checklist (Mokkink et al, 2017). The quality of each study was separately evaluated using the box corresponding to the measurement property in the COSMIN risk of bias checklist (see Table 1). The COSMIN Risk of Bias checklist was used as a modular tool; only those boxes were completed for the measurement properties, which were actually evaluated in the study. Each study was rated on a 4-point quality system as very good, adequate, doubtful or inadequate. The overall rating of the quality of each single study on a measurement property was determined by the lowest rating of any standard in the box assessed (i.e." the worse score counts' principle) (Terwee et al, 2012).

Summarising the evidence and application of GRADE

The evidence was summarised per measurement property. The overall result was rated against the criteria for good measurement properties and the quality of the evidence was graded using the **G**rading of **R**ecommendations **A**ssessment, **D**evelopment, and **E**valuation (GRADE) approach (Prinsen et al, 2018). COSMIN has modified the GRADE approach to be adapted for systematic reviews on measurement properties. The modified GRADE approach used in COSMIN manual takes into account four factors for determining the quality of evidence. (Mokkink et al, 2018). These are as follows:

(1) Risk of bias (i.e. the methodological quality of the studies)

(2) Inconsistency (i.e. unexplained inconsistency of results across studies)

(3) Imprecision (i.e. total sample size of the available studies)

(4) Indirectness (i.e. evidence from different populations than the population of interest in the review).

Table 1: Boxes of COSMIN risk of Bias checklist and order in which the measurement properties are assessed (Mokkink et al, 2018)

	Content validity
Box 1.	Instrument development*
Box 2.	Content validity
	Internal structure
Box 3.	Structural validity
Box 4.	Internal consistency
Box 5.	Cross-cultural validity\measurement invariance
	Remaining measurement properties
Box 6.	Reliability
Box 7.	Measurement error
Box 8.	Criterion validity
Box 9.	Hypotheses testing for construct validity
Box 10.	Responsiveness

* not a measurement property but taken into account when evaluating content validity.

2.4 Results

2.4.1 Study characteristics

The data was extracted and populated in the pre-formed overview tables developed in the Qualtrics software. Following information was retrieved:

-Characteristics of the included samples i.e. study design, sampling methods, mean age of infant and mean age of parent, primary carer, gender of infant, setting, country, language etc (Appendix 14). Out of a total 41 studies included, 14 studies were carried out in the UK, 5 in Portugal, 5 in Germany, 4 in Canada, 5 in Switzerland, 2 in Italy, and 1 study each in Chile, Denmark, Finland, Japan and USA, while two studies was conducted simultaneously in France, Germany and UK. All of the included studies were in English language. The age range of the infant population ranged from birth up to 15 months, which is the recommended age range for operation of Infant CARE-Index. The study designs included cross-sectional studies, longitudinal studies, quasi-experimental studies, randomised controlled trials, observational studies, cross-sectional studies, pilot studies, exploratory studies and mixed-methods studies.

2.4.2 Measurement properties assessed and data synthesis

Three measurement properties were identified across the included studies, which were reliability, hypothesis testing for construct validity and responsiveness.

Reliability evidence was retrieved from almost all the included studies mainly in the form of inter-rater reliability. Information on reliability levels and other reliability evidence (such as number of coders) was directly copy pasted from the studies into preformed tables (see Appendix 14). The exact levels of reliability (I, II, III, IV) according to Infant CARE-Index developer guidelines were not mentioned in any of the studies. In those studies, where it was mentioned that coders were trained by Crittenden (developer of Infant CARE-Index), it was assumed to be at level II, III or IV. Furthermore, due to limited details provided, this inter-rater reliability was assumed to be between coders in the study and not based on the official Infant CARE-Index reliability test, which all Infant CARE-Index trainees undergo.

To evaluate **hypotheses testing and responsiveness** for construct validity, hypotheses need to be formulated a priori. Since many authors had not explicitly formulated the hypotheses in advance, the researcher defined the hypotheses (see

Table 2) taking guidance from the COSMIN manual (Mokkink et al, 2018). The more hypotheses being tested, the more the evidence gathered. In many studies, more than one hypothesis was being tested, hence accounting for a greater number of hypotheses than the number of studies in the present review. The hypotheses a priori and findings regarding the correlation of Infant CARE-Index with other measures and its ability to discriminate between important patient subgroups were considered for extraction. (Appendix 5).

For **responsiveness**, studies with more than one observation and findings regarding the correlation of Infant CARE-Index with other measures, its ability to discriminate between important patient subgroups and before/after intervention evidence were considered for extraction. (Appendix 6).

Table 2: Generic hypotheses to evaluate construct validity and responsiveness

(Mokkink et al, 2018)

	Generic hypotheses
1	Correlations with (changes in) instruments measuring similar constructs should be ≥0.50
2	Correlations with (changes in) instruments measuring related, but dissimilar constructs should be lower, i.e. 0.30-0.50
3	Correlations with (changes in) instruments measuring unrelated constructs should be <0.30
4	Correlations defined under 1, 2, and 3 should differ by a minimum of 0.10.
5	Meaningful changes between relevant (sub)groups (e.g. patients with expected high vs low levels of the construct of interest).
6	For responsiveness, AUC* should be ≥ 0.70

In this review, the other measurement properties mentioned in Table 1 could not be assessed adequately due to the nature of Infant CARE-Index. Brief explanations for why they were not assessed are provided to aid the reader in developing a better understanding:

Content validity has been defined by COSMIN group as the first and most important property of measures as it is essential to the measures' usefulness in doing the job it is designed to do, and it influences all other measurement properties (Prinsen, 2018). A thorough assessment of a measure's content validity is essential and should include not only those studies which present information on content validity in the population of interest, but also consider initial development papers of a measure and the contents of the measure. However, none of the included studies were designed to test

specifically the measurement properties of Infant CARE-Index, thus content validity could not be identified in any of the included studies.

The second most important form of the validity assessment of a measure is **structural validity** (Chiarotto, 2018), which describes the degree to which the scores of a measure are an adequate reflection of the dimensionality of the construct to be measured (Prinsen et al, 2016). The term *index* is used for a measure comprising of multiple dimensions, which can be summarised in one score. However, Infant CARE-Index measures multidimensional constructs, i.e. each item represents a different dimension. The sum scores are based on observable constructs (Henrica et al, 2011). For this reason, **Internal consistency** could not be identified as CARE-Index is a multidimensional index and not a scale.

Similarly, no gold standard for Infant CARE-Index is explicitly stated in literature, hence **criterion validity** could not be assessed either.

Cross cultural validity was not applicable, as it concerns translation of the items of a measure. This is one of the domains where, COSMIN is restricted to PROMS, as the focus is on translating the questions/items for respondents, which is not the case for an observational measure like the Infant CARE-Index.

2.4.3 Methodological quality of the studies

Reliability was mentioned in **40** studies (out of a total 41 included in the review), in the form of Intraclass correlation coefficient (23 studies), Alpha coefficient (3 studies), Pearsons' correlation (5 studies), kappa (4 studies) and percentage agreement (5 studies). The methodological quality of each study was rated as either very good, adequate, doubtful or inadequate quality. **(none of the studies were very good, only 4 were adequate, 25 studies were doubtful and 11 were inadequate).**

Hypothesis testing for construct validity was carried out in **31** studies (out of a total 41 included studies in the review). Each study was rated as either very good, adequate, doubtful or inadequate quality. **(5 studies were very good, 1 was adequate, 21 studies were doubtful and 4 were inadequate).**

Responsiveness was assessed in total of **11** studies (out of a total 41 included in the review). Each study was rated as either very good, adequate, doubtful or inadequate quality. (**1 was very good, 9 studies were doubtful and 1 was inadequate).**

2.4.4 Quality of the assessed measurement properties

For **reliability,** COSMIN recommends ICC or weighted kappa as an acceptable measure. Each result was rated against updated criteria for good measurement properties and categorised as either sufficient (+) if ICC or weighted Kappa ≥ 0.70; insufficient (-) if ICC or weighted Kappa <0.70; or indeterminate(?) if ICC or weighted Kappa not reported (Terwee et al, 2007) (see Table 3). Applying this criteria, 27 studies were rated as sufficient and 13 studies were rated as indeterminate (see Table 3). These results for each single study on reliability can be visualised in Appendix 4.

For **Hypothesis testing,** the result of each single study on Hypotheses testing was rated against updated criteria for good measurement properties. Each result was rated as either sufficient (+), insufficient (−), or indeterminate (?). Applying this criteria, 31 studies were rated as sufficient and 9 studies were rated as insufficient (see Table 3). These results for each single study reporting on hypothesis testing can be visualised in Appendix 5.

For **Responsiveness, t**he result of each single study on Hypotheses testing was rated against updated criteria for good measurement properties. Each result was rated as either sufficient (+), insufficient (−), or indeterminate (?). Applying this criteria, 12 hypotheses were rated as sufficient and 3 hypotheses were rated as insufficient (see Table 3). The results for each single study reporting on responsiveness can be visualized in Appendix 6.

Table 3: Updated criteria for good measurement properties and summary results for Infant CARE-Index based on the synthesized evidence from the 41 articles reviewed.

(Terwee et al, 2007; Prinsen et al, 2016)

Measurement Property	Rating[1]	Criteria	Quality of the assessed measurement properties	Applying criteria for good measurement properties to the summarised result
Reliability	+	ICC or weighted Kappa ≥ 0.70	27 studies rated as sufficient	Sufficient
	?	ICC or weighted Kappa not reported	13 studies rated as indeterminate	
	-	ICC or weighted Kappa <0.70	None of the studies rated as insufficient	

Hypotheses testing for construct validity	+	The result is in accordance with the hypothesis[2]	31 studies rated as sufficient	Sufficient
	?	No hypothesis defined (by the review team)	None of the studies rated as indeterminate	
	-	The result is not in accordance with the hypothesis	9 studies rated as insufficient	
Responsiveness	+	The result in accordance with the hypothesis[2]	12 studies rated as sufficient	Sufficient
	?	No hypothesis defined (by the review team)	None of the studies rated as indeterminate	
	-	The result is not in accordance with the hypothesis[2]	3 studies rated as insufficient	

Key: [1] "+" = sufficient, " –" = insufficient, "?" = indeterminate

[2] The results of all studies should be taken together, and it should then be decided if 75% of the results are in accordance with the hypotheses

2.4.5 Summarising the evidence

The previous section dealt with the quality of **single studies** on measurement properties of Infant CARE-Index, this section will focus on the quality of evidence on Infant CARE-Index **across all the studies combined.**

Applying criteria for good measurement properties to the summarised result

The summarised result per measurement property for Infant CARE-Index were **again** rated against the same quality criteria for good measurement properties by Terwee et al (Table 3)(Prinsen et al, 2016; Prinsen et al, 2018; Terwee et al, 2007). The overall rating for the summarised result was categorised as sufficient (+), insufficient (-), inconsistent (+/-), or indeterminate (?). This rating led to the formulation of summarised results on the measurement properties of Infant CARE-Index.

Reliability: Terwee's criteria for good measurement properties (Terwee et al, 2007) considers *ICC or weighted Kappa ≥ 0.70* as sufficient, *ICC or weighted Kappa not reported* as insufficient and *ICC or weighted Kappa <0.70* as indeterminate (see Table 3).

If taking into account (ICC) which was mentioned in 23 studies, the range (lowest and highest) of results for maternal scales, infant scales and dyadic synchrony is as follows:

-ICC maternal: 0.70-0.98

-ICC Infant: 0.81-0.92

-ICC Dyadic synchrony: 0.78-0.89

The above ranges indicate that ICC for all the parameters is >_ 0.70, thus leading the overall rating to be **sufficient.** (see Appendix 4)

Hypothesis testing: Terwee criteria for good measurement properties considers sufficient (+) as when the result is in accordance with the hypothesis, Insufficient(-) when no hypothesis defined (by the review team) and indeterminate (?) as when the result is not in accordance with the hypothesis.

Furthermore, for the overall result to be sufficient quality, COSMIN recommends that at least 75% of the results should be in accordance with the hypotheses. (Table 3)

-32 out of 41 results on hypotheses testing(78%) were in accordance with the hypothesis, thus leading to an overall rating of sufficient. (see Appendix 5)

Responsiveness: Terwee criteria for good measurement properties considers sufficient (+) as when the result is in accordance with the hypothesis, Insufficient(-) when no hypothesis defined (by the review team) and indeterminate (?) as when the result is not in accordance with the hypothesis.

Furthermore, for the overall result to be sufficient quality, COSMIN recommends that at least 75% of the results should be in accordance with the hypotheses. (Table 3).

-12 out of 15 results on responsiveness (80%) were in accordance with the hypothesis, thus leading to an overall rating of sufficient. (see Appendix 6)

2.4.6 Grading the quality of evidence

After summarising the overall evidence per measurement property for Infant CARE-Index and rating the summarised result against the criteria for good measurement properties, the final step was to grade the quality of evidence, which was an indicator of how trustworthy the summarised results were. The grading of quality was formulated upon the (GRADE) approach for systematic reviews of clinical trials (GRADE, 2013). Accordingly, the quality of evidence was graded as high, moderate, low, or very low. (see Table 4 for definitions).

The quality of evidence according to modified GRADE approach was based upon four factors namely, risk of bias, inconsistency, imprecision and indirectness (see Table 5) for the summarised result of the three measurement properties assessed in this review (i.e. Reliability, Hypothesis testing for construct validity & Responsiveness).

Table 4: Definitions of quality levels

(Mokkink et al, 2018; GRADE, 2013)

Quality level	Definition
High	The researcher is very confident that the true measurement property lies close to that of the estimate* of the measurement property.
Moderate	The researcher is moderately confident in the measurement property estimate: the true measurement property is likely to be close to the estimate of the measurement property, but there is a possibility that it is substantially different.
Low	The researcher's confidence in the measurement estimate is limited: the true measurement property may be substantially different from the estimate of the measurement property.
Very low	The researcher has very little confidence in the measurement property estimate: the true measurement is likely to be substantially different from the estimate of the measurement property.

*Estimate of measurement property refers to the **summary of results** of the measurement property of Infant CARE-Index

Table 5: Modified GRADE approach for grading the quality of evidence

(Mokkink et al, 2018)

Quality of evidence	Lower if	
High	Risk of bias 1. Serious 2. Very serious 3. Extremely serious	Methodological quality of studies
Moderate		
Low	Inconsistency 1. Serious 2. Very serious	Inconsistency in results of studies
Very Low	Imprecision 1. Total n=50-100 2. Total n<50	Total sample size of the available studies
	Indirectness 1. Serious 2. Very serious	Evidence from populations other than populations of interest

N=Sample size

The application of these four GRADE factors in evaluating the quality of evidence on Infant CARE-Index is described as follows:

1) Risk of bias

This occurs when the COSMIN risk of Bias checklist assesses the quality of the study as doubtful or inadequate or if only one study of adequate quality is available.

Reliability: The quality of evidence available for **reliability** was in the form of 4 adequate studies, 25 doubtful studies and 11 inadequate studies. According to COSMIN guidelines (see Table 5), there was no risk of bias, as there were multiple studies of at least adequate quality (4 studies in this case) (see Table 6 for definitions of risk of bias). Thus, overall there was **high quality of evidence** available on the **reliability** of Infant CARE-Index, when considering risk of bias.

Hypotheses testing: Similarly, the same guidelines were followed for **Hypotheses testing.** There were 5 very good studies, 1 adequate study, 21 doubtful studies and 4 inadequate studies. According to COSMIN guidelines (see Table 5), there was no risk of bias, as 'there were multiple studies of at least adequate quality (1 study in this case), or there was one study of very good quality (5 studies in this case) (see Table 6 for definitions of risk of bias). Thus, overall there was **high quality of evidence** available on the hypotheses testing of Infant CARE-Index, when considering risk of bias.

Responsiveness*:* Again, the same guidelines were followed for r**esponsiveness.** There was 1 very good quality study, 9 doubtful studies and 1 inadequate study. According to COSMIN guidelines (see Table 5), there was no risk of bias, as there was one study of very good quality available. Thus, the overall there was **high quality of evidence** available on responsiveness of Infant CARE-Index, when considering risk of bias.

Table 6: Guidelines and definitions on downgrading for Risk of Bias

(Mokkink et al, 2018)

Risk of bias	Downgrading for Risk of Bias
No	There are multiple studies of at least adequate quality, or there is one study of very good quality available
Serious	There are multiple studies of doubtful quality available, or there is only one study of adequate quality
Very serious	There are multiple studies of inadequate quality, or there is only one study of doubtful quality available
Extremely serious	There is only one study of inadequate quality available

2) Inconsistency

COSMIN recommends that in order to establish an overall conclusion on the quality of a measure, it needs to be decided whether the results of all available studies per measurement property are **consistent** or **inconsistent.**

Handling consistent results:

For consistent results, COSMIN recommends that the results may be qualitatively summarised, followed by comparison against the criteria for good measurement properties (Terwee et al, 2012). This will lead to determining whether overall, the measurement property of the measure is sufficient (+), insufficient (-), inconsistent (+-), or indeterminate (?). (see Table 3).

The results for **hypothesis testing for construct validity and responsiveness** followed a consistent pattern in results, hence the results were qualitatively summarised to provide overall ratings. Hence there was **no need to downgrade for inconsistency for hypothesis testing and responsiveness** (see Table 5)

Handling inconsistent results:

For inconsistent results, COSMIN recommends three pathways:

a) Give explanation and summarise per subgroup;

b) Do not summarise the results and do not grade the evidence;

c) The conclusions may be based on the majority of consistent results, alongside down grading for inconsistency.

The results for **reliability evidence** followed the above-mentioned inconsistent pathway (c). The reason being that the reliability evidence was mentioned in the form of different parameters like ICC (23 studies), Alpha coefficient (3 studies), Pearson's correlation (5 studies), kappa (4 studies) and percentage agreement (5 studies). This inconsistency was resolved by including the results of only those studies which followed Terwee's criteria on good measurement properties, (see Table 3) i.e. the reliability evidence presented in the form of ICC (for continuous scores) and kappa (for categorical scores) was only eligible for inclusion as a measurement property for systematic reviews on measurement properties. For simplicity, the 23 studies which had reliability mentioned in the form of ICC were utilised for summarizing the overall results on reliability.

Furthermore, ICC was reported in three forms:

-ICC for maternal component of Infant CARE-Index

-ICC for infant component of Infant CARE-Index

-ICC for dyadic/global synchrony

According to Terwee's criteria for good measurement properties; ICC or weighted kappa should be >_ 0.70 to be considered as sufficient measure of reliability (see Table 3) (Terwee et al, 2007; Mokkink et al, 2018). The ranges (lowest to highest) for all these above-mentioned forms of ICC's were accumulated together from the 23 studies mentioning 'Inter-rater reliability' in the form of ICC.

-ICC maternal : 0.70-0.98

-ICC Infant: 0.81-0.92

-ICC Dyadic synchrony: 0.78-0.89

As is evident from the above-mentioned ranges, the ICC for all the forms i.e. maternal, infant and dyadic synchrony are **all above 0.70**, which is the value considered acceptable to rate the measurement property of reliability as of **sufficient** quality. Hence, **no downgrading for inconsistency was required for reliability evidence**.

3) Imprecision

This is related to the total sample size of the included studies. COSMIN recommends to down grade the quality of evidence, if the total sample size of the summarised studies is below 100, however, the total sample size of the summarised studies was well above 100. Hence there was **no need to downgrade for imprecision** (see Table 5).

4) Indirectness

In the context of this systematic review, indirectness occurred as the studies included in the review were performed in another context of use than the context of use of interest in the systematic review, as none of the included studies were designed to precisely assess the measurement properties of Infant CARE-Index. Thus, **it was decided to downgrade with two levels for very serious indirectness** (see Table 5).

To determine the overall grading for the quality of evidence, the GRADE factors concerning the quality of the evidence were summarised separately for reliability, hypotheses testing and responsiveness as follows:

Overall quality of evidence using GRADE approach:

Reliability: First, the *risk of bias* was considered and there was high quality of evidence present, hence no downgrading was done. Next factor considered was *inconsistency*, for which no downgrading was needed. Similarly, for *imprecision,* again, no downgrading was needed. Lastly, due to *indirectness*, the evidence was downgraded by two levels, taking the overall evidence level from high to low for reliability. Thus, according to GRADE approach, there is **'low'** quality evidence that reliability is sufficient. (Table 5 & 7).

Hypothesis testing: First, the *risk of bias* was considered and there was high quality of evidence present, hence no downgrading was deemed necessary. Next factor considered was *inconsistency*, for which no downgrading was needed. Similarly, for *imprecision,* again, no downgrading was needed. Lastly, due to *indirectness*, the evidence was downgraded by two levels, taking the overall evidence level from high to low for hypotheses testing. Thus, according to GRADE approach, there is **'low'** quality evidence that Hypotheses testing is sufficient. (Table 5 & 8).

Responsiveness: First, the *risk of bias* was considered and there was high quality of evidence present, hence no downgrading was deemed necessary. Next factor was i*nconsistency*, for which no downgrading was needed. Similarly, for *imprecision,* again, no downgrading was needed. Lastly, due to *indirectness*, the evidence was downgraded by two levels, taking the overall evidence level from high to low for responsiveness. Thus, according to GRADE approach, there is '**low'** quality evidence that responsiveness is sufficient.(Table 5 & 9).

Summarising the evidence qualitatively

After applying GRADE approach in determining the overall quality of evidence on Infant CARE-Index, the results of the studies were qualitatively summarised. Following the COSMIN guidelines, the summary of Findings (SoF) (see Tables 7, 8 & 9) were constructed which comprised of:

-The summarised results per measurement property of Infant CARE-Index.

-An overall rating on the measurement properties of Infant CARE-Index as sufficient (+), insufficient (-), inconsistent (+/-) or indeterminate(?).

-The quality of evidence according to GRADE approach (high, moderate, low, very low).

The SoF tables were eventually used in formulating recommendations for the use of Infant CARE-Index.

Reliability: The majority of the studies (27 out of a total 40) reported the reliability evidence in the form of intraclass coefficient (ICC)(23 studies) and kappa (4 studies), which is the preferred statistical method of reporting reliability (Mokkink et al, 2018). The range of reliability scores across the 23 studies for the maternal and infant scales individually and when taken together in the form of Dyadic synchrony were in the range considered as **sufficient** according to criteria for good measurement properties (Terwee et al, 2012). When the GRADE approach was applied, the quality of evidence was rated as **'low'**. (See Table 7).

Table 7: Summary of Findings table (Reliability)

Reliability	Summarised results (Range)	Overall rating on measurement properties	The Quality of evidence according to GRADE approach
Infant CARE-Index	ICC maternal: 0.70-0.98 ICC Infant: 0.81-0.92 ICC Dyadic synchrony: 0.78-0.89	Sufficient	Low

Hypothesis testing: The evidence on hypotheses testing was based on the principle that at least 75% of the results should be in accordance with the hypotheses to rate the overall results as sufficient. In the present review, this criterion was duly met, as 32 out of 41 hypotheses (78%) were confirmed, leading to an overall rating of **sufficient**. Since, some studies tested more than one hypothesis; hence they were counted individually leading to more hypothesis than the number of studies assessing hypothesis testing in the review. When the GRADE approach was applied, the quality of evidence was rated as **'low'**. (See Table 8).

Table 8: Summary of findings table (Hypotheses testing)

Hypotheses testing	Summarised results	Overall rating on measurement properties	The Quality of evidence according to GRADE approach
Infant CARE-Index	32 out of 41 hypotheses confirmed (78%)	Sufficient	Low

Responsiveness: The evidence on responsiveness was based on the principle that at least 75% of the results should be in accordance with the hypotheses to rate the overall results as sufficient. In the present review, this criterion was almost met, as 12 out of 15 hypotheses (80%) were confirmed in the studies assessing responsiveness, leading to an overall rating of **sufficient**.(see Table 9).

Table 9: Summary of findings table (Responsiveness)

Responsiveness	Summarised results	Overall rating on measurement properties	The Quality of evidence according to GRADE approach
Infant CARE-Index	12 out of 15 hypotheses confirmed (80%)	Sufficient	Low

2.4.7 Formulation of recommendations according to COSMIN

COSMIN recommends classifying a measure into three categories:

(A) A measure with evidence for sufficient content validity.

(B) A measure categorised not in A or C.

(C) Measures with high quality evidence for an insufficient measurement property.

Measures categorised in category (A) can be recommended, while measures categorised in the category (C) should not be recommended for use. Whereas, measures coming under the category (B) have the potential to be recommended for use. In this review, no evidence whatsoever was identified for content validity of Infant CARE-Index. However, sufficient measurement properties were reported for reliability, hypotheses testing and responsiveness (see Tables 7, 8 & 9). Hence, Infant CARE-index was placed in category (B). Thus, there is a need for further evidence on measurement properties of Infant CARE-Index before it can be recommended for use in routine and research settings for assessing parent-infant interactions from birth up to 15 months.

2.5 Discussion

In this systematic review, the published scientific evidence on the measurement properties of Infant CARE-Index was thoroughly evaluated. Extensive searching techniques, using a combination of initially broad and then more focused terminology (Appendix 2), led to a total of forty-one papers being identified. COSMIN was adapted to the context of this research and combined the assessment of quality and risk of bias for each of the included papers. Results revealed evidence on three measurement properties i.e. reliability, hypotheses testing for construct validity and responsiveness of the Infant CARE-Index.

A clear lack of background knowledge on how to report the reliability evidence on Infant CARE-Index was evident amongst most of the included studies. COSMIN recommends that reliability be reported as intraclass correlation coefficient (ICC) for continuous scores and Cohen's kappa for dichotomous scores. However, many authors presented the reliability evidence in the form of percentage agreement, Pearsons' correlation and also alpha coefficient, which resulted in exclusion of such studies in the final synthesis of evidence on reliability of Infant CARE-Index. Moreover, several studies did not use the developer mentioned target age range (0-15 months) for the use of Infant CARE-Index, hence leading to many studies being screened out (see Figure 2). Similarly, many studies failed to mention explicitly on the number of coders involved in coding Infant CARE-Index videos and the level of reliability of such coders to the standards laid down by FRI. Many authors did not clearly describe the test conditions, location of administration of Infant CARE-Index and the time period between two administrations (applicable in case of two observations for assessing responsiveness).

Many instruments that are being utilised in practice are without the accompanying good quality evidence that supports their validity for this task (Mokkink et al, 2018). When the evidence is available, most of it is either of a low quality, featuring insufficient detail in the published articles to make systematic and comprehensive assessments on validity and reliability as demanded by COSMIN, leading to doubtful ratings eventually. (Mokkink et al, 2018).

2.5.1 Strengths

The strengths of this study include the comprehensive literature search in various databases, as well as the systematic approach applied during the entire review process. In addition, this study is the first to critically review various measurement properties of Infant CARE-Index including the assessment of the methodological quality of studies reporting on these measurement properties. For this purpose, standardised criteria were used to assess both the methodological quality of the included studies using the COSMIN checklist (Mokkink et al, 2018), as well as the quality of the measurement properties using quality criteria that were proposed by Terwee. (Terwee et al, 2007).

This review highlighted the areas where evidence on measurement properties of Infant CARE-Index is lacking. It has also sign posted the elements that are often missed, when authors report the measurement properties of infant CARE-Index in a research study. Many of the studies included had samples both from clinical and non-clinical population, along with representation from various ethnic backgrounds (Appendix 3). This ability to bring together diverse study populations and interpret their shared findings in a wider, more diverse context is considered a strength of a systematic review (Mulrow, 1994).

2.5.2 Limitations

The limitations faced in this review was that after carrying out the searches, the results revealed that almost none of the studies were specifically designed to assess measurement properties of Infant CARE-Index, due to which only three measurement properties could be identified.

The results on measurement properties were further scrutinised by COSMIN's worst score counts system leading to many of the studies being identified as doubtful, due to subjective analysis for some questions. Hence, it is important to recognise that although the methodological approach of the review was robust and followed COSMIN guidelines, but it must be acknowledged that the rating criteria was equally strict.

For Hypothesis testing, COSMIN recommends use of correlations between the instrument under study (Infant CARE-Index) and comparator instrument and changes

in mean scores as the preferred statistical method of presenting the results. However, the evidence on hypotheses testing for construct validity was faced with challenge when the results from individual studies was accumulated, due to the variation in use of statistical methods by the authors in their respective studies.

Another limitation encountered was on the fact that Cochrane guidance recommends the need for at least two researchers to carry out assessments of risk of bias (Higgins and Green, 2011), but this was not feasible due to time and resource constraints. For this reason, every effort was made to ensure that all aspects were as transparent as possible.

2.5.3 Recommendations

Infant CARE-Index was developed when the measurement properties were not explicitly researched and published in a systematic checklist manner. The content validity was based on theory and clinical expertise, rather than being investigated on a pre-defined systematic checklist. The methodology of reporting on the initial stages of instrument development has gained significance since last decade or so, thus leading to extensive research and delphi studies, of which COSMIN is one prime example (Mokkink et al, 2010). Although such checklists provide a step by step guide to assess the measurement properties in detail, but researchers are faced with the limitation in terms of journal restrictions on word count, which leads to limited knowledge being transferred to the reader. However, there is still scope for systematic validation of these instruments in appropriately designed studies, which should follow standardised checklists (Mokkink et al, 2019).

Future studies should focus on reliability (inter-rater, intra-rater and test-retest) and validity evidence with a precise description of the construct which is being measured. Multiple measurements performed in stable patients by at least two reliable observers should be the focus of upcoming validation/ reliability studies. Furthermore, in order to monitor an individual dyad over time in clinical or research settings, agreement parameters such as measurement error and the smallest detectable change should be calculated.

The authors must practice consistency and utmost care must be taken in using Infant CARE-Index for the recommended age range from birth up to 15 months. While, the version of CARE-Index (Infant or Toddler) and the reliability level of the coders must be precisely mentioned in the studies.

Furthermore, majority of the studies focussed specifically on the mother-infant interactions, further studies should seek to assess interactions amongst father-infant or other caregiver-infant interactions, as Infant CARE-Index is designed such that it can be used with any caregiver and infant. Again, good quality validity studies need to be the initiating point for such research.

2.5.4 Conclusions

This is one of the first few reviews that has utilised the updated COSMIN guidelines published in 2018 and has used the latest recommended methods in identification of evidence, data extraction, data assessment and data synthesis. The robust but strict rating system of COSMIN led to low ratings of Infant CARE-Index. But as mentioned earlier, this instrument was developed well before the emergence of literature reporting on the instrument development methods. Hence, there is no evidence to state that development of Infant CARE-Index was inconsistent; if compared to the present criteria established by COSMIN on judging the reliability and validity of a measure.

Moreover, this work is limited by its reliance on studies that were not designed with psychometric modelling in mind or to capture all relevant aspects of measurement properties of Infant CARE-Index. Nevertheless, the evidence on measurement properties that has been retrieved was rated as sufficient and hence offers a potential means of monitoring and improving the course of infants' social and emotional development, if further explored in good quality validity studies.

Out of a total of 41 studies included in this review, 14 studies had been carried out in the UK. Although, several other studies carried out in the UK were also identified while carrying out searches (but excluded due to not meeting the inclusion criteria of the review). Hence, this indicates that although there has been ongoing research in the UK which has been utilising Infant CARE-Index, but no study according to the researcher's knowledge has comprehensively explored the measurement properties of Infant CARE-Index.

For a measure to be used in routine practice, it is essential to study its measurement properties and also its practical acceptability from the perspectives of professionals involved in its use. Keeping this in mind, the researcher conducted a qualitative analysis of the acceptability of Infant CARE-Index from the perspective of health professionals involved in the use of Infant CARE-Index within the UK context. The next section of this thesis will describe the qualitative analysis. In broader context, this

research will inform the gap in evidence base for a measure needed for assessing parent-infant interactions in the first year of life and how acceptable that measure will be in routine practice to achieve the overall aim of socioemotional wellbeing of children in the early years.

Chapter 3: Qualitative analysis on the acceptability of Infant CARE-Index in the UK

3.1 Introduction

Although the Infant CARE-Index has been utilised in assessing parent-infant interactions as a part of various studies and trials in the UK (Svanberg et al., 2010, Barlow et al., 2016, Steadman et al., 2007, Kenny et al., 2013, Parfitt et al., 2013, Stephenson et al., 2018, Rigby et al., 2016, Barnes et al., 2017), little is known about its acceptability from the perspective of health professionals (who provide health services to 0-19 years) and researchers using the Infant CARE-Index in the UK. Regular feedback and recommendations from service providers and recipients of an assessment measure helps in minimising the unintended consequences and in turn improves the accuracy, validity and clinical utility of assessment measures (Hackmann, 2017 & Kendall et al, 2019).

This study intends to explore, if the Infant CARE-Index is deemed acceptable in the UK settings. The term **acceptability** is an increasingly important factor being considered in the design, evaluation and implementation of healthcare interventions and technologies/measures (Sekhon, 2017). In this study, 'acceptability' is being broadly operationalised to indicate the individual perspective or judgement about the practicalities of Infant CARE-Index.

3.1.1 Aims

The primary aim is to conduct a qualitative analysis on the acceptability of an observational measure (Infant CARE-Index) from the perspective of professionals involved in providing health services and research within the UK context. This will help in addressing the gaps in evidence (regarding the use of an appropriate assessment measure for assessing parent-infant interactions in children 0-12 months) as indicated in NICE recommendations for 'Assessing attachment difficulties in children and young people in all health and social care settings' (NICE guidelines, 2015).

This study was carried out by a post-graduate student, as a part of her research thesis and had previously studied a module on Qualitative research methods.

3.1.2 Research questions

Overarching research question: How acceptable is the Infant CARE-Index in the UK?

Following are the sub questions:

i) How is the Infant CARE-Index used in combination with other measures?

ii) What are the current assessment procedures and support provided to the parent-infant dyads (0-15 months) in the UK services?

iii) Is Infant CARE-Index video recording acceptable to parents generally?

iv) What are/might be the barriers/facilitators in implementation of Infant CARE-Index in the UK?

3.2 Methods

Two focus groups were conducted with two independent samples:

i) Infant CARE-Index trainees
ii) Health professionals not trained in Infant CARE-Index

3.2.1 Eligibility criteria

Inclusion criteria for focus groups

- Adult Participants (over the age of 18 years).
- Able to communicate in English language and through email.

Exclusion criteria for focus groups

- Not able to communicate in English and through email.

3.2.2 Participant selection, sample size and data collection

Purposive sampling was used in selecting participants, on the assumption that they would provide rich, significant and diverse data in relation to the research questions. This kind of sampling was selected as it involved choosing study participants based on the purpose of their involvement in the study. Guest et al define various types of purposive sampling and the most appropriate type of sampling for this study was ‘convenience’ (Guest et al, 2017). In this type of sampling, the people are sampled simply because they will act as convenient and approachable source of data for the researcher.

Guest (2017) recommends carrying out at least three focus groups per population. It was estimated that 1-2 focus group (having 3-5 participants) per the two sets of populations would be sufficient. The focus groups would comprise of semi-structured discussions conducted in accordance with a predetermined set of prompts and questions. Five types of questions were formulated i.e. opening question, introductory question, transition questions, key questions and ending questions to frame the focus group into a systematic process (Krueger, 2002).

Focus group with Infant CARE-Index trainees: The first focus group was designed to be held with individuals who undertook the training in Infant CARE-Index at University of York; but had not undertaken the reliability test. The aim was to explore the training of Infant CARE-Index and their experiences while learning the use of this measure, alongside their future perspectives regarding the acceptability of Infant CARE-Index in UK settings. The researcher had previously undertaken Infant CARE-Index training (held a Level IV+ reliability certificate) and was familiar with the content of the training.

Focus group with health professionals: The second focus group was designed to be held with health professionals who provide health services to 0-19 years (not trained in Infant CARE-Index). The aim was to explore the current assessments being carried out in the UK for assessing parent-child interactions in this age group (0-15 months) alongside exploring the acceptability of Infant CARE-Index.

3.3 Procedure

Due to time constraints and availability of the participants, only one focus group per population could be facilitated.

3.3.1 Focus group with Infant CARE-Index trainees*:* Out of the six female participants approached through email, five agreed to participate in the focus group. The participant who did not participate had other commitments. Out the five participants, three participants were involved in academic research, while the remaining two were practising psychologists. This focus group was moderated by the researcher at the University of York premises, having no non-participants or observers; on the second last day of the 9-day training course of Infant CARE-Index in July 2019. Permission to conduct the focus group was sought from the Department of Health Sciences, University of York. The participants were provided with a brief introduction on the topic

by the researcher and also provided with participant information sheet/consent forms prior to the initiation of focus groups (see Appendices 7 & 8).

3.3.2 Focus group with health professionals*:* The participants were approached through the '0-19 Research Network: Yorkshire and the Humber'(http://clahrc-yh.nihr.ac.uk/resources/e-newsletter-clahrc-partnership-programme/-0-19-research-network-yorkshire-and-the-humber), which is an innovative initiative that aims to enhance the research engagement and capacity of the 0-19 public health workforce in the Yorkshire and Humber region. Its audience includes health visitors, social workers, school nurses etc. The researcher conducting this research, attended one of these meetings and briefly mentioned about the proposed study and some of the health professionals showed interest in participating. One of these health professionals then went on to organise the focus group.

This focus group was moderated by the researcher at the beginning of a routine general meeting (Locala partnership) of health visitors at a selected venue at Huddersfield in October 2019, with no non-participants or outsiders. Twenty female health professionals comprising of health visitors and school nurses took part. The participants were provided with brief introduction on the topic by the researcher alongside the participant information sheets/consent forms to aid them in understanding the researcher's focus and to encourage the participants in expressing their view- points.(see Appendices 10 & 11).

In both focus groups, the researcher used prompt questions to probe answers in more detail (Appendices 9 & 12). Both the focus groups lasted approximately 20-25 minutes. The duration was based on the feasibility of the participants taking into view their professional schedule and commitments. Both the focus group discussions were audio-recorded after consent and were transcribed verbatim according to a simplified transcription system by the researcher before carrying out analysis.

3.3 Analysis

Various analytic methods were explored to select the most suitable method for carrying out the analysis in this research study. Initially, Interpretative phenomenological analysis (IPA) was studied, but it was concerned on how people make sense of their lived experiences, and how people perceive and speak about objects and events and researcher is supposed to make sense of these experiences

through interpretation (Clarke & Braun, 2013), however, this was not the case in this study. Similarly, grounded theory was not applicable either, as it is concerned with developing theory from data and there is emphasis on understanding social processes (Charmaz, 2006). Discourse analysis was not appropriate either, as it was concerned with patterns in language connected to the social production of reality, and with perceiving how the accounts of objects and events are created in specific manner (Coyle, 2006). Finally, Thematic analysis was explored, which appeared to be a better fit. It is a method for identifying patterns or themes, selecting which are of interest, analysing and then describing them to the reader. This approach implies that a real and knowable world lies behind the individual and socially positioned information which a researcher can access. The experiences and perspectives could be speculated in a straightforward manner (Potter & Wetherell, 1987; Widdcombe & Wooffitt, 1995).

The flexibility of thematic analysis helped in identifying the themes centred in researcher's area of focus. Hence coding was carried out for the specific research questions. The researcher was interested in identifying how the acceptability of Infant CARE-Index displayed across the data and focussed on that particular feature in coding the data. The patterned responses that the researcher concentrated on was not dependent on the frequency of occurrence of the identified themes, rather than the quality of identified themes. Researcher's judgement played a crucial role in determining what counted as a theme specifically in relation to answering the research questions posed in the study.

Furthermore, the themes were identified at semantic or explicit level, in which the researcher focused explicitly on the surface meaning of what the participants said. This involved rich description of one particular theme, or group of themes, within the data rather than the entire data. This was followed by interpretation in relation to previous literature. (Braun & Clarke, 2006).

3.3.1 Phases or steps of thematic analysis

The use of multiple coders is recommended, however, due to time and resource constraint, the analysis of data was carried out by the researcher only. Qualitative analysis comprises of a recursive process, rather than a simple linear process. Analysis was guided by the steps outlined by Braun and Clarke (2006). This allowed access to the 'thick description' of the data set and uncovered insights into the perspectives of the participants. (Braun & Clarke 2006).

Familiarising with the data: The researcher transcribed the data into written form. The transcription was then entered into NVivo, a qualitative data analysis software. The researcher constantly read and re-read the data for analysis. The extracts were written down by the researcher from the very beginning of the analysis and continued up to the end.

Generating initial codes: This was followed by generating initial list of ideas demonstrating what the data was about. This led to formulation of initial codes for the data, which depicted the basic element of the raw data, which was then organised and interpreted in a meaningful way to answer the research questions (Tuckett, 2005). The coding structure was developed to address themes of acceptability, utility and barriers/facilitators of Infant CARE-Index. Further themes were added to take into account of emerging topics raised by the research participants to maximise the researcher's understanding of their perspective.

Searching for themes: The next phase involved searching for themes, when all the data had been coded (Braun & Clarke, 2006). Different codes were classified within potential themes. This led to development of a relationship between codes, between themes and between different levels of themes (overarching themes/ sub-themes).

Reviewing the themes: Next, the themes were reviewed and refined. Initially reviewing was done at level of coded extracts. The researcher investigated whether the themes formed a coherent pattern or not. This led to the formulation of a refined thematic map. This was followed by reviewing the entire data set. This process was concerned with whether the refined thematic map accurately reflected the meanings evident in the data set as a whole.

Defining and naming the themes: Once a satisfactory thematic map had been developed, the themes were named, followed by reporting the findings.

3.4 Findings

In general, the health professionals and the infant CARE-Index trainees recognised the Infant CARE-Index as a measure that would provide useful information on the quality of parent-infant interactions during the first year of an infant's life. Following the thematic analysis on the focus groups data; the findings were presented under various themes:

3.5.1 Need of an assessment measure

Health visitors and Infant CARE-Index trainees both were of the view that by using Infant CARE-Index, they would be able to provide solid evidence-based assessment for further interventions in families at risk.

"I have a case who has just gone under child protection and the main concern by the social worker is that the parent is not interacting with the child, so this is just us saying, there is no, urmm there is nothing we have done in terms of tool or concrete evidence for those interaction, so I guess in that case or scenario , child protection or assessment.. we might be able to say like, 'look at baby, we have evidence! Yes'" (Health professional 1).

"We work with mothers and babies, so if there is any PIIOS identified act or risk or history; mother would be referred to do a CARE-Index. So as to develop more sensitive relationship between mothers and babies." (Infant CARE-Index trainee 3)

"So, we want to look at whether parenting intervention has any bearing on the CARE-Index (sensitivity) score over a period of time." (Infant CARE-Index trainee 5)

Upon investigating, whether there actually was a need of a measure for assessing parent-infant interactions, the health professionals expressed that there was a definite need of such a measure. They described the need as:

"Massive!" (Health professional 2)

"Imperative!" (Health professional 5)

The need of a proper assessment measure was also highlighted, when, some health visitors expressed their concerns, that often, a child was taken up by the social services and separated from his/her parent solely on subjective assessments.

"But, actually in terms of leaving the parent from their young children. Different practitioners might not measure the interaction. It can't be subjective, can it be?" (Health professional 5)

Whereas, a few of the health professionals were of the view that although they were already identifying the parent-infant dyads at risk without using a specific measure and referring them for further interventions. But if a measure was introduced, it would provide a strong evidence base.

"We are doing that anyway aren't we. We are assessing. We are just not saying that "we are ticking the box". We are assessing them anyway and documenting. Would this necessarily be introduced as "oh we are going to assess you", who (mother) knows we are coming anyway. But we add a little bit more as you said for safe guarding purposes and things, a little bit more

of a concrete tool that we could say that actually we are going to assess you using this tool, then it's probably better in the long term."(Health professional 9).

3.5.2 Current practices

When the participants in the focus group with Infant CARE-Index trainees were asked, if they were aware of or using any measure in assessing parent-infant interactions, only one of the participants was trained and using Maternal post-natal attachment scale (MPAS) and Parent-infant interaction observational scale (PIIOS).

"I am using it clinically for mothers with mental health problems and infants for planning interventions. I am simultaneously training in PIIOS assessment, which is the brought down version of CARE-Index really, so we plan on using that (CARE-Index). Everybody in the team is trained in the PIIOS and myself and then the CARE-Index." (Infant CARE-Index trainee 3)

"Sorry, I forgot to mention about MPAS as well, it is an assessment for measuring mother baby interactions, we use that as well. Had some previous training for it." (Infant CARE-Index trainee 3)

While, the rest of the participants in the focus group with Infant CARE-Index trainees were neither aware nor using any other measure in assessing parent-infant interactions.

Similarly, the participants of focus group with health professionals were neither aware nor using any specific measure for assessing parent-infant interactions. However, some health professionals did mention the use of ASQ:SE-2, although it did not strictly measure parent-infant interactions but rather assessed the social emotional wellbeing of the child.

They based their judgement on subjective assessment relating to eye contact, physical touch and verbal communication between the parent and the infant.

"It's more like when we are looking out for things that are wrong. But you normally, you know yeah. For example, finding something, like the parent is not making eye contact, or the mom did not interact at all with the baby or respond to the baby, for example the baby cries and the mom does not pick up the baby." (Health professional 3)

"We use it as a part of global interaction…"(Health professional 4)

"It's not a standalone thing, other things are involved like substance abuse, domestic violence."(Health professional 6)

"You know everything is going well and you have the perfect play. But then, if you have spotted any concerns. Or it might be that later on something happens and you might have to call back in again. Like the involvement of social care services or might be when they go to the nursery, (the nursery staff) they might pick things up." (Health professional 6)

"It's usually a small piece of a bigger picture relating to the complexities going on, sometimes you really have to focus on that, the child who's forgotten…." (health professional 1)

"But what we are doing is considering is the voice of the child. So, in our assessments, we are looking at family dynamics and other complexities that are present and mental health." (Health professional 7)

From the discussion, it was observed that there was no strict criteria or protocol being followed by the health professional in carrying out their assessments.

3.5.3 Need of structural changes

The health professionals were of the view, that just introducing a new measure was not enough, structural changes needed to place in order to see the overall outcome in terms of improved quality of parent-infant interactions. This involved more contact with the families, more funding from the higher authorities to achieve better outcomes

"I think it needs some structural changes. We need more money from the government to fund more of us, so we could do more of these things . I don't think its case of looking at us or this organisation, you are looking at structural causes within the society and what the government behind the scene is important and I am thinking to feel that children' health especially young children's health is very low in the agenda now." (Health professional 3)

3.5.4 Relationship with the client family

Recent changes in the home visits (by health visitors) introduced by the health system, had led to reduced home visits by the health professionals and each time there was a different health professional having a varying subjective assessment, hence relationship with a family was hardly developed.

"I suppose what you could argue though is, that it's different practitioners seeing different , some of it is going to be evidence based and some of it is going to be subjective as well. Whereas when it is the same person that is seeing at pre one, that she has seen historically, that would make a difference." (Health professional 8)

Furthermore, reduced amount of contact between the health visitors and the families was one of the main reasons why many potentially 'at risk' dyads were being missed.

"It did used to work better, when we used to see the clients regularly, because, you know, if you had seen them antenatally, we used to do a lot of visiting in days gone by. You could pick and talk, and it would be quicker, then what it would be now." (Health professional 9)

"*Yeah, you could supervise for weeks and compare the interactions and calculate the next." (Health professional 12)*

The health visitors explained that under previous regime they would know the families antenatally and would have developed a relationship of trust and those mothers who had difficulty in developing healthy relationship with their infant or had any other issues would express their concerns with the health visitors, but with reduced home visits under the revised regimes had made it difficult to identify such dyads who were in the need of extra support.

"It's just about meeting and making a relationship, that's it, relationship building" (Health professional 10)

"Once you have got a relationship with a family that child was made to perform or not perform. Anyway, it's the whole relationship" (Health professional 8)

"Varies from person to person, some are quite honest about it like not feeling the bonding or having some form of depression." (Health professional 5)

3.5.5 Anticipated apprehension and ease of giving instructions for Infant CARE-Index video recordings

This theme describes the experience of video recording the parent-infant interactions. The participants in the Infant CARE-Index trainees focus groups had anticipated that they would find it difficult to obtain consent and carry out the video recordings of the parent-infant dyads. They thought the reason behind it may be that 'at risk' dyads would already be seen by social services and any sort of assessment would put them under further scrutiny.

"I think when I use it clinically, I will get questioned a lot, because people are often under threat. They feel like the scrutiny and threatened about what things mean, you know, that kind of I think that'll be a bit more tricky to sell, you know, the CARE- Index when families have already got social service."(Infant CARE-Index trainee 3)

The participants assumed their clients would give different explanations to their interactions to offset the threat and avoid the video recording.

" 'He's fine. We are doing fine.' 'It's not him. It's my mental health.' 'He's developing ok', 'she's developing ok.' 'And so, what will it mean?' 'If they (social services) see something, what will

happen and what might happen might not be good'….. like that… just imagining" (Infant CARE-Index trainee 3)

"Because I thought people would have like, they would feel a bit more like, ' oh no, I really don't want you to video and then be watching it' and then you know…." (Infant CARE-Index trainee 2)

Keeping in mind these anticipated reasons, the participants had opted for normative dyads (to video record two dyads as a part of Infant CARE-Index training course requirement), which was the easier option and giving instructions to parents and video recording came out as a good experience for them.

"It was quite easy to do, probably because the people I chose were quite happy. They weren't too many questions; they were not objectively very uncomfortable about it. It's quite… quite easy, really." (Infant CARE-Index trainee 5)

"They've been quite open about it. And just, I think it now it's very, very common, like people because of phones and stuff. Everyone's taking videos, maybe so I think maybe it's just like, Yeah, sure. I've got 10 of those. You can take another one. Yeah, I don't know. So yeah." (Infant CARE-Index trainee 2)

Whereas, the participants of the focus group with health professionals were of the opinion that their clients (parent-infant dyads) would not give them the consent to do Infant CARE-Index video recordings (due to cultural reasons). However, one of the participants mentioned that she knew two mental health nurses who were using Infant CARE-Index with their clients and it was working well.

"Two of mental health nurses have been trained in it. And they use it. They show the parents the videos and they are like wow ! It works really well, so yeah.." (Health professional 10)

3.5.6 Coding methodology

Almost all the participants in the Infant CARE-Index trainee focus group admitted that Infant CARE-Index had a rather complex and difficult coding methodology.

"Quite difficult, I am starting to make more sense of the patterns. A lot of patterns have looked similar at times to me, but at day 8 (2nd last day of training) it's becoming a bit clearer." (Infant CARE-Index trainee 5)

"I find it very complex, coding !"(Infant CARE-Index trainee 3)

When asked how confident the Infant CARE-Index trainees felt, while carrying out the coding of video recordings of parent-infant interactions; majority of the participants

had varying confidence. The participants felt more confident while discussing the coding in the class under supervision of the instructor, but when they were coding on their own, they would often miss the patterns within the interactions.

"Not very, but yesterday and today before I feel a bit more confident. I think I would say, I am getting a bit more confident. Really because I think I spotted some patterns right. But last week I just got fed up, when I was getting it all wrong last week. So, I think I am getting a bit more confident, but not very…."(Infant CARE-Index trainee 5)

"I would say I am not confident at all and like only a minority of us know like what is happening in interaction, that's held my confidence in terms of seeing things that are helpful, but then I go mad when I score them and she's (Instructor) right sadly… But then I go that was difficult, rather than see the overall picture, so again maybe….."(Infant CARE-Index trainee 3)

"I am confident, when I get it right, when I know that I have got that. O yeah, great, got it! And when I get it wrong, then it's like I am not confident, it's just like that.. " (Infant CARE-Index trainee 1)

The participants expressed the need of continued supervision in terms of gaining confidence and achieving reliability in coding the Infant CARE-Index video recordings.

"I feel like I need a lot of support for coding" (Infant CARE-Index trainee 2)

"I need ongoing supervision for every case I use the CARE-Index for the next ten years!" (Infant CARE-Index trainee 3)

"So, I think it's like kind of you need that ongoing support and challenging me, 'Oh you think it's that, what about that?'" (Infant CARE-Index trainee 4)

3.5.7 Time constraints

This theme presented the views of participants of Infant CARE-Index trainees in relation to the time needed to develop an understanding of the Infant CARE-Index methodology. The participants were of the view that they needed more time to reflect on what was taught in the training course.

"I feel I need a lot more time to reflect on things. The training has been quite quick for me with my work.. Having the time to read through my notes or reflect on them." (Infant CARE-Index trainee 4)

"Having that time to focus individually, but very hard at the same time. Video after video after… You know you need that time to reflect and then you know, give your information." (Infant CARE-Index trainee 2)

The participants were of the view that the time period between the sessions should have been more, so they could attempt coding the videos by reflecting on them and also gain appropriate time to go through feedback.

"Apart from that the sessions are just a bit too close together and that has made it harder because I would have liked longer stand on each video there is like suggested that half an hour or something to spend for each video for the homework but for me that's not long enough. That's a small… I need an hour. Yeah, to get my head around it…"(Infant CARE-Index trainee 5)

The participants considered the course to be quite intensive and they were not quite prepared for continuously watching the video clips of the parent-infant interactions and having additional video clips to code when they went home after a six-hour session.

"Expectation in terms of course evaluation on the CARE-Index and more time between the sessions. And maybe being prepared that you'll need a day between each one to do your homework sort of thing would have really helped ease the stress on me." (Infant CARE-Index trainee 3)

"A bit more of timely spread out. When you are sitting in a room all day and you are looking at all these videos and you get back three or another two. You know you are not really going to give; you know something you have got, as it's been a long day. Yeah and you do need that time to really, to come away from it, but also like time to be able to do that as well." (Infant CARE-Index trainee 1)

"There was obviously a lot of homework to do." (Infant CARE-Index trainee 3)

One of the participants expressed her concern that she nor her supervisor were aware of the amount of work required during the training and she could not balance the course work and her job at the same time.

"And also, yeah, I don't know if this is just me or I didn't read everything properly before the course. But my supervisor was quite surprised at how much work there was in between each week, because she hadn't accounted for any of that and neither had I and so I think it maybe, it was somewhere, I missed it written somewhere, but a bit more of a heads up on that just in terms of preparing myself for the time I need to actually really invest in and do it properly, I think" (Infant CARE-Index trainee 2)

3.5.8 Provision of training

The health professionals were eager to attend training course for Infant CARE-Index to improve their skills and gain the confidence to perform their assessments, however,

they were not very optimistic about obtaining any substantial amount of funding from the higher authorities.

"It would be nice to having the training and the confidence to go and do the assessment."(Health professional 6)

"At the minute, the commissioning budget would absolutely not support it. You definitely have to have backing from the commissioners that they want us to do it, to pay for it, fund it and pay for training. Within our current budget, it will be absolutely no!"(Health professional 4)

"So, from our perspective, we are welcome to anything, but it's about constraints. So, it's about, really you have to get the commissioners involved, public health to make that decision. Is that what they want from our service?"(Health professional 2)

3.5.9 Wider implications of assessments on parent-infant interactions

The health professionals expressed their concern, that only assessing and identifying parent-infant dyads 'at risk' was not enough. They stressed on the need of providing appropriate interventions, once the 'at risk' dyads are identified.

"Then you see all the other services like sure start, they have all been constricted, so if you identify these issues, you also have a duty of care to be able to be able do something about them and the services that are there are not robust enough to deal with them because of long waiting lists. Because of this you have to be really careful on using a tool. All the other things that go along with it." (Health professional 4)

When asked by the researcher, that what could make a difference in improving the assessments of parent-infant interactions? The health professional identified the bigger picture in terms of various measures that needed to be adapted to achieve sustained results within the resources available.

"More contact with the health visitors" (Health professional 10)

"Investment" (Health professional 11)

"Improving the services" (Health professional 13)

"Not just health stuff, social stuff" (Health professional 7)

"It's like not assessing and assessing and having no service available."(Health professional 6)

3.6 Discussion

This study was designed to provide an insight into whether Infant CARE-Index is acceptable in assessing parent-infant interactions within a UK context. To meet this objective, it was hypothesised that barriers and facilitators in the use of Infant CARE-Index need to be identified and addressed, which would eventually help in addressing the gaps in evidence (regarding the use of an appropriate measure for assessing parent-infant interactions in children under the age of 15 months). In this study, we explored the perspectives of different practitioners, regarding various existing methodologies for assessing parent-infant interaction and the use of Infant CARE index in their settings. Additionally, the factors that could facilitate a multidisciplinary, multi-sectoral large scale nationwide roll out of Infant CARE-Index were also explored.

It was observed that generally the health providing organisations delivering antenatal and postnatal care were following a non-standardised approach in carrying out parent-infant interactional assessments. Due to the reduced number of visits to the families since the revised health policies, the health professionals were finding it difficult to develop a relationship with the families and carry out appropriate assessments. Moreover, most of the prevailing assessments were subjective and not in accordance with the latest guidelines issued by NICE (NICE, 2015). The health professionals advocated the need of appropriate health policies with financing as the top-most criteria to facilitate effective assessments and subsequent interventions.

Relationship building with the families was considered an important factor in determining the quality of parent-infant interactions. In a broader perspective, the reason for ineffective assessments and interventions could also be related to the characteristics of the communities such as ethnic minorities, and in families with drug abuse and domestic violence where often the pre-requisites of health (food, shelter, sustainable income, safety) were not met. Women often fear losing the custody of their children and hence put up false positive behaviour with their children, so as to satisfy the health professionals. However, such cases usually come back in the later years with considerable mental health issues in the children. Therefore, there was a need to develop trustworthy relationships with such ‘at risk’ families and simultaneously address the determinants of socioemotional wellbeing rather than just carrying out assessments with a specific measure.

The views expressed by the trainees of Infant CARE-Index training course were mainly in relation to the limited time between the three sessions and the amount of course work during these sessions alongside the need of continued supervision.

3.6.1 Trustworthiness and rigor of the study

In this study, trustworthiness and rigor were maintained through explicit measures to ensure credibility, transferability and through a reflexive approach to the inquiry and analysis. To confirm credibility, during the data collection, the researcher ensured that the participants were allowed to express themselves openly during the discussions. During and at the end of each focus group discussion, the researcher revisited the main points and any additional points which were missed; that came out of the discussions to confirm agreements, disagreements, additions, and corrections from the participants. To ensure the credibility of the study findings, the researcher provided a comprehensive description of the study methodology and the purpose of such focus groups to the participants.

The Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups COREQ (Tong, 2007) and 'Standards for reporting Qualitative Research' (O' Brien, 2014) was used extensively by the researcher in reporting important aspects of the study methods, context of the study, findings, analysis and interpretation for this study (See Appendix 13).

3.6.2 Validating the data and results

The aim was to check the quality and accuracy of the data, the results and the interpretation from both the data sets. In practice, this could be carried out through various pathways. Member-checking was one approach, in which the researcher could take the summaries of findings to the participants and they would be asked to comment on the findings and whether the findings were an accurate reflection of their experiences. However, due to time constraints the approach adapted by the researcher involved, requesting her two supervisors to review the database and the qualitative results using their own criteria. Lastly, the researcher reported the disconfirming evidence to add transparency to data analysis, with the rationale that in real life, evidence does diverge from positive results.

3.6.3 Strengths and limitations

In this study, focus groups were used as a method of inquiry, which allowed the researcher to get simultaneous responses from multiple participants. Furthermore, it allowed participants to validate each other's responses. A strength of this study was

that participants were recruited from different settings such as psychologists, health visitors, researchers. This helped in obtaining perspectives from different angles and add breadth to the study.

It has been argued that quantitative research is weak in understanding the context in which participants express their output. The researcher is unable to personally reach out directly to the voices of participants in quantitative research. This qualitative study was designed to bring forward the reflexivity, personal biases and interpretations of the researcher, which are seldom discussed in standalone quantitative research alongside gaining details and in-depth output by the participants.

One of the limitations of this study was the limited time allocated for moderating the focus groups (approximately 20 mins), the reason being the busy schedule of the health visitors and the trainees of Infant CARE-Index course. Another limitation was the time constraint due to which the participant checking of the final data could not be carried out.

3.6.4 Conclusions

Despite the presence of various measures for the assessments of parent-infant interactions, the prevalence of attachment issues amongst parent and their children (above one year of age) remains high in the UK. This can be certainly prevented, if standardised observational measures are carried out with at risk families during the first year of an infant's life. Therefore, the need for a policy to ensure coordinated efforts for the prevention and management of such issues from the very beginning is essential for social emotional wellbeing of children in the later years.

The Infant CARE-Index training course needs to be to spread out across a longer time period with enough time for the trainees to understand and reflect upon the taught material. Similarly, adequate time needs to be allocated for the course work. The structure and expectations from the course need to be conveyed to the trainees in a schedule prior to the starting of the course, so the trainees can be mentally prepared for the course and the associated course workload.

3.6.5 Ethical concerns

This study received ethical approval from the Research Governance Committee, Department of Health Sciences, University of York. All participants received an information sheet briefly outlining the purpose of study; their voluntary participation and their data protection. Informed consent was obtained from all participants. Right

to refuse to answer to any question during focus groups was offered to the participants. Details about data anonymisation of the audio recorded transcripts was also explained to the participants. The confidential data will be kept for 5 years and only the researcher and supervisors will have access to the data on University of York Health Sciences secure network. The identification of the participants will not be revealed in any publication or dissemination.

Chapter 4: Overall Discussion

This chapter will identify how the initial aims and objectives have been met and seek to ground these findings within a broader research context. The strengths and limitations of the findings of the review and the qualitative analysis will also be discussed.

Combined, the research question, aims and objectives of the review and the qualitative analysis sought to determine a means of assessing the parent-infant interactions during the first year of an infant's life in the context of an observational measure, the Infant CARE-Index. This has been achieved through the following;

A detailed systematic review assessing the measurement properties of Infant CARE-Index and subsequently a qualitative analysis focusing on the real-life perspectives of practitioners on the acceptability of this measure.

The results, as summarised, will be interpreted and applied to the context and needs of routine home visits by health professionals to families postnatally during the first year and also to community parent-infant programmes. Such programmes, aid in fostering infants' social and emotional development alongside analysing the parent-infant interactions as both an indicator and influencer of this crucial aspect of childhood well-being. In targeting these early interactions, these home visits and programmes require a measure which is easy to administer and cost-effective, whilst also providing an accurate and reliable measure of a complex construct.

Initial exploration of the theoretical underpinnings of the parent-infant interactions and methods used to assess them lay the foundations for the systematic review. Initial scoping of a breadth of literature on attachment theory was performed prior to planning the research methodology of the review. This was central to gaining an informed overview of research context, and identifying where specific research was required, and would help in meeting the aims of this research.

4.1 Summary of results

Varied search techniques, using different combinations of broad and more focused terminologies led to a total of forty-one papers being identified. These detailed the

use of Infant CARE-Index in children aged from birth up to 15 months. A consensus - based standards for the selection of health measurement instruments (COSMIN) methodology for systematic reviews of patient-reported outcome measures (Mokkink et al, 2018) was adapted to the context of this research and incorporated further considerations to enable assessment of measurement properties of an observational measure-Infant CARE-Index. Although it was widely stated that Infant CARE-Index had been well validated across many studies, however across the board, evidence was clearly lacking for validity and reliability, alongside significant doubt on the methodological rigour of its evidence base.

Similarly, acceptability was a significant concern, with difficult training and associated high cost, highlighting a major obstacle to its use in the community settings.

4.1.1 Applying the findings to a broader context

Of particular significance, researchers involved in Enhancing Social-Emotional Health and wellbeing in the Early years (E-SEE) trial (Bywater et al, 2018) have recently used the Infant CARE-Index at baseline, 2 months follow up and 9 months follow up home visits to parent-infant dyads. This measure has been used in acknowledgment that issues surrounding parent-infant interactions can have major detrimental impacts on both mother and infant.

Designed for use during the critical first 15 months of an infant's life, it has the potential to be used to assess parent-infant interactions in resource-limited and time-restricted community settings, once a few more robust validity studies are published.

Although, as pointed out in the qualitative analysis, Infant CARE-Index does have an intensive and expensive training, but once reliability is achieved, this measure is reasonably easy to use and only requires three minutes to video record and another 20-25 minutes to code. This is central, especially for the community-based researchers and healthcare practitioners which this research intends to inform.

Unlike some of the other measures, the Infant CARE-Index is firmly grounded in the pivotal attachment theory. The maternal behaviour as an indicator of 'attachment' has been the longstanding focus of existing research. This term 'attachment' in the first year of infant's life is used interchangeably as interactions, relationship and also sometimes as bonding in both the literature and by the wider public. In developing the Infant CARE-Index, Crittenden shifted the focus towards dyadic synchrony rather than focusing on the two participants as separate individuals. These emotive responses

within a dyadic context offer a crucial and subjective insight into the interactions which is so integral to early child development. This brought with it the need for a comprehensive, observational instrument as opposed to previously lengthy observational methods like the 'Strange situation procedure' making their practicality within the community setting as impossible. The parent and infant responses are divided into affective (affect and arousal, turn taking, control and activity) and cognitive (facial expression, vocal expression and position) components. The theoretical foundations of this measure therefore acknowledge the dynamicity of the relationship, whilst making it clear as to which element it focuses on.

Yet, equally, the Infant CARE-Index does not ignore the behaviours which have for decades been the basis of attachment research, rather it seeks to measure and quantify the emotions and rationale *behind* such behaviours. The end score lead to an overall assessment in terms of dyadic synchrony patterns which are most commonly sensitive mother and cooperative infant, unresponsive mother and passive infant or controlling mother and difficult/compulsive infant. Although different variants of these main patterns are often encountered. Infant CARE-Index therefore sets out to measure the interrelatedness of the emotional indicators which drive both the parental and infant interactions in a dyadic context.

4.1.2 Parent-infant interactions as a measurement focus

Measurement properties aside, placing parent-infant interactions on the qualitative analysis as the measurement focus, amongst the wide range of other interrelated concepts may guide parent-infant support programmes and health practitioners in understanding what exactly the Infant CARE-Index is measuring. It is difficult to keep the various elements conceptually distinct, and the boundaries of affective and cognitive components closely overlap. Yet, abstract as it may sound, in highlighting the strengths of the Infant CARE-Index, the qualitative element has brought forward the need of distinctly recognising the infants' social, emotional and neurocognitive wellbeing:

"*sometimes you really have to focus on that, the child who's forgotten…." (health professional 1)*

This came along with the accompanying parental mental health and well-being, which plays an important role in providing sensitive parenting to their infants:

"So, in our assessments, we are looking at family dynamics and other complexities that are present and mental health." (Health professional 7)

Thus, with ever growing behavioural and mental health problems in the children of school going age; careful assessment of the parent-infant interactions during the early part of a child's life has become even more relevant to programmes aiming to strengthen and support the parent-infant relationship. (Bywater et al, 2018)

4.2 Quality of evidence

In focusing purely on the coding methodology of Infant CARE-Index, the issue of reported reliability levels of the coders and the methodology used to report the reliability between coders was inherent across the studies, thus leading to downgrading of many studies.

Some studies included within the review reported how the mothers put up their best behaviours, which was confirmed by the qualitative analysis also, where the participants had anticipated such behaviours on the part of mothers due to threat of having to give up their children. Having said that, Infant CARE-Index has the unique ability to discriminate such behaviours, however, it is also very much dependent on the coder's reliability level, which was seldom mentioned clearly in the studies included in the review.

4.3 Strengths & limitations

This review was the first to assess the measurement properties of Infant CARE-Index across the literature to date (March 2020) and highlight the areas where evidence was lacking. Furthermore, this was one of the first few reviews which have used the updated COSMIN checklist allowing for thorough assessment of methodological quality and rigour (Mokkink et al, 2018).

It is also important to note that the vast majority of studies included in the review, comprised of the representative study populations associated with different risk factors such as single women, adolescent mothers, women with different levels of education, drug abuse, premature births, maternal depression etc. Hence, when the study samples were combined, the overall population represented a diverse range of dyads and the results of these studies could be generalised to the diverse range of

mothers who access community parenting programmes and with whom the health visitors come across during their routine home visits.

Cochrane guidance states the need for at least two researchers to conduct and agree upon assessments of risk of bias (Higgins and Green, 2011). This was not feasible due to time and resource constraints. For this reason, every effort was made to ensure that all aspects were as transparent as possible, though this is acknowledged as a weakness and a source of potential bias.

While, evidence which supported the use and need of this measure came into light through the results of qualitative analysis, it also drew attention to potential biases and issues relating to usability, allowing researchers and healthcare practitioners to make an informed and evidence-based decision as to the suitability of Infant CARE-Index in research and routine practice alongside highlighting a dearth of robust observational measures focusing on this age range.

4.4 Recommendations for future research

With almost no studies exclusively validating its use and its acceptability within the UK settings; Infant CARE-Index require a much greater evidence- basis and needs to be made more readily accessible before it can be used on a large scale in clinical and research context. Whilst the Infant CARE-Index is well established and used across different cultures and populations across the world and well suited to use in the community setting, newer downgraded versions of Infant CARE-Index such as the Parent-infant observation scale (PIIOS) is offering a potential for fast and effective assessment of the maternal-infant interactions consisting of fewer items, less intensive and less expensive training and online reliability tests (Svanberg et al, 2013). Hence, with an outburst of new measures in the field of assessing the parent-infant interactions, it is essential to regularly update the existing knowledge and update it according to the latest guidelines with good quality longitudinal studies with a rationale to gain maximum benefit from decades of research and efforts that have been incorporated into the present form of Infant CARE-Index.

As a more general recommendation, a validation study using the COSMIN Study Design checklist for PROMS (Mokkink et al, 2019) would add the much needed evidence on the content validity, hypothesis testing, responsiveness and reliability

(Inter-rater, test-retest, and intra-rater reliability) of Infant CARE-Index with clear mentioning of the number of coders involved and their reliability levels,

Furthermore, whilst the review revealed majority of the studies on mother-infant interactions, as does the large majority of research in the field, further studies should seek to assess interactions in more diverse parent or carer-infant interactions.

A general trend observed in the process of literature review for the systematic review revealed that most authors used the 'maternal sensitivity' scale of Infant CARE-Index. Even though other components (maternal unresponsiveness, maternal control, infant difficulty, infant passiveness, infant compliance and infant cooperation), all offer insight into more distinct aspects of interactional thoughts, feelings and behaviours. Hence, there is a need of studies utilising the Infant CARE-Index across all the domains rather than just focusing on the maternal sensitivity.

The qualitative analysis on the acceptability of Infant CARE-Index may benefit from a wider population involved in using this measure. Hence, conducting focus groups with the trained Infant CARE-Index coders and social workers will add further evidence on the use and acceptability of this measure across the UK.

This research added to existing literature in further defining the parent-infant relationship, highlighting maternal-infant interaction as a crucial component which warrants focus in the search for an accurate and informative measure of the interactions. This study offers a potential means of monitoring and improving the trajectory of childhood social and emotional development with the multiple uses of Infant CARE-Index such as screening, assessing and planning intervention with 'at risk' dyads. In addition, it is recommended that further research is needed to evaluate current practices and interventions systematically, to see the extent they have on preventing the emergence of attachment difficulties in children.

Appendix 1

Definitions used for measurement properties

(R. Goodman, 1997; Mokkink et al., 2010; Streiner & Norman, 2008)

Content validity: The degree to which the content of an instrument is an adequate reflection of the construct to be measured.
Construct validity: *Discriminative validity:* Ability of a tool to discriminate between two extreme groups. *Convergent validity*: The degree to which the scores of the (new) scale relate to scores on other measures to which it should be related. *Discriminant/divergent validity*: The degree to which the scores of the (new) scale do not relate to scores on another scale that measures dissimilar constructs. *Structural validity*: The degree to which the scores of an instrument are an adequate reflection of the dimensionality of the construct to be measured. *Cross-cultural validity*: The degree to which the performance of the items on a translated or culturally adapted instrument are an adequate reflection of the performance of the items of the original version of the instrument.
Criterion validity: *Concurrent validity*: The correlation of the instrument with a ‘‘gold standard’’ criterion administered at the same time. *Predictive validity*: The correlation of the instrument with a ‘‘gold standard’’ criterion that will be available in the future.
Internal consistency: The degree of the interrelatedness among the items.
Reliability: *Intra-rater reliability*: The extent to which scores for people who have not changed are the same for repeated measurement by the same rater. *Inter-rater reliability*: The extent to which scores for people who have not changed are the same for repeated measurement by different raters (of the same type) on the same occasion. *Cross-informant consistency*: The extent to which scores for people who have not changed are the same for repeated measurement by different types of raters on the same occasion. *Test-retest reliability*: The extent to which scores on the same version of questionnaire for people who have not changed are the same for repeated measurement over time.
Measurement error: The systematic and random error of a person’s score that is not attributed to true changes in the construct to be measured.
Responsiveness: The ability of an instrument to detect change over time in the construct to be measured

Appendix 2

Example search strategy

Using the following seven databases:

Econlit, Embase, Maternity & Infant Care Database (MIDIRS), OVID MEDLINE®, APA PsycInfo, Social policy and practice & APA PsycArticles Full text.

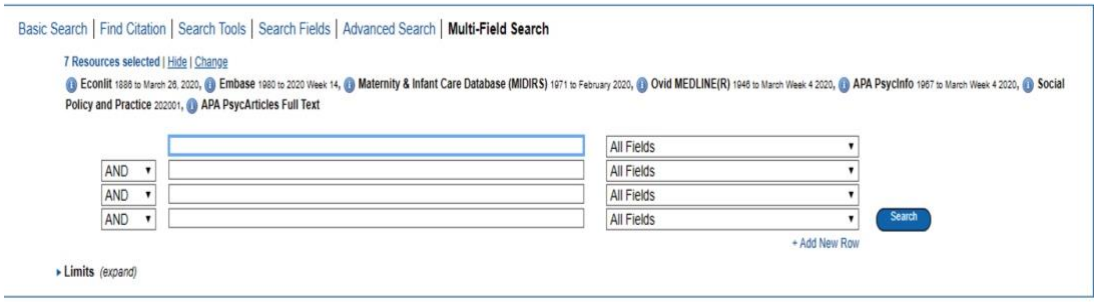
Using the following search terms:

1. CARE-Index

2. parent or mother or father or carer

3. infant or child or baby or toddler

4. interaction or relationship or attachment



Appendix 3

Study summaries (Characteristics of included study populations)

Sr no.	Author, Year, Country	Study type	No. of measurements for Infant CARE-Index	Construct measured	N Sample size (dyads) reported in analysis	Attrition	Main Carer	Mean age of carer (SD)	Mean child age (SD)	% female (Parent)	% female (Child)	Pre-dominant Ethnicity	Study/Recruitment settings	Location of observations	Recruitment Method
1)	Azar, 2006 Canada	Epidemiological	1	Maternal control	212	NK	Mother	16.9 (1.0)	4.38 months (0.4)	100%	57	White	longitudinal	Laboratory	Convenience
2)	Forcada 2010 Switzerland	cohort	**1**	Dyadic interaction	72	NK	Mother	NK	6 months	100%	55	White	longitudinal	NK	Convenience
3)	Pillhofer, 2014 Germany	Pilot study	**4** preintervention postintervention 6 months age 2 months age	Maternal sensitivity	83	0.01-0.07% n=14	Mother	22.5 (6.1)	birth-12 months	100%	51.80%	White	Community	Home	Convenience
4)	Tester-Jones, 2017 England	(Intervention study) mixed measures design	**2**	Maternal sensitivity	79	mentioned	Mother	33.2 (4.4)	7.2 months (2.9)	100%	NK	White	Community	Home /University	Random
5)	Svanberg, 2010, England	Developmental and intervention study	**2** Baseline 12 months follow up	Maternal sensitivity Dyadic interaction	192	mentioned	Mother	26.1 (5.7)	3months-12 months	100%	48%	NK	Community	Home	Convenience

Sr no.	Author, Year, Country	Study type	No. of measurements for Infant CARE-Index	Construct measured	N Sample size (dyads) reported in analysis	Attrition	Main Carer	Mean age of carer (SD)	Mean child age (SD)	% female (Parent)	% female (Child)	Pre-dominant Ethnicity	Study/Recruitment settings	Location of observations	Recruitment Method
6)	Soares, 2018, Portugal	Sub study of a larger quasi-experimental and longitudinal study	1	Maternal sensitivity	86	mentioned	Mother	30.63 (6.40)	12.4 months (6.19)	100%	53.40%	Portuguese	Primary health care center	Laboratory	Convenience
7)	Neri, 2017, Italy	exploratory study	1	Dyadic interaction	92	not mentioned	mother/ father	Mothers 37.1 (5.6) Father 39.3 (5.1)	3 months	58.60%	54.20%	Italian	Hospital NICU	laboratory	Convenience
8)	Kristensen, 2017, Denmark	Quasi-experimental study	1	Dyadic interaction, Maternal sensitivity, Infant cooperation	132	mentioned	mother	30.3	3-6months	100%	54%	German	Community	Home	Convenient
9)	Barlow, 2016, UK	Pilot RCT	1	Dyadic interaction excluding compulsivity in analysis	31	mentioned	mother/ father	32	NK	93.50%	38.70%	British	Hospital/NICU	Home	Convenience
10)	Steadman, 2007, UK	Pilot/ exploratory study	1	Dyadic interaction	18	mentioned	mother	34	5-18 weeks	100%	28%	British	Longitudinal study/National child-birth trust	Home	Convenience
11)	Letourneau, 2017, Canada	Secondary analysis from a subsample of a cohort	1	Maternal sensitivity, Maternal control Maternal unresponsiveness	242	mentioned	mother	31.2	6 months	100%	50%	Canadian	Longitudinal	Clinic	Convenience

Sr no.	Author, Year, Country	Study type	No. of measurements for Infant CARE-Index	Construct measured	N Sample size (dyads) reported in analysis	Attrition	Main Carer	Mean age of carer (SD)	Mean child age (SD)	% female (Parent)	% female (Child)	Pre-dominant Ethnicity	Study/Recruitment settings	Location of observations	Recruitment Method
12)	Feurtes, 2009, Portugal	Longitudinal study	1	Maternal sensitivity, Maternal control Maternal unresponsiveness	48	mentioned	mother	29.7	9 months	100%	40%	Caucasian	NICU	Laboratory	Convenience
13)	Kenny, 2013, UK	Case-control	1	Maternal sensitivity, Maternal unresponsiveness, Infant passivity, Infant cooperation	138	not mentioned	mother	30.6	15.1 weeks	100%	47%	Caucasian	Hospital, Community	Lab/ home	convenience
14)	Hofer, 2007, Germany	Exploratory	1	Maternal sensitivity, Maternal unresponsiveness, maternal cooperation	56	mentioned	mother	33	6 months	100%	59%	Caucasian	General population	Laboratory	Convenience
15)	Zwonitzer, 2015, Germany	pilot study/intervention study	4	Maternal sensitivity	53	mentioned	mother	28.5	6-12 months	100%	NK	NK (German?)	child welfare services	Home	Convenience
16)	Fuertes, 2009, Portugal	cross-sectional	1	Maternal sensitivity, Maternal control Maternal unresponsiveness, Infant cooperative, Infant compulsive, Infant difficult Infant passive	53	not mentioned	mother	17-35+ yrs	6.9 months	100%	46.30%	NK (Portuguese?)	Hospital/ Early interventions team	Home assumed	Convenience

Sr no.	Author, Year, Country	Study type	No. of measurements for Infant CARE-Index	Construct measured	N Sample size (dyads) reported in analysis	Attrition	Main Carer	Mean age of carer (SD)	Mean child age (SD)	% female (Parent)	% female (Child)	Pre-dominant Ethnicity	Study/Recruitment settings	Location of observations	Recruitment Method
17)	Sidor, 2011, Germany	Quasi-experimental	1	Maternal sensitivity, Maternal control, Maternal unresponsiveness	133	mentioned	mother	26.5	19 weeks	100%	45%	NK (German?)	Hospital/maternal health clinics, youth and social welfare offices	Home	Convenience
18)	Parfitt, 2013, UK	Prospective study/observational study	1	Maternal sensitivity, Maternal control Maternal unresponsiveness, Infant cooperative, Infant compulsive, Infant difficulty Infant passivity	84	mentioned	mother/father	Mother (33.12) Father (34.6)	3 months	97%	62%	Caucasian	Longitudinal /hospital/community antenatal clinics and antenatal classes	Home	Convenience
19)	Stephenson, 2018, UK	Observational study	2	Maternal sensitivity, Maternal unresponsiveness, Infant cooperative, Infant passiveness	62	mentioned	mother	30.9	99 days	100%	45%	White	Hospital/ NICU	Hospital	Convenience
20)	Underdown, 2013, UK	Mixed-methods study	2	Maternal sensitivity	33	mentioned	mother	16-41 yrs	10 weeks	100%	51%	White	Children centre	Home	Convenience
21)	Rigby, 2016, UK	exploratory study	1	Maternal sensitivity	40	mentioned	mother	31.5 (5.1)	21.2 weeks	100%	40%	White	Hospital/ NICU	Hospital	Random
22)	Hohenberger, 2012, UK,	Quasi-experimental study	2	Maternal sensitivity, Maternal control, Maternal	84 (t1) 58(t2)	mentioned	mother	Not mentioned	6 months (t1) 10 months (t2)	100%	46%(t1), 43%(t2)	European (English, French, German)	Longitudinal. Participants recruited from subject pools of 3 labs in	Laboratory	Convenience

Sr no.	Author, Year, Country	Study type	No. of measurements for Infant CARE-Index	Construct measured	N Sample size (dyads) reported in analysis	Attrition	Main Carer	Mean age of carer (SD)	Mean child age (SD)	% female (Parent)	% female (Child)	Pre-dominant Ethnicity	Study/Recruitment settings	Location of observations	Recruitment Method
	France, Germany			unresponsiveness									Munich, London and Paris		
23)	Jorgensen, 2011, Switzerland	Exploratory, Longitudinal	1	Dyadic interaction	32	mentioned	mother	33.4(4.4)	4 months	100%	51%	European Union, Swiss	Hospital	Laboratory	Convenience
24)	Ward, 1995, USA	Longitudinal study	2	Maternal sensitivity, Maternal control, Maternal unresponsiveness	84 (3 months) 81 (9 months)	mentioned	mother	16.5 yrs (0.99)	3 months 6 months	100%	NK	76% (African American)	NK	NK	convenience
25)	Fuertes, 2011, Portugal	Exploratory study	1	Maternal sensitivity, Maternal control, Maternal unresponsiveness	46	NK	mother	29.7 yrs (5.7)	3 months	100%	39.14%	Portuguese Caucasian	NICU	Laboratory	Convenience
26)	Flykt, 2010, Finland	Exploratory, Longitudinal	1	Maternal sensitivity, Maternal control, Maternal unresponsiveness, Infant cooperative, Infant compulsive, Infant difficult, Infant passive	49	mentioned	Mother	29 yrs (5)	14 months (7.76)	100%	48%	Finnish	Maternal health care centers	Home	Convenience
27)	Marina Fuertes, 2016, Portugal	Longitudinal study	2 (at 9 and 15 months)	Maternal & Paternal sensitivity, control, unresponsive	82 Triads	mentioned	mother/ father	Mother 29.84 yrs (5.19) Father	9 months. 15 months	50%	36.50%	Portuguese Caucasian	Longitudinal (2 yr)NICU	Lab in Hospital	Convenience

Sr no.	Author, Year, Country	Study type	No. of measurements for Infant CARE-Index	Construct measured	N Sample size (dyads) reported in analysis	Attrition	Main Carer	Mean age of carer (SD)	Mean child age (SD)	% female (Parent)	% female (Child)	Pre-dominant Ethnicity	Study/Recruitment settings	Location of observations	Recruitment Method
				ness, Infant cooperative, compulsive, Infant difficult, Infant passive				33.89 yrs (6.58)							
28)	Conroy, 2009, UK	Exploratory study	1	Maternal sensitivity	200	mentioned	mother	30.7 yrs (6.58)	8 weeks	100%	52%	British White	Hospital NICU	Home	consecutive
29)	Barlow, 2007, UK	Randomised controlled trial	1	Maternal sensitivity, infant cooperation	131	mentioned	Mother	NK	0-12 months	100%	NK	British White	Primary health care center	Home	Random
30)	Komoto, 2015, Japan	Quasi-experimental study	4	Maternal sensitivity, Maternal control, Maternal unresponsiveness, Infant cooperative, Infant compulsive, difficult, Infant passive, Dyadic Interactions	135	mentioned	Mother	33.86 yrs (Intervention) 32.76 (control)	1-12 months	100%	53%	Japanese	Primary health care center	clinic	Convenience
31)	Cameron, 2018, UK	Exploratory study	2	Dyadic interactions	66	mentioned	Mother	18-42 yrs	6.9 weeks	100%	NK	White (Northern Ireland)	child welfare services	Home	Convenience

Sr no.	Author, Year, Country	Study type	No. of measurements for Infant CARE-Index	Construct measured	N Sample size (dyads) reported in analysis	Attrition	Main Carer	Mean age of carer (SD)	Mean child age (SD)	% female (Parent)	% female (Child)	Pre-dominant Ethnicity	Study/Recruitment settings	Location of observations	Recruitment Method
32)	O'Shaughnessy, 2012, UK	Mixed-methods study, pilot study	2	Dyadic interactions	5	mentioned	Mother	17-32 yrs	6-12 months	100%	NK	West African	Community, maternal health centers, primary care, asylums screening centers, refugee pre/post-natal support group	Therapeutic group sessions	Convenience
33)	Karmiloff, 2010, UK, France, Germany	Longitudinal study	2	Dyadic interaction	173 (t1) 155 (t2)	mentioned	Mother	Not mentioned	(t1) 6 months (t2) 10 months	100%	(t1) 45% (t2) 44%	Caucasian	Longitudinal, subject pools of 3 labs in Munich, London and Paris	Similarly set up Labs at Munich, London & Paris	Convenience
34)	Neuhauser, 2018, Switzerland	Cross-sectional	1	Maternal sensitivity	211	mentioned	Mother	30yrs (5.77)	2.69 months (1.54)	100%	59%	25% Balkan. 14% Swiss. 11% Turkish. 9% Portugal. 9% Central & East Asia	Maternity wards, medical & social services, parenting counselling offices in suburbs of Zurich	Home	Convenience
35)	Sidor, 2015, Germany	Longitudinal controlled	2 (T1 &T3)	Maternal sensitivity, Maternal control, Maternal unresponsiveness, Infant cooperative, Infant compulsive, difficult, Infant passive	150 (Intervention) 152 (control)	mentioned	Mother	Intervention group: 24.5yrs (6.7) Control group: 28.2 (6.4)	Age at T1. Intervention group: 19.3 wks ((3.32) Control group: 19 wks (2.39)	100%	Intervention group: 48.3%. Control group: 48%	German	-Regular midwives. -maternity clinics -gynecologists and pediatricians -welfare offices -counselling centers	Home	Convenience

Sr no.	Author, Year, Country	Study type	No. of measurements for Infant CARE-Index	Construct measured	N Sample size (dyads) reported in analysis	Attrition	Main Carer	Mean age of carer (SD)	Mean child age (SD)	% female (Parent)	% female (Child)	Pre-dominant Ethnicity	Study/Recruitment settings	Location of observations	Recruitment Method
36)	Udry-Jørgensen, 2015, Switzerland	Exploratory	1	Maternal sensitivity, Paternal sensitivity	65	mentioned	Mother & Father	Mother: 32.3(4.2) Father: 34.7(5.7)	98.7 days(9.5) 3.3months	100%	47.7%	Swiss	-Maternity ward of Hospital. -General Register office	Laboratory	Convenience
37)	Borghini, 2014, Switzerland	Interventional	2	Maternal sensitivity, Maternal control, Maternal unresponsiveness, Infant cooperative, Infant compulsive, difficult, Infant passive	**78** Preterm with interventions: 26. Preterm without interventions: 29 Full-term without interventions: 23	mentioned	Mother	Pre-term infants with interventions: 33(3) Pre-term infants without interventions: 33(5). Full-term without interventions: 31(5)	4 months	100%	Pre-term infants with interventions: 50% Pre-term infants without interventions 52%. Full-term without interventions 48%	Swiss/EU	NICU, Maternity ward	Not mentioned	Convenience
38)	Binda, 2019, Chile	Observational study	1	Maternal sensitivity, Maternal control, Maternal unresponsiveness, Infant cooperative, Infant compulsive, difficult, Infant passive	177	Mentioned	Mother	25.3(7.09)	7.37months (2.12)	100%	49.2%	Chilean	Two Primary health centers	Corresponding Primary health centers	Convenience

Sr no.	Author, Year, Country	Study type	No. of measurements for Infant CARE-Index	Construct measured	N Sample size (dyads) reported in analysis	Attrition	Main Carer	Mean age of carer (SD)	Mean child age (SD)	% female (Parent)	% female (Child)	Pre-dominant Ethnicity	Study/Recruitment settings	Location of observations	Recruitment Method
39)	Crugnola, 2016, Italy	Pilot study	3 T1(3 months infant age) T2 (6 months infant age) T3 (9 months infant age)	Maternal sensitivity, Maternal control, Maternal unresponsiveness, Infant cooperative, Infant compulsive, difficult, Infant passive	T1:48 32(Intervention) 16(control) T3:28. 18(Intervention). 10(control)	Mentioned		Intervention grp:18.75(1.43) Control grp:17.94(1.94)	T1:3months T2:6months T3:9months	100%	T1:Intervention grp (62%) Control grp (62%)	Italian	-Maternity Ward. -A Family counselling services	Out-patient unit of Hospital	Convenience
40)	Emery, 2008, Canada	Part of Longitudinal research project	1	Maternal sensitivity	134	mentioned	mother	16.89 (1.01)	4.4 months	100%	57.1%	Caucasian(North American)	-Specialized school for pregnant women -4 Group homes for young mothers	Laboratory	Convenience
41)	Paquette, 2004, Canada	Exploratory study	1	Maternal sensitivity Maternal control, Maternal unresponsiveness, Infant cooperative, Infant compulsive, difficult, Infant passive	96	None	mother	16.9(1.2)	4.4 months(0.5)	100%	Not mentioned	Caucasian(Canadian)	-4 Rehabilitation centre School for adolescent mothers to finish high school	University	Convenience

Appendix 4

Results of studies on Reliability evidence

Sr No.	Author, year, country	N	No. of coders	Reliability levels	Construct measured	Reliability evidence	COSMIN methodological quality	Results & Quality of results (Terwee Rating)
1)	Azar, 2006, Canada	212	2	Trained by Crittenden	Maternal control	Inter-rater reliability (Good-Excellent) Alpha coefficients 0.84-0.95	Inadequate	alpha coefficient 0.84-0.95 [?]
2)	Forcada, 2010, Switzerland	72	2	One coder Certified by Crittenden	Dyadic interaction	ICC 0.87 (Maternal sensitivity, control and unresponsiveness). ICC 0.86 (Infant cooperation, compulsive compliance, difficult and passive)	Doubtful	ICC= 0.87 (Maternal) ICC=0.86 (Infant) [+]
3)	Pillhofer, 2014, Germany	83	7	Trained coders	Maternal sensitivity	Inter-rater reliability assessed on 11 videos from the main sample. ICC 0.78 p<0.001(Maternal sensitivity)	Doubtful	ICC= 0.78 (Maternal) [+]
4)	Tester-Jones, 2017, UK	79	3	Trained to reliability	Maternal sensitivity	Inter rater reliability on 31 (15%) videos from the main sample assessed by third coder. Inter-rater reliability (Strong) k=0.88	Doubtful	k=0.88 [+]
5)	Svanberg, 2010, UK	192	2	Trained to reliability	Maternal sensitivity Dyadic interaction	Correlation at baseline r=0.82,p<0.001, n=30 by third coder (maternal sensitivity) Correlation at 12 month follow up r=0.87, n=92, p<0.001 (maternal sensitivity) Correlation for the intervention grp r=0.83, n=47, p<0.001 Correlation for the comparison group r=0.90, n= 45, p<0.001	Doubtful	r=0.82-0.90 [?]
6)	Soares, 2018, Portugal	86	3	trained to reliability	Maternal sensitivity	Reliability amongst two coders: ICC 0.94 (maternal sensitivity) ICC 0.89 (infant cooperation) Reliability amongst three coders ICC 0.91 (maternal sensitivity) ICC 0.86 (Infant cooperation)	Doubtful	ICC=0.91-0.94 (Maternal) ICC=0.86-0.89 (Infant) [+]

Sr No.	Author, year, country	N	No. of coders	Reliability levels	Construct measured	Reliability evidence	COSMIN methodological quality	Results & Quality of results (Terwee Rating)
7)	Neri, 2017, Italy	92	2	one coder trained to reliability No info for the second coder	Dyadic interaction	Inter-rater reliability amongst two coders = mean Cronbach's alpha=0.77; range 0.71-0.83	Inadequate	Cronbach α =0.77 [?]
8)	Kristensen, 2017, Denmark	132	3	Reliable	Dyadic Interaction, Maternal sensitivity, Infant cooperation	Inter-rater reliability amongst three coders on random 20% sample = Cronbach's alpha=0.88 (dyadic synchrony) Cronbach's alpha =0.70-0.88 (maternal sensitivity, control, unresponsiveness) Cronbach's alpha 0.82-0.92 (infant cooperation, compulsive, difficult and passive)	Doubtful	Cronbach α =0.88 [?]
9)	Barlow, 2016, UK	31	NK	NK	Dyadic interaction excluding compulsivity in analysis	Inter-rater reliability for 10% sample taken from both groups 0.77 (Maternal sensitivity) 0.81 (Infant cooperation)	Inadequate	ICC=0.77 (Maternal) ICC= 0.81 (Infant) [+]
10)	Steadman, 2007, UK	18	2	Reliable	Dyadic interaction	inter-rater reliability (ICC) for 22% random sample 0.81 Maternal sensitivity 0.87 maternal unresponsiveness 0.85 maternal control 0.57 infant cooperation 0.96 infant compulsive 0.99 infant difficult 0.98 infant passive	Doubtful	ICC=0.81 [+]
11)	Letourneau, 2017, Canada	242	NK	Reliable	Maternal sensitivity, Maternal control, Maternal unresponsivenes	Inter-rater reliability 94.4% (maternal sensitivity, control and un-responsiveness)	Inadequate	Inter rater reliability 94.4% [?]
12)	Feurtes, 2009, Portugal	48	2	Reliable	Maternal sensitivity, Maternal control, Maternal unresponsivenes s	Inter-rater reliability for 33.3% random sample Cohen's k Maternal M k 0.90 Infant M k 0.85 Overall Mk 0.87 Individual Means ks 0.95 maternal sensitivity ks 0.91 maternal control. ks 0.84 maternal unresponsiveness ks 0.93 infant cooperation ks 0.80 infant compliance ks 0.88 infant difficulty ks 0.80 infant passivity	Doubtful	k=0.87 [+]

Sr No.	Author, year, country	N	No. of coders	Reliability levels	Construct measured	Reliability evidence	COSMIN methodological quality	Results & Quality of results (Terwee Rating)
13)	Kenny, 2013, UK	138	4	Reliable	Maternal sensitivity, Maternal unresponsiveness, Infant passivity, Infant cooperation	Inter-rater reliability (ICC) for 15% sample = ICC 0.90 maternal sensitivity	Doubtful	ICC=0.90 (Maternal) [+]
14)	Hofer, 2007, Germany	56	2	Reliable	maternal sensitivity, Maternal unresponsiveness, Maternal cooperation	Inter-rater reliability (ICC) for 20% random sample. ICC r=0.89 Maternal sensitivity. ICC r=0.90 Maternal control ICC r=0.90 Maternal control. ICC r=0.94 Maternal unresponsiveness	Adequate	ICC r=0.89-0.94 (Maternal) [+]
15)	Zwonitzer, 2015, Germany	53	7	Reliable	Maternal sensitivity	Inter rater reliability on 21 (40%) videos of the main sample. Maternal sensitivity ICC r= 0.74-87	Doubtful	ICC r=0.74-0.87 (Maternal) [+]
16)	Fuertes, 2009, Portugal	53	3	1 reliable	Maternal sensitivity, Maternal control, Maternal unresponsiveness, Infant cooperative, Infant compulsive, Infant difficult, Infant passive	Inter rater reliability for major classification above 80%	Inadequate	Inter-rater reliability 80% among 3 coders [?]
17)	Sidor, 2011, Germany	133	2	1 (screening level reliable) 2 (provisional screening level reliable)	Maternal sensitivity, Maternal control, Maternal unresponsiveness	first coder reliability 0.65 second coder reliability 0.49 Mean reliability scores maternal sensitivity r=0.65 maternal control r=0.77 maternal unresponsiveness=0.84 infant cooperation r= 0.56 infant compulsiveness r=0.15 infant difficultness' r=0.61 infant passivity r=0.58	Inadequate	r=0.65 -0.84(Maternal) [?]

Sr No.	Author, year, country	N	No. of coders	Reliability levels	Construct measured	Reliability evidence	COSMIN methodological quality	Results & Quality of results (Terwee Rating)
18)	Parfitt, 2013, UK	84	2	Reliable	maternal sensitivity, control and unresponsiveness, Infant cooperative, compulsive, difficult and passive	Inter rater reliability on 12% videos. ICC parental sensitivity=0.82, parental control= 0.84, parental unresponsiveness=0.70, child cooperation=0.86, child difficulty=0.90, child compulsivity=0.92, child passivity=0.85	Doubtful	ICC=0.70-0.84 (maternal) ICC= 0.85-0.92 (Infant) [+]
19)	Stephenson, 2018, UK	62	4	Reliable	Maternal sensitivity, Maternal unresponsiveness, Infant cooperative, Infant passiveness	Inter rater reliability on 15% videos. ICC maternal sensitivity 0.78 ICC infant cooperativeness 0.78	Adequate	ICC=0.78 [+]
20)	Underdown, 2013, UK	33	not mentioned	NK	Maternal sensitivity	Inter rater reliability on 36% (n=12) videos. Kappa= 0.73	Doubtful	k=0.73 (Maternal) [+]
21)	Rigby, 2016, UK	40	3	Reliable	Maternal sensitivity	3 coders. One coder scored all the videos. Second coder scored 10 (25%) of the videos for reliability. A third coder solved the discrepancy between the first two coders. In the end inter rater reliability was ICC=0.87	Doubtful	ICC=0.87 (maternal) [+]
22)	Hohenberger, 2012, UK, France, Germany	84(t1) 58(t2)	2	NK	Maternal sensitivity, Maternal control, Maternal unresponsiveness	Munich lab n=10 (M sens r=0.975, M Cont r=0.974, M unresp= 0.986). Munich-London n=12 (M sens r=0.525, M cont r=0.461, M . unresp r=0.946). Munich-Paris n=21 (M sens r=0.77, M cont r=0.38, M unresp r=0.84)	Doubtful	r=0.38-0.98 (Maternal) [?]
23)	Laura Udry-Jorgensen, 2011, Switzerland	32	3	2 blinded coders coded the videos under the supervision of a certified coder.	Dyadic interaction	Coding by 2 blinded independent coders under the supervision of a certified coder. ICC= 0.87 (Maternal characteristics), ICC= 0.86 (Infant characteristics). No mention of % of videos on which Inter rater reliability was checked.	Doubtful	ICC=0.87 (Maternal) ICC=0.86 (infant) [+]

Sr No.	Author, year, country	N	No. of coders	Reliability levels	Construct measured	Reliability evidence	COSMIN methodological quality	Results & Quality of results (Terwee Rating)
24)	Ward, 1995, USA	84 (3 mnths) 81 (9mnths)	3		Maternal sensitivity, Maternal control, Maternal unresponsiveness	Inter rater agreement for 20% videos (16 out of total 80 videos at t2). Maternal sensitivity kappa=0.86. Maternal control kappa=0.847. Maternal unresponsiveness kappa= 0.88. Interrater reliability for the aggregated sensitivity score used in analyses r=0.88. [formula for sensitivity score= (sensitivity*2)-(control + unresponsiveness)]	Doubtful	ICC=0.88 (Maternal) [+]
25)	Fuertes, 2011, Portugal	46	2	One reliable coder coded all the videos and one reliable coder coded 16 random cases	Maternal sensitivity, Maternal control, Maternal unresponsiveness	Coding done by reliable coder. Other coder blindly coded 16 randomly selected cases. Mean Cohen Kappa coefficient for inter rater agreement for M sensitivity=0.95, M control=0.91, M unresponsive=0.84. Total mean of Kappa's from maternal scales is 0.90	Doubtful	k=0.90 (Maternal) [+]
26)	Flykt, 2010, Finland	49	2	One reliable coder coded all the videos and one reliable coder coded n=9 (18%) cases	Maternal sensitivity, Maternal control, Maternal unresponsiveness, Infant cooperative, Infant compulsive, Infant difficult, Infant passive	Both coders were reliable. Inter rater reliabilities indicated by Pearson's correlation coefficient were satisfactory. Maternal sensitivity=0.85 Maternal unresponsiveness=0.65 Infant cooperation=0.79, Infant compulsiveness=0.98 Infant difficulty=0.69 Infant passivity=0.97	Doubtful	ICC r=0.65-0.87 (Maternal). r= 0.79-0.98 (Infant) [+]
27)	Marina Fuertes, 2016, Portugal	82 Triads	3	2 coders scored all the videos (one was trained to reliability while the other was not). 3rd coder who checked 20 random videos was just mentioned as trained.	Maternal and Paternal sensitivity, control and unresponsiveness, Infant cooperative, compulsive, difficult and passive	The two coders who scored the videos (one was trained to reliability, while the other was not). The third coder was trained (no mention of reliability). ICC Maternal sensitivity =0.94 Maternal control=0.96 Maternal unresponsiveness=0.98 Paternal sensitivity=0.98 Paternal control=0.97 Paternal unresponsiveness=0.97ICC Infant (cooperative, compulsive-compliant, difficult, passive) range 0.89 to 0.92.	Doubtful	ICC = 0.94-0.98 (Maternal). ICC=0.97-0.98 (Paternal). ICC=0.89-0.92 (Infant) [+]

Sr No.	Author, year, country	N	No. of coders	Reliability levels	Construct measured	Reliability evidence	COSMIN methodological quality	Results & Quality of results (Terwee Rating)
28)	Conroy, 2009, UK	200	2	One trained coder coded all the videos. A second independent coder coded (n= 15) 27% of the videos.	Maternal sensitivity, unresponsive, controlling	Inter rater reliability (ICC) between two coders. Maternal sensitivity= 0.90	Doubtful	ICC=0.90 (Maternal) [+]
29)	Barlow, 2007, UK	131	2	Not mentioned	Maternal sensitivity, Infant cooperation	A <_2-point difference was observed for 92% of codings for maternal sensitivity and 75% of codings for infant cooperativeness	Inadequate	A <_ 2-point difference was observed for 92% of codings for maternal sensitivity and for 75% of codings for infant cooperativeness [?]
30)	Komoto, 2015, Japan	135	2	Reliable	Maternal and Paternal sensitivity, control and unresponsiveness, Infant cooperative, compulsive, difficult and passive, Dyadic synchrony	Two reliable coders	Inadequate	The coding generated >90% inter-rater reliability on a randomly selected 20% of the videos [?]
31)	Cameron, 2018, UK	66	5	Reliable	Dyadic synchrony	Coder agreement was checked initially and periodically [?]	Inadequate	Coder agreement was checked initially and periodically [?]
32)	O'Shaughnessy, 2012, UK	5	2	Reliable	Dyadic synchrony	No evidence on reliability		
33)	Karmiloff, 2010, UK, France, Germany	T1 (173) T2 (155)	6? 2 coders each in London, Paris & Munich labs	Trained coders	Dyadic synchrony	Inter rater reliability reached between two coders in each of the 3 labs. Both coders first rated n = 10 participants independently. After reaching agreement on divergent scores, participants' behaviours were re-coded. Reliability within each of the three labs was very high (above r = .9 for the sensitive/ cooperative scale, the control/compliant scale and the unresponsiveness/passivity scale.	doubtful	r>0.90 [?]

Sr No.	Author, year, country	N	No. of coders	Reliability levels	Construct measured	Reliability evidence	COSMIN methodological quality	Results & Quality of results (Terwee Rating)
34)	Neuhauser, 2018, Switzerland	211	3	Trained & certified by Crittenden	Maternal sensitivity	Three coders were trained and certified by the author of the measure with an intraclass correlation coefficient (ICC) > .80 for each coder. Additionally, each coder scored a test set with 25 videos of the sample independently and blind to family background information.	Adequate	ICC (maternal sensitivity)= 0.82 among the three coders. [+]
35)	Sidor, 2015, Germany	T1: 302 T3: 274	2	Screening Level reliability (at least two scales >0.70)	Maternal sensitivity	Fisher's r to z transformation provided the following mean Pearson correlations with Crittenden values: maternal sensitivity r=0.65, maternal control r=0.77, maternal unresponsiveness r=0.84, child's cooperation r=0.56, child's compulsiveness r=0.15, child's difficulty r=0.61, child's passivity r=0.58. Coders were not blind to the family intervention status. The data refer to the first and third measurement times.	Inadequate	maternal sensitivity r=0.65, maternal control r=0.77, maternal unresponsiveness r=0.84, child's cooperation r=0.56, child's compulsiveness r=0.15, child's difficulty r=0.61, child's passivity r=0.58 [?]
36)	Udry-Jørgensen, 2015, Switzerland	65	3	Not mentioned	Maternal sensitivity. Paternal sensitivity	One coder assessed parental sensitivity in the Dyadic context for all mothers and fathers in the sample. A second coder double-coded 20 randomly selected videotapes of mother– child interactions, obtaining satisfactory interrater reliability, with an intraclass correlation coefficient (ICC)=0.77. A third coder followed the same procedure for father–child interactions, obtaining an ICC=0.85.	Doubtful	ICC=0.77 maternal sensitivity for double coding 20 videos out of a total 65 videos (between coder 1 and coder 2). ICC=0.85 paternal sensitivity for double coding 20 videos out of a total 65 videos (between coder 1 and coder 3). [+]
37)	Borghini, 2014, Switzerland	78	2 blind & indepen dent coders	One coder Certified by Crittenden	Maternal sensitivity, control and unresponsiveness, Infant cooperative, compulsive, difficult and passive	Video recordings of the interactions were coded by two blinded and independent coders who were not involved in the intervention program. ICC= 0.87 for maternal characteristics ICC=0.86 for infant characteristics	Doubtful	ICC= 0.87 for maternal characteristics ICC=0.86 for infant characteristics [+]
38)	Binda, 2019, Chile	177	3	Trained by Crittenden and passed the reliability test	Maternal sensitivity, control and unresponsiveness, Infant cooperative, compulsive, difficult and passive	The training was for coding Hispanic American dyads, and the coders lived in the same cultural context of the participants. Scoring was conducted independently. ICC for all three coders was > 0.9 for every subscale, $p < 0.001$.	Adequate	ICC for all three coders was > 0.9 for every subscale [+]

Sr No.	Author, year, country	N	No. of coders	Reliability levels	Construct measured	Reliability evidence	COSMIN methodological quality	Results & Quality of results (Terwee Rating)
39)	Crugnola, 2016, Italy	48	2	Not mentioned	Maternal sensitivity, control and unresponsiveness, Infant cooperative, compulsive, difficult and passive	Reliability between two observers was calculated on 20% of the observations of the dyads through the intraclass correlation coefficient ICC = 0.88 for maternal behavior ICC = 0.83 for infant behavior.	Inadequate	ICC = 0.88 for maternal behavior ICC = 0.83 for infant behavior [+]
40)	Emery, 2008, Canada	134	2	Trained by Crittenden	Maternal sensitivity	The 2 Coders were blind to each other's results, obtained a: ICC =0.93 for the maternal sensitivity	Doubtful	ICC=0.93 maternal sensitivity [+]
41)	Paquette, 2004, Canada	96	2	Trained by Crittenden	Maternal sensitivity, control and unresponsiveness, Infant cooperative, compulsive, difficult and passive	Interrater reliability ranged from good to excellent: The correlations obtained for each of the seven scales varied from 0.84 to 0.95.	Doubtful	ICC=0.84-0.95 (for all the scales) [+]

Appendix 5

Results of studies on Hypothesis testing

Sr.no.	Author, year, country	N	Comparator instrument Comparison with other outcome measurement instruments (Convergent validity)	Subgroup Comparison between subgroups (Discriminative or known-groups validity)	Generic Hypotheses	COSMIN methodological quality	Results & Quality of results (Terwee rating)
1)	Azar, 2006, Canada	212	No hypothesis testing for construct validity				
2)	Forcada, 2010, Switzerland	72		Mothers of pre-term infants would be less sensitive in their interactions with their infants , relative to mothers of full-term infants.	Meaningful changes between relevant subgroups. Prematurity has an impact on the dyadic quality of mother-infant interaction.	Doubtful	Result in line with 1 hypo [1+]
3)	Pillhofer, 2014, Germany	83		To see whether mothers would show an improvement in sensitivity (adequate score) to their infants' after intervention	Meaningful changes between relevant subgroups. Mothers who were at high risk for child maltreatment and who completed the program showed enhanced sensitivity after intervention.	Doubtful	Result in line with 1 hypo [1+]
4)	Tester-Jones, 2017 , England	79		Mothers induced to ruminate by thinking and focusing on a goal-based problem would be less sensitive to their infants than mothers who were induced to think about a goal-based problem they had solved	Meaningful changes between relevant subgroups. Rumination causally reduces maternal sensitivity to infants	Doubtful	Result in line with 1 hypo [1+]
5)	Svanberg, 2010, England	192	No hypothesis testing for construct validity				
6)	Soares, 2018, Portugal	86	To investigate the quality of mother-infant interaction with infant development.		Correlations with instruments measuring related, but dissimilar construct should be lower i.e. 0.30-0.50. To assess correlation between maternal sensitivity and infant cooperation (CARE-Index) with infant development score (using SGS II subscales). Significant correlations ranged from 0.215-0.408	Very good	Result in line with 1 hypo [1+]

Sr.no.	Author, year, country	N	Comparator instrument Comparison with other outcome measurement instruments (Convergent validity)	Subgroup Comparison between subgroups (Discriminative or known-groups validity)	Generic Hypotheses	COSMIN methodological quality	Results & Quality of results (Terwee rating)
7)	Neri, 2017, Italy	92		To understand the role of parental gender on the quality of early parent-infant interactions with preterm babies. It was expected to find more difficulties in father-infant interactions compared with mother-infant interactions.	Meaningful changes between relevant subgroups. Result: There was a lack of significant differences in the level of sensitivity between mothers and fathers	Inadequate	Result not in line 1 the hypo [1-]
8)	Kristensen, 2017, Denmark	132	To investigate the correlation between observational outcome (CARE-Index) and self-reported outcome measures; (Karitane parenting confidence scale (KPCS)), Parental stress scale (PSS), Edinburgh Postnatal Depression scale (EPDS) and Ages & Stages Questionnaires (ASQ:SE) after an intervention programme.		Co-relations with instruments measuring related, but dissimilar constructs should be lower, i.e. 0.30-0.50. Results: The improved results on the observational measure were supported by the self-reported data that showed a significantly higher level of maternal confidence (KPCS) and significantly lower level of parental stress (PSS). No significant association with ASQ:SE and EPDS	Doubtful	Result in line with 1 hypo [1+]
9)	Barlow, 2016, UK	31	Whether video feedback intervention could improve the sensitivity of parents of premature babies (CARE-Index),anxiety & depression (HADS), stress (PSI) and traumatic experience (PC-PTSD); once they had left NICU and returned home.		Co-relations with instruments measuring related, but dissimilar constructs should be lower, i.e. 0.30-0.50. Results: Almost all of the data showed non-significant findings, but effect sizes were large on the whole. There was large effect size for CARE-Index scoring and PSI.	Inadequate	Result not in line 1 the hypo [1-]
10)	Steadman, 2007, UK	18	To examine the relationship between maternal mental illness, maternal cognitive function (speed of memory-CDR assessment system) and the quality of mother-infant interaction (maternal sensitivity-CARE-Index).		Co-relations with instruments measuring related, but dissimilar constructs should be lower, i.e. 0.30-0.50. Results: Negative correlation i.e. reduced maternal sensitivity in the presence of maternal mental illness (spearman's rho =-0.49) and slower speed of memory (spearman rho=-0.57)	Inadequate	Result in line with 1 hypo [1+]
11)	Letourneau, 2017, Canada		No hypothesis testing for construct validity				
12)	Feurtes, 2009, Portugal		No hypothesis testing for construct validity				

Sr.no.	Author, year, country	N	Comparator instrument Comparison with other outcome measurement instruments (Convergent validity)	Subgroup Comparison between subgroups (Discriminative or known-groups validity)	Generic Hypotheses	COSMIN methodological quality	Results & Quality of results (Terwee rating)
13)	Kenny, 2013, UK	138		Mothers with mental illness (Schizophrenia, depression, mania) admitted in the MBU and receiving intervention in the form of video feedback would show a change in their CARE-Index scores upon discharge.	Meaningful changes between relevant (sub)groups. Results: No significant effect of diagnosis on the change in either maternal sensitivity or unresponsive scores from admission to discharge. All 3 subgroups showed similar changes in scores.	Doubtful	Result in line with 1 hypo [1+]
14)	Hofer, 2007, Germany	56	No hypothesis testing for construct validity				
15)	Zwonitzer, 2015, Germany	53		Mothers in the high risk for child abuse and neglect group, a significant increase in maternal sensitivity was seen immediately post intervention	Meaningful changes between relevant (sub)groups. Result: F (3,78)=2.7;p=0.048)	Inadequate	Result in line 1 the hypo [1+]
16)	Fuertes, 2009, Portugal	53		To gather 4 independent samples combining 2 risk vs non risk conditions: preterm vs full term and low vs middle SES to study the impact of birth and income status on maternal and infant interactive behaviour. (Full term , middle class; premature, middle class; Full term, low income; premature, low income).	Meaningful changes between relevant (sub)groups. Results: Maternal sensitivity positively correlated with infant cooperation, more so in full term infants from middle class homes (r=0.885). Maternal controlling and infant compliance in all 4 samples (r=0.348, 0.634,0.687,0.731). Infant difficulty positively correlated with maternal unresponsiveness in full term, middle class (r=0.223), but not other three groups. In the other groups, maternal unresponsiveness was correlated with infant passivity (r= 0.476, 0.575, 0.624)	Doubtful	Result in line with 1 hypo [1+]
17)	Sidor, 2011, Germany	133	(1) To test the link between maternal depression (EPDS), stress (PSI) and maternal sensitivity, unresponsiveness and control (CARE-Index) in dyadic interaction.	(2) To test the impact of infant gender and maternal depression on maternal sensitivity	(1) Co-relations with instruments measuring related, but dissimilar constructs should be lower, i.e. 0.30-0.50. Results: EPDS score (maternal depressive symptoms) does not significantly correlate with CARE-Index score (maternal sensitivity, control and unresponsiveness). (2) Meaningful changes between relevant (sub)groups. Results: Two-way ANOVA with 'infant gender' and maternal postpartum depressive symptoms as between -subject factors had no interaction effect on maternal sensitivity	Doubtful	Results not in line with 3 hypo [3-]

Sr.no.	Author, year, country	N	Comparator instrument Comparison with other outcome measurement instruments (Convergent validity)	Subgroup Comparison between subgroups (Discriminative or known-groups validity)	Generic Hypotheses	COSMIN methodological quality	Results & Quality of results (Terwee rating)
18)	Parfitt, 2013, UK	84	To test the impact of pre/postnatal parental (mothers &fathers) mental (anxiety, depression & PTSD) health problems (measured by HADS, PDS) on parent-infant interactive patterns (CARE-Index)		Co-relations with instruments measuring related, but dissimilar constructs should be lower, i.e. 0.30-0.50. Results: No association found between pre/postnatal **depression** and maternal-child interactions. High level of maternal prenatal **anxiety** was associated with less optimal mother-infant interaction. While prenatal (**depression & anxiety**) in the father predicted lower paternal control and infant difficulty but higher paternal unresponsiveness and infant passivity. Paternal postnatal **depression and anxiety** correlated with higher infant difficulty and lower levels of infant passivity.	Doubtful	Result not in line with 1 hypo [1-] Result in line with 1 hypo [1+] Result in line with 1 hypo [1+] Result in line with 1 hypo [1+]
19)	Stephenson, 2018, UK	62	To assess the impact of interventions given at MBU on mother-child interactions (CARE-Index) and maternal mental health status (BPRS & HoNOS).		Correlations with instruments measuring unrelated constructs should be <0.30 Results: No significant correlations between BPRS and CARE-Index scores at admission or discharge or in the changes during the admission period. Significant positive correlation between total HoNOS score and infant passiveness (r=0.26) at admission. No significant correlations between total HoNOS scores and CARE-Index at discharge. Significant correlations between improvement in the HoNOS scores over the admission and the improvement in CARE-Index score for mat sensitivity (r=-0.36), mat unrespon (r=+0.44), infant passivity (r=+0.28)	Very good	Result in line with 1 hypo [1+]
20)	Underdown, 2013, UK	33	No hypothesis testing for construct validity				
21)	Rigby, 2016, UK	40	(1)To assess the correlation between ToM (Frith-Happe Animations) and maternal sensitivity (CARE-Index)	(2) To assess the impact of severe mental illness (Anxiety, Depression, Bipolar affective disorder, Schizo-affective disorder, Schizophrenia) on maternal sensitivity.	(1) Correlations with instruments measuring related, but dissimilar constructs should be lower, i.e. 0.30-0.50.The two scores derived from Frith- Happe Animations were each found to correlate significantly with maternal sensitivity (Intentionality r=0.49, Accuracy r=0.46)	Doubtful	Result in line with 2 hypo [2+]

Sr.no.	Author, year, country	N	Comparator instrument Comparison with other outcome measurement instruments (Convergent validity)	Subgroup Comparison between subgroups (Discriminative or known-groups validity)	Generic Hypotheses	COSMIN methodological quality	Results & Quality of results (Terwee rating)
					(2) Meaningful changes between relevant (sub)groups. Results: Mothers with diagnosis of schizophrenia were rated as least sensitive (3.29) and the difference was significant in comparison with mothers with depression (6.21) and those with bipolar disorder (6.10). Schizo-affective diagnosed mothers obtained highest scores for sensitivity (9.75).		
22)	Hohenberger, 2012, UK. France, Germany	84 (t1) 58(t2)	To assess the correlation of mother-infant interaction quality (CARE-Index) with human goal-directed action (Back of Hand task) and physical causality (CAUS) at 6 and 10 months		Co-relations with instruments measuring unrelated constructs should be <0.30 Results: For human goal-directed action the peak influence is reached at 6 months of age and the interactional quality is moderate control, whereas for understanding of physical causality the peak is at around 10 months and the interactional quality is high sensitivity.	very good	Result in line with 1 hypo [1+]
23)	Laura Udry-Jorgensen, 2011, Switzerland	32	(1) To assess the correlation between dyadic interaction (CARE-Index) at 4 months with attachment security (SSP) at 12 months. (2) To explore the relationship of perinatal risk (PERI) with maternal sensitivity CARE-Index) at 4 months and attachment security (SSP) at 12 months.		(1) Co-relations with instruments measuring similar constructs should be _>0.50. Results: Sensitive dyadic patterns (Sensitive mother +cooperative infant) at 4 months associated with secure attachment at 12 months. While controlling dyadic pattern (Controlling mother + compulsive-compliant infant) at 4 months associated with insecure attachment at 12 months. Dyadic interactive patterns are significantly associated with quality of attachment x2 =8.66 , p=0.034. (2) Correlations with instruments measuring related, but dissimilar constructs should be lower; 0.30-0.50 Results: High PERI and controlling mother at 4 months led to insecure attachment at 12 months. R2=0.32 (Cox &Snell)p=0.006	Adequate	Result in line with 1 hypo [1+]
24)	Ward, 1995, USA	84 (3 months) 81 (9 months)	(1) To assess associations between adult attachment representations (AAI) in adolescent mothers prenatally and maternal sensitivity (Crittenden's measure of sensitivity) at 3 and 9 months postnatally. (2) To assess association between maternal sensitivity at 3 and 9 months with infant-mother attachment (SSP) at 15 months		(1) Co-relations with instruments measuring related, but dissimilar constructs should be lower, i.e. 0.30-0.50. Results: Moderate associations exist between AAI classifications and the two sensitivity scores (3 months r=0.28; 9 months r=0.32, p<0.05). (2) Correlation with instruments measuring similar constructs should be _> 0.50. No significant association	doubtful	Result in line with 1 hypo [1+]

Sr.no.	Author, year, country	N	Comparator instrument Comparison with other outcome measurement instruments (Convergent validity)	Subgroup Comparison between subgroups (Discriminative or known-groups validity)	Generic Hypotheses	COSMIN methodological quality	Results & Quality of results (Terwee rating)
					between SSP at 15 months and sensitivity scores (3 months r=0.06 and 9 months r=0.01, p=n.s)		
25)	Fuertes, 2011, Portugal	46	To investigate the association of maternal behaviour (sensitivity, control and unresponsiveness) (CARE-Index) with infant coping behaviour (Positive other-directed, negative other-directed and self-directed coping behaviour) (FFSF) in a preterm sample at 3 months of age.		Correlations with instruments measuring related, but dissimilar constructs should be lower, i.e. 0.30-0.50. Results: Positive Other-directed style of coping positively correlated with maternal sensitive behaviour (r=0.31, p<0.05). Self-Directed style of coping positively correlated with controlling maternal behaviour (r=0.308, p<0.05)	doubtful	Result in line with 1 hypo [1+]
26)	Flykt, 2010, Finland	49	(1) To assess the association of prenatal and postnatal depression (EPDS) with the dyadic interaction quality (CARE-Index).	(2) To assess the association between maternal attachment style(secure-autonomous, insecure-pre-occupied, insecure-dismissing) and depressive symptoms and dyadic interactions.	(1) Correlation with instruments measuring related but dissimilar constructs should be lower: 0.50-0.30. Results: High level of depressive symptoms both prenatally and postnatally, leads to unresponsiveness towards child. While having high prenatal but low postnatal depressive symptoms is associated with relatively higher maternal unresponsiveness than having high postnatal but low prenatal depressive symptoms. (2) Meaningful changes between relevant subgroups. Results: In dyads with secure autonomous (EPDS) mothers , the levels of maternal sensitivity and child cooperation (CARE-Index) remained adequate (_>7) despite increase in maternal depressive symptoms. Preoccupied mothers were adequately sensitive and their children adequately co-operative(_>7), when mothers had low postnatal depressive symptoms. However, when depressive symptoms increased, the dyads with pre-occupied mothers began to show less than adequate levels of maternal sensitivity and child cooperation (<7)	doubtful	Result in line with 2 hypo [2+]
27)	Marina Fuertes, 2016, Portugal	82 Triads		To evaluate the differences in paternal and maternal sensitivity at 9 and 15 months	Meaningful changes between relevant subgroups: Mothers were rated as being more sensitive than fathers during parent-infant free play at both 9 and 15 months	doubtful	Result in line with 1 hypo [1+]
28)	Conroy, 2009, UK	200		To assess both the independent and combined effects of maternal depression and personality disorder (PD) on infant care. PD cluster A: paranoid, schizoid, schizotypal. PD cluster B: borderline, antisocial, histrionic,	Meaningful changes between relevant (sub)groups. Results: Depression had a detrimental effect on maternal sensitivity only where women had a Personality disorder cluster A (paranoid, schizoid,	doubtful	Result in line with 1 hypo [1+] . Result not in line

Sr.no.	Author, year, country	N	Comparator instrument Comparison with other outcome measurement instruments (Convergent validity)	Subgroup Comparison between subgroups (Discriminative or known-groups validity)	Generic Hypotheses	COSMIN methodological quality	Results & Quality of results (Terwee rating)
				narcissistic. Cluster C: avoidant, dependent, obsessive-compulsive	schizotypal). No problems with maternal sensitivity among women with cluster B PD		with 1 hypo [1-]
29)	Barlow, 2007, UK	131		To assess the effectiveness of an intensive home visiting programme in improving maternal sensitivity and infant cooperativeness in vulnerable families compared with standard treatment.	Meaningful changes between relevant (sub) groups. Results: At 12 months differences favouring the home-visited group were observed on assessment of maternal sensitivity (p<0.04) and infant cooperativeness (p<0.02)	doubtful	Result in line with 1 hypo [1+]
30)	Komoto, 2015, Japan	135	No hypothesis testing for construct validity				
31)	Cameron, 2018, UK	66		To compare the effectiveness of an intervention at improving maternal infant interaction with a control group not receiving the intervention.	Meaningful changes between relevant (sub)groups. The mean Infant CARE-Index scores for both groups were in high-risk range at both intake and post-intervention.	doubtful	Result not in line 1 the hypo [1-]
32)	O'Shaughnessy, 2012, UK	05		To explore the impact of intervention on the quality of mother-infant relationship	Meaningful changes between relevant (sub)groups. Results of Infant CARE-Index analysis coupled with women's narratives indicated a positive shift in the quality of attachment relationships after intervention.	doubtful	Result in line with 1 hypo [1+]
33)	Karmiloff, 2010, UK, France, Germany	T1 (173) T2 (155)	No hypothesis testing for construct validity				
34)	Neuhauser, 2018, Switzerland	211	How does the magnitude of psychosocial stress (Heidelberg stress scale) affect maternal sensitivity (Infant CARE-Index)?		Co-relations with instruments measuring unrelated constructs should be <0.30 Results: Increased psychosocial stress decreases maternal sensitivity (b= -0.22, p=0.00).	Very good	Result in line with 1 hypo [1+]
35)	Sidor, 2015, Germany	T1: 302 T3: 274		To explore the effects of intervention on the quality of maternal sensitivity between intervention and control group	Meaningful changes between relevant (sub)groups. Results: The sensitivity increased in both the control and the intervention group	Doubtful	Result not in line 1 the hypo [1-]
36)	Udry-Jorgensen, 2015, Switzerland	65	To explore the effect of family alliance as measured by Family Alliance assessment scale (FAAS) on parental sensitivity (Infant CARE-Index).		Correlation with instruments measuring related but dissimilar constructs should be lower: 0.50-0.30.	Doubtful	Result in line with one hypo [1+]

Sr.no.	Author, year, country	N	Comparator instrument Comparison with other outcome measurement instruments (Convergent validity)	Subgroup Comparison between subgroups (Discriminative or known-groups validity)	Generic Hypotheses	COSMIN methodological quality	Results & Quality of results (Terwee rating)
					Result: A significant main effect for family alliance: parental sensitivity is globally higher in high-coordination families (M = 9.97, SD = 2.35) than in low-coordination families (M = 8.24, SD = 3.15).		
37)	Borghini, 2014, Switzerland	78		To investigate the impact of an early intervention on the quality of mother–infant interactions, in a sample of very preterm infants and their mothers and compare it preterm infants without interventions and full-term infants without interventions.	Meaningful changes between relevant (sub)groups. Results: At 4 months, during the third mother–infant interactive play, maternal sensitivity and infant cooperation were found to be significantly better, and infants also significantly less difficult in the group of preterm infants with intervention, compared with mother and infant interactive qualities during the first play.	Doubtful	Result in line with one hypo[1+]
38)	Binda, 2019, Chile	177	To investigate postnatal depressive symptoms (EPDS) would be associated with a lower maternal sensitivity (Infant CARE-Index)		Correlation with instruments measuring related but dissimilar constructs should be lower: 0.50-0.30. Results: Low quality mother–infant interaction in mothers with postnatal depression symptoms demonstrated a low maternal sensitivity and higher maternal controlling scores.	Very good	Result in line with one hypo[1+]
39)	Crugnola, 2016, Italy	4		To investigate whether, the adolescent mothers who use the interventions will increase their Sensitivity style and decrease their Controlling style and whether their infants will have more Cooperative and fewer Passive styles after 3 and 6 months of intervention compared to mothers and infants in the control group?	Meaningful changes between relevant (sub)groups Results: At the post-intervention assessment by the Care-Index, the intervention group at 9 months reached an average score of 8.6 for mothers and 8.2 for infants, indicating an adequate quality of mother and infant interaction, while the control group went down to an average score of 4.0 for the mothers and 3.5 for the infants, remaining in the “in need of further intervention” category.	Doubtful	Result in line with one hypo[1+]
40)	Emery, 2008, Canada	134	No Hypothesis testing for construct validity				
41)	Paquette, 2004, Canada	93		To compare the parental sensitivity of 33 adolescent mothers in group homes with that of a school group of 63 adolescent mothers divided into two subgroups according to the presence(28) or absence(27) of a conduct disorder diagnosis.	Meaningful changes between relevant (sub)groups Results: The Sensitivity mean score of the school group without a conduct disorder diagnosis tended to be higher than that of the other two groups ($F(2, 80) = 3.04$, $p < .06$).	Doubtful	Result in line with one hypo[1+]

Appendix 6

Results of studies on Responsiveness

Sr No.	Author, year, country	N	Number of observations	Construct approach (Hypotheses testing; Comparison with other outcome measurement instruments)	Construct approach: (Hypotheses testing: Comparison between subgroups)	Construct approach: (Hypotheses testing: before and after intervention)	Generic hypotheses	COSMIN methodological quality	Results & Quality of results(Terwee rating)
1)	Stephenson, 2018, UK	62	2	1)To assess correlation between Infant CARE-Index score and BPRS Results: No significant correlations between BPRS and CARE-Index scores at admission or discharge or in the changes during the admission period. 2)To assess correlation between Infant CARE-Index score and HoNoS Result: Significant positive correlation between total HoNOS score and infant passiveness (r=0.26) at admission. No significant correlations between total HoNOS scores and CARE-Index at discharge. Significant correlations between improvement in the HoNOS scores over the admission and the improvement in CARE-Index score for mat sensitivity (r=-0.36), mat unresponsiveness (r=+0.44), infant passivity (r=+0.28) (CARE-Index) Mean (SD) Change in Infant CARE-Index scores after intervention (Maternal sensitivity, maternal unresponsiveness and infant cooperation) -significant correlation. Mean (SD) Change in Infant CARE-Index score after intervention (Infant passiveness) - no significant change.		To assess the impact of interventions given at MBU on mother-child interactions (CARE-Index) Mean (S.D) Change in Infant CARE-Index scores after intervention (Mat sens, mat unresp and infant cooperation) -significant correlation Mean (S.D) Change in Infant CARE-Index score after intervention (Infant passiveness) - no significant change	Correlations with instruments measuring unrelated constructs should be <0.30	Doubtful	1)Result not line with 1 hypo [1-] 2)Result in line with 1 hypo [1+] 3)Result in line with 1 hypo[1+] 4)Result not in line with 1 hypo [1-]

Sr No.	Author, year, country	N	Number of observations	Construct approach (Hypotheses testing; Comparison with other outcome measurement instruments)	Construct approach: (Hypotheses testing: Comparison between subgroups)	Construct approach: (Hypotheses testing: before and after intervention)	Generic hypotheses	COSMIN methodological quality	Results & Quality of results(Terwee rating)
2)	Underdown, 2013, UK	33	2			Which parent-infant dyads (Low/Moderate/high-risk) will benefit the most from (intervention) infant massage programme (Good/fair/poor program) regarding change in maternal sensitivity? Result: Moderate risk dyads attending a Good quality program. Sensitivity Score: 7 (pre-intervention) Sensitivity Score: 8 (post-intervention)		doubtful	Result in line with 1 hypo [1+]
3)	Hohenberger, 2012 UK, France, Germany	84 (t1) 58(t2)	2	To assess the correlation of mother-infant interaction quality (CARE-Index) with human goal-directed action (Back of Hand task) and physical causality (CAUS) at 6 and 10 months Results: For human goal-directed action the peak influence is reached at 6 months of age and the interactional quality is moderate control. Whereas for understanding of physical causality the peak is at around 10 months and the interactional quality is high sensitivity.			Co-relations with instruments measuring unrelated constructs should be <0.30	Very good	Result in line with 1 hypo [1+]
4)	Ward, 1995, USA	85(t1) 80(t2)	2	To examine association between adolescents prenatal AAI classifications at 3 and 9 months postnatally. Results: Moderate association exists between AAI classification and two sensitivity scores(3 months: r=0.28; 9 months: r=0.32)			Correlations with (changes in) instruments measuring unrelated construct should be<0.30	Doubtful	Result in line with 1 hypo [1+]
5)	Komoto, 2015 Japan	135	4			Scores for dyadic synchrony on Infant CARE-Index would improve after intervention Result: Improve in dyadic synchrony scores Mean (SD) 4.20(1.61) -pre-intervention 1- 3months, 6.60 (1.80)-6 months 6.73(1.28)- 9 months 7.47 (1.81)- 12 months		Doubtful	Result in line with 1 hypo [1+]

Sr No.	Author, year, country	N	Number of observations	Construct approach (Hypotheses testing; Comparison with other outcome measurement instruments)	Construct approach: (Hypotheses testing: Comparison between subgroups)	Construct approach: (Hypotheses testing: before and after intervention)	Generic hypotheses	COSMIN methodological quality	Results & Quality of results(Terwee rating)
6)	Cameron, 2018, UK	66	2	Comparison with other measurement instrument		Scores for dyadic synchrony on Infant CARE-Index would improve after intervention Result: The mean scores remained in high risk range at intake and post intervention 3.42-pre-intervention 4.19-post-intervention		Doubtful	Result not in line 1 the hypo [1-]
7)	O' Shaughnessy, 2012, UK	4 to 12	2		Results of Infant CARE-Index analysis coupled with women's narratives indicated a positive shift in the quality of attachment relationships after intervention		Meaningful changes between relevant (sub)groups.	Doubtful	Result in line with 1 hypo [1+]
8)	Zwonitzer, 2015, Germany	53	4			To investigate effectiveness of intervention on maternal sensitivity. Observations at pre-intervention, at end of intervention, child 6 months, child 12 months of age. Results: High risk mothers had short term increase in sensitivity after intervention seen at infant's 12 months age.		Inadequate	Result in line with 1 hypo [1+]
9)	Marina Fuertes, 2016, Portugal	82 triads	2		To evaluate the differences in paternal and maternal sensitivity at 9 and 15 months. Result: Mothers were rated as being more sensitive than fathers during parent-infant free play at both 9 and 15 months. 9 months Sensitivity: Mean (SD). Mother: 8.19 (2.2). Father: 7.41(2.3). 15 months sensitivity: Mother:8.50 (2.2) Father: 7.79 (2.2)		Meaningful changes between relevant subgroups:	Doubtful	Result in line with one hypo [1+]

Sr No.	Author, year, country	N	Number of observations	Construct approach (Hypotheses testing; Comparison with other outcome measurement instruments)	Construct approach: (Hypotheses testing: Comparison between subgroups)	Construct approach: (Hypotheses testing: before and after intervention)	Generic hypotheses	COSMIN methodological quality	Results & Quality of results(Terwee rating)
10)	Sidor, 2015, Germnay	T1: 302 T3: 274	2			To investigate effectiveness of intervention on maternal sensitivity. Results: There was an improvement in maternal sensitivity between pre and post data collection. Time point: M (SD) T1: 5.52(2.60) T3: 5.82(2.29)	Correlations with changes in instruments measuring similar construct should differ by a minimum of 0.10	Doubtful	Result in line with one hypo [1+]
11)	Borghini, 2014, Switzerland	78	2			To investigate effect of interventions on maternal sensitivity and infant cooperation Result: Significant increases in maternal sensitivity and infant cooperation as well as a significant decrease in infant difficulty and a marginal decrease in maternal control were found. Paired difference between 1st & 2nd Infant CARE-Index Mat Sens=1.84 (0.68). Mat Cont=-1.32 (0.67). Mat Unresp=-0.53(0.39). Inf Coop=2.95(0.60). Inf compl=-0.53(0.45) Inf Diff=-1.68(0.61). Inf pass=-0.74(0.64)	Correlations with changes in instruments measuring similar construct should differ by a minimum of 0.10	Doubtful	Result in line with one hypo [1+]
12)	Crugnola, 2016, Italy	T1: 48 T3: 28	2			To investigate the effect of interventions on maternal and infant behaviours. Result: At pre-intervention (3 months), Sensitivity score was 5 and 4.3 for infant cooperation. At the post-intervention assessment by the CARE-Index, the	Correlations with changes in instruments measuring similar construct should differ by a minimum of 0.10	Doubtful	Result in line with one hypo [1+]

Sr No.	Author, year, country	N	Number of observations	Construct approach (Hypotheses testing; Comparison with other outcome measurement instruments)	Construct approach: (Hypotheses testing: Comparison between subgroups)	Construct approach: (Hypotheses testing: before and after intervention)	Generic hypotheses	COSMIN methodological quality	Results & Quality of results(Terwee rating)
						intervention group at 9 months reached an average score of 8.6 for mothers and 8.2 for infants, which indicates an adequate quality of mother and infant interaction.			

Appendix 7

Participant Information Sheet for focus groups with Infant CARE-Index trainees

Title of study: Qualitative analysis on the acceptability of Infant CARE-Index in the UK

We would like to invite you to take part in the above-named study but before you decide, please read the following information.

What is the purpose of this study?

This study is a part of a larger research study investigating the properties and the use of Infant CARE-Index in UK settings. The purpose of this study is to explore the acceptability of Infant CARE-Index in routine care from the perspective of Infant CARE-Index trainees.

The focus groups will briefly focus on questions related to the measures used for assessing parent-infant interactions, Infant CARE-Index training and use of Infant CARE-Index in professional /research settings.

Who is doing the study?

This study will be conducted by Ms XYZ, a research student at Department of Health Sciences, University of York, UK. This study will be supervised by Professor ABC and Dr DEF, Department of Health Sciences, University of York.

Why have I been asked to participate?

You have been requested to participate in this study because you have either previously attended Infant CARE-Index training. We are interested in finding about your experience and perspectives.

Do I have to take part?

Participation in this study is entirely voluntary. You will need to complete the consent form, if you are willing to participate.

What will be involved if I take part in this study?

You will be required to participate in a focus group, which will last 25-30 minutes approximately at University of York. This will be on one of the last three days of your nine-day training course in Infant CARE-Index. The research student will be moderating this focus group. Topics covered will mainly be in relation to your training experience and how you intend to use Infant CARE-Index in the future.

What are the advantages/ benefits of taking part?

Although there are no direct benefits to participants for taking part; however, it is hoped that this work will have a beneficial impact on increasing the knowledge of how, where and why and in what capacity is Infant CARE Index is used within different UK settings. The results will be shared with participants in order to inform their professional/research work.

What are the disadvantages/risks of taking part?

Participating in the research is not anticipated to cause you any risk or disadvantage.
Can I withdraw from the study at any time?

You have the right to withdraw **during** the focus group without giving any prior reason. Beyond that time point, the recorded data will be transcribed and anonymised and it will not be possible to extract and remove an individual's data from the dataset.

How will the information and personal data I give be handled?

Any sensitive/personal data that might come forward will be handled in accordance with the General Data Protection Regulation (GDPR) ***https://www.york.ac.uk/records-management/dp/.*** The focus groups will be recorded and transcribed. However, you will not be identified in the recordings – we will give you a number at the start of the focus group and refer to you by number throughout. We will not tell anyone that you have taken part in the focus group, although there is of course a possibility that another member of the group might recognise you. We will also not name you in any of our reports or publications. In addition, all participants will refer to one another by number (e.g. participant 4, participant 6). In our reports and publications, we will use quotes from the focus groups to help illustrate the points that are being made. Some of these quotes may come from you. We will not use any quotes that might reveal who you are. Your anonymised data will be held confidentially, with access restricted to staff working on the E-SEE project. All recordings will be destroyed once transcripts have been made and checked. The transcribed data may be preserved (for a minimum of 10 years) at the end of M Res. The consent forms will be kept in a locked filing cabinet in the Department of Health Sciences, York.

What will happen to the results of the study?

The findings of this study are intended to be disseminated at conferences and published in a maternal and child wellbeing journal, within one year of completion this research degree.

Who has reviewed and approved this study?

The study has been granted ethical approval by University of York's Health Sciences Research Governance Committee.

Who do I contact for more information about the study?

For further information please contact:

Ms XYZ

M Res student

Department of Health Sciences

University of York

Heslington

YO10 5HDD

Email: MsXYZ@york.ac.uk

Who do I contact in the event of a complaint?

For general complaints, please contact:

Prof ABC: profabc@york.ac.uk

If participants are unhappy with the way their personal data has been handled, they have a right to complain to University of York's Data Protection Officer at: dataprotection@york.ac.uk

If still unsatisfied, the participants have a right to report concerns to the Information Commissioner's Office at:

www.ico.org.uk/concerns

Thank you for taking the time to read this information sheet

Appendix 8

Participant Consent Form for focus groups with Infant CARE-Index trainees

Title of study: Qualitative analysis on the acceptability of Infant CARE-Index in the UK

	Please confirm agreement to each statement by putting your initials in the boxes below
I have read and understood the participant information sheet [date ..., version ...]	
In case of query, I have had the opportunity to ask questions and discuss this study	
I have received satisfactory answers to all of my questions	
I have received enough information about the study	
I understand my participation in the study is voluntary and that I am free to withdraw from the study: -At any time during the focus group without giving a reason for withdrawing.	
I understand that relevant sections of my responses collected during the study may be looked at by researchers. I give permission for these individuals to have access to my responses.	
I understand that any information I provide, including personal data, will be kept confidential, stored securely and only accessed by those carrying out the study.	
I understand that any information I give may be included in published documents, but all information will be anonymised.	
I agree to take part in this study	
Participant Signature …… Date	
Name of Participant	
Researcher Signature …….. Date	
Name of Researcher	

Appendix 9

Focus group guide for Infant CARE-Index trainees

Title of Study: Qualitative analysis on the acceptability of Infant CARE-Index in the UK

Expected duration:25-30 mins

General questions in relation to assessments procedures for parent-infant interactions

1) What assessment methods (observational /parent reported/interviews) are you aware of, other than Infant CARE-Index, or have used for assessing parent-infant interactions in the age group 0-15 months?

Specific questions in relation to Infant CARE-Index training

2) What was the primary purpose of your Infant CARE-Index training?
3) Is the coding methodology and scoring the various maternal and infant behaviours difficult or easy to carry out?
4) How confident are you at using this tool and interpreting the results?
5) How easy or difficult is it to convey the instructions to parent-infant dyads for video recording of Infant CARE-Index during your training?
6) Is there anything that you would like to add, or thought was left out during our discussion?

Appendix 10

Participant Information Sheet for focus groups with Health professionals

Title of Study: Qualitative analysis on the acceptability of Infant CARE-Index in the UK

We would like to invite you to take part in the above-named study but before you decide, please read the following information.

What is the purpose of this study?

This study is a part of a larger research study investigating the properties and the use of Infant CARE-Index in UK settings. The purpose of this study is to explore the acceptability of Infant CARE-Index in routine care from the perspective of Health professionals.

The focus groups will briefly focus on questions related to the measures used for assessing parent-infant interactions and use of Infant CARE-Index in professional /research settings.

Who is doing the study?

This study is being conducted by Ms XYZ, a research student at Department of Health Sciences, University of York, UK. This study is being supervised by Professor ABC and Dr DEF Department of Health Sciences, University of York.

Why have I been asked to participate?

You have been requested to participate in this study because you are involved in providing health services for 0-19yrs age group. We are interested in finding about your experience and perspectives.

Do I have to take part?

Participation in this study is entirely voluntary. You will need to complete the consent form, if you are willing to participate.

What will be involved if I take part in this study?

You will be required to participate in a focus group, which will last 25-30 minutes approximately on the day of your 0-19 workshop/ meeting. The research student shall be moderating this focus group.

What are the advantages/ benefits of taking part?

Although there are no direct benefits to participants for taking part; however, it is hoped that this work will have a beneficial impact on increasing the knowledge of how, where and why and in what capacity is Infant CARE Index is or can be used within different UK settings. The results will be shared with participants in order to inform their professional/research work.

What are the disadvantages/risks of taking part?

Participating in the research is not anticipated to cause you any risk or disadvantage.

Can I withdraw from the study at any time?

You have the right to withdraw **during** the focus group without giving any prior reason. Beyond that time point, the recorded data will be transcribed and anonymised and it will not be possible to extract and remove an individual's data from the dataset. The data collected will be used

for research purposes, as stated on the Consent Form and every detail for anonymity will be maintained. All the information that we collect will be kept strictly confidential. You will not be identified or identifiable in any reports or publications. Your organisation will also not be identified or identifiable. Data collected may be shared in an anonymised form to allow reuse by the research team and other third parties.

How will the information and personal data I give be handled?

Any sensitive/personal data that might come forward will be handled in accordance with the General Data Protection Regulation (GDPR) https://www.york.ac.uk/records-management/dp/. The focus groups will be recorded and transcribed. We will not tell anyone that you have taken part in the focus group, although there is of course a possibility that another member of the group might recognise you. We will also not name you in any of our reports or publications. In our reports and publications, we will use quotes from the focus groups to help illustrate the points that are being made. Some of these quotes may come from you. However, we will not use any quotes that might reveal who you are. Your anonymised data will be held confidentially, with access restricted to staff working on the (Enhancing social and emotional wellbeing in the early years) E-SEE project, which is being run from University of York and is using Infant CARE-Index. All recordings will be destroyed once transcripts have been made and checked. The consent forms will be kept in a locked filing cabinet in the Department of Health Sciences, York.

What will happen to the results of the study?

The findings of this study are intended to be disseminated at conferences and published in a maternal and child wellbeing journal, within one year of completion this research degree.

Who has reviewed and approved this study?

The study has been granted ethical approval by University of York's Health Sciences Research Governance Committee.

Who do I contact for more information about the study?

For further information please contact:

Ms XYZ

M Res student

Department of Health Sciences

University of York

Heslington

YO10 5HDD

Email: MsXYZ@york.ac.uk

Who do I contact in the event of a complaint?

For general complaints, please contact: Prof ABC: profabc@york.ac.uk

If participants are unhappy with the way their personal data has been handled, they have a right to complain to University of York's Data Protection Officer at: dataprotection@york.ac.uk

If still unsatisfied, the participants have a right to report concerns to the Information Commissioner's Office at: www.ico.org.uk/concerns

Thank you for taking the time to read this information sheet.

Appendix 11

Participant Consent Form for focus groups with Health professionals

Title of study: Qualitative analysis on the acceptability of Infant CARE-Index in the UK

	Please confirm agreement to each statement by putting your initials in the boxes below
I have read and understood the participant information sheet [date ..., version ...]	
In case of query, I have had the opportunity to ask questions and discuss this study	
I have received satisfactory answers to all of my questions	
I have received enough information about the study	
I understand my participation in the study is voluntary and that I am free to withdraw from the study: -At any time during the focus group without giving a reason for withdrawing.	
I understand that relevant sections of my responses collected during the study may be looked at by researchers. I give permission for these individuals to have access to my responses.	
I understand that any information I provide, including personal data, will be kept confidential, stored securely and only accessed by those carrying out the study.	
I understand that any information I give may be included in published documents, but all information will be anonymised.	
I agree to take part in this study	
Participant Signature …… Date	
Name of Participant	
Researcher Signature …….. Date	
Name of Researcher	

Appendix 12

Focus group guide for Health professionals

Title of study: Qualitative analysis on the acceptability of Infant CARE-Index in the UK

Expected duration: 20-25 minutes

General questions about parent-infant interaction assessments

1) Have you been involved in assessing parent-infant interactions with the help of an observational tool in the age range 0-15 months?
2) How do you use these assessments, e.g. to screen and identify problems so support could be offered, or to monitor change over time etc ?
3) What are the different measurement tools that you use in assessing parent-infant interactions for the age range birth-15months?
4) What are the key behaviours that you look for regarding parent-infant interactions, when you are carrying out your assessments?
5) Would you be interested in learning or investing in using an observational tool?
6) Has your organisation ever arranged a course or workshop to learn the use of an observational or any other tool?

A brief description of Infant CARE-Index (3 minutes) will be provided to the participants, before asking the next questions.

Specific questions relating to the Infant CARE-Index

7) Have you heard about the Infant CARE-Index?
8) Would you ever consider investing in this tool for routine assessments?

   prompt - if not/yes why?

9) Do you think cost of training/continuous support might be a contributing factor towards uptake by Health professionals?
10) Do you think CARE-Index might or might not compliment in carrying out your assessments/ interventions/ program outcomes evaluations?
11) Do you think parents will be willing to participate in video recording for CARE-Index?
12) What barriers do you think you might face in carrying out this kind of assessment?
13) What would help families overcome these barriers?
14) Is there anything that you would like to add that you thought was left out during our discussion?

Appendix 13

Consolidated criteria for reporting qualitative research (COREQ): A checklist for focus groups

No. Item	Guide questions/description	Reported in section no.
Domain 1: Research team and reflexivity *Personal Characteristics*		
1.Focusgroup facilitator	Which author/s conducted the focus group?	Section 3.1.1 Researcher who conducted the study
2. Credentials	What were the researcher's credentials? E.g. PhD, MD	Section 3.1.1 M Res student
3. Occupation	What was their occupation at the time of the study?	Section 3.1.1 Student
4. Gender	Was the researcher male or female?	Section 3.1.1 Female
5.Experience and training	What experience or training did the researcher have?	Section 3.1.1 The researcher undertook a Qualitative research methods module
Relationship with participants		
6.Relationship established	Was a relationship established prior to study commencement	Section 3.3 Yes
7.Participant knowledge of the interviewer	What did the participants know about the researcher? e.g. personal goals, reasons for doing the research	Section 3.3 Participant information sheet and Consent form
8.Interviewer characteristics	What characteristics were reported about the inter viewer/facilitator? e.g. Bias, assumptions, reasons and interests in the research topic	Section 3.1.1 & 3.3 The participants knew that the researcher was a student
Domain 2: Study design. *Theoretical framework*		
9.Methodological orientation and Theory	What methodological orientation was stated to underpin the study? e.g. grounded theory, discourse analysis, ethnography, phenomenology, content analysis	Section3.1.1 – Exploratory qualitative design
Participant selection		

10.Sampling	How were participants selected? e.g. purposive, convenience, consecutive, snowball	Section 3.2.2 Purposive sampling
11.Method of approach	How were participants approached? e.g. face-to-face, telephone, mail, email	Section3.2.2 face-to-face, email
12. Sample size	How many participants were in the study?	Section 3.3 Focus group with Infant CARE-Index trainees- 5 participants Focus group with health professionals- 20 participants
13.Non- participation	How many people refused to participate or dropped out? Reasons?	Section3.3 One (It was because of other commitments)
Setting		
14.Setting of data collection	Where was the data collected? e.g. home, clinic, workplace	Section 3.3 -FG with trainees, University of York -FG with health professionals, Huddersfield Town Hall
15. Presence of non-participants	Was anyone else present besides the participants and researchers?	Section 3.3. No
16.Description of sample	What are the important characteristics of the sample? e.g. demographic data	Section 3.3 Psychologists, health visitors, researchers. All participants were females
Data collection		
17.Focus group guide	Were questions, prompts, guides provided by the authors? Was it pilot tested?	Section 3.3 Focus groups guides used (see Appendices 9 & 12) Not pilot tested
18. Repeat interviews	Were repeat interviews carried out? If yes, how many?	N/A
19. Audio/visual recording	Did the research use audio or visual recording to collect the data?	Section 3.3 Yes (Audio recording)
20. Field notes	Were field notes made during and/or after the interview or focus group?	Yes
21. Duration	What was the duration of the inter views or focus group?	Section 3.3 20-25 min
22. Data saturation	Was data saturation discussed?	No
23.Transcripts returned	Were transcripts returned to participants for comment and/or correction?	No

Domain 3: analysis and finding *Data analysis*		
24. Number of data coders	How many data coders coded the data?	Section 3.4 One
25. Description of the coding tree	Did authors provide a description of the coding tree?	No
26.Derivation of themes	Were themes identified in advance or derived from the data?	Section3.4 Derived from the data
27. Software	What software, if applicable, was used to manage the data?	Section 3.4 NVivo
28.Participant checking	Did participants provide feedback on the findings?	No
Reporting		
29. Quotations presented	Were participant quotations presented to illustrate the themes/findings? Was each quotation identified? e.g. participant number	Section 3.5 Yes
30. Data and findings consistent	Was there consistency between the data presented and the findings?	Section 3.5 Yes
31. Clarity of major themes	Were major themes clearly presented in the findings?	Section 3.5 Yes
32. Clarity of minor themes	Is there a description of diverse cases or discussion of minor themes?	Section 3.5 Yes

References

Abedi, Aidin., Mokkink, Lidwine. B., Zadegan, Shayan, Abdollah., Paholpak, Permsak. , Tamai, Koji., Wang, Jeffrey C., and Buser, Zorica. (2019). Reliability and Validity of the AOSpine Thoracolumbar Injury Classification System: A Systematic Review. *Global spine journal*, 9(2), pp.231–242.

Ainsworth, M.D.S. & Bell, S.M.(1970). Attachment, exploration, and separation: Illustrated by the behavior of one-year-olds in a strange situation. *Child Develop*, 41, 1, 49-67, 70. Available at: https://pdfs.semanticscholar.org/8272/bd76f36d195023f245735e23e6b5c8b19afd.pdf.

Ainsworth, M.D.S., Blehar, M.C., Waters, E. and Wall, S. (1978). *Patterns of Attachment: A Psychological Study of the Strange Situation.* Hillside: Erlbaum

Barlow, J., Schrader McMillan, A., Kirkpatrick, S., Ghate, D., Smith, M., Coram, T. and Barnes, J. (2008). Health-Led Parenting Interventions in Early Years. Research Report DCSF RW070. *Department for Schools, Children and Families*

Barlow, J., Smailagic, N., Huband, N., Roloff, V., & Bennett, C. (2012). Group-based parent training programmes for improving parental psychosocial health. A systematic review, 15. doi: 0.4073/csr.2012/15

Barlow, J. et al. (2013). An evaluation of the parents under pressure programme: a study protocol for an RCT into its clinical and cost effectiveness. *Trials*, 14, p.210

Barlow, J., Sembi, S., & Underdown, A. (2016). Pilot RCT of the use of video interactive guidance with preterm babies. Journal of Reproductive and Infant Psychology, 34(5), 511–524.

Barlow, J. & Coren, E. (2018). The Effectiveness of Parenting Programs: A Review of Campbell Reviews. *Research on social work practice*, 28(1), pp.99–102.

Barnes, J. et al. (2013). First steps: study protocol for a randomized controlled trial of the effectiveness of the Group Family Nurse Partnership (gFNP) program compared to routine care in improving outcomes for high-risk mothers and their children and preventing abuse. *Trials*, 14, p.285.

Barnes, J., Stuart, J., Allen, E., Petrou, S., Sturgess, J., Barlow, J. Elbourne, D. (2017). Randomized controlled trial and economic evaluation of nurse-led group support for young mothers during pregnancy and the first year postpartum versus usual care. Trials, 18(1), 508.

Biringen, Z., Derscheid, D., Vliegen, N., Closson, L, M., Easterbrooks, A. (2014).Emotional availability (EA): Theoretical background, empirical research using the EA Scales, and clinical applications. Developmental review. 34:2.pp 114-167

Bowlby, J. (1953). *Child Care the Growth of Love.* Baltimore, MD: Pelican Books.

Bradley, R. H., & Caldwell, B. M. (1977). Home observation for measurement of the environment: a validation study of screening efficiency. American Journal of Mental Deficiency 81(5), 417–420.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology, Qualitative Research in Psychology, 3:2, 77-101

Bywater, T., Berry, V., Blower, S., *et al* (2018). Enhancing Social-Emotional Health and Wellbeing in the Early Years (E-SEE): a study protocol of a community-based randomised controlled trial with process and economic evaluations of the incredible years infant and toddler parenting programmes, delivered in a proportionate universal model. *BMJ Open.* **8:**e026906. doi:10.1136/bmjopen-2018-026906

Cassidy, B., Zoccolillo, M., & Hughes, S. (1996). Psychopathology in Adolescent Mothers and Its Effects on Mother–Infant Interactions: A Pilot Study. *Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie*, *41*.

Charmaz, K. (2006). Constructing grounded theory: A practical guide through qualitative analysis. Thousand Oaks, CA: SAGE.

Chiarotto, A., Ostelo, R, W., Boers, M., et al. (2018). A systematic review highlights the need to investigate the content validity of patient-reported outcome measures for physical functioning in patients with low back pain. *J Clin Epidemiol*. 95:73-93.

Clarke, V & Braun, V. (2013). Successful Qualitative Research: A Practical Guide for Beginners. SAGE publications.

Comfort, M., Gordon, P.R., and Naples. D. (2011). KIPS: An evidence-based tool for assessing parenting strengths and needs in diverse families. Infants and young children: An interdisciplinary journal of early childhood interventions. 24(1). 56-74

Coyle, A. (2006). Discourse analysis. In G.M. Breakwell, S. Hammond & C. Fife-Schaw (Eds), Research methods in psychology (3rd ed., pp366-387). London: Sage.

CRD. (2009).Systematic Reviews: CRD"s guidance for undertaking reviews in health care. https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf

Crittenden, P. M. (1981). Abusing, Neglecting, Problematic, and Adequate Dyads: Differentiating by Patterns of Interaction. Merrill-Palmer Quarterly of Behaviour and Development, 27(3), 201–21

Crittenden, P.M., Bonvillian, J. D. (1984). The relationship between maternal risk status and maternal sensitivity. *Amer. J. Orthopsychiat*. 54(2)

Crittenden, P. M. (1984). Sibling interaction: Evidence of a generational effect in maltreating infants. *Child Abuse & Neglect.* Vol. 8: 433-438.

Crittenden, P. M. (1985). Maltreated infants: Vulnerability and resilience. *J. Child Psychol. Psychiat*. 26(1), 85-96

Crittenden, P. M. (1987). Non-organic failure to thrive: Deprivation or Distortion? *Infant Mental Health Journal.* 8(1)

Crittenden, P.M. (1988). Distorted patterns of relationship in maltreating families: The role of internal representation models. *Journal of Reproductive and Infant Psychology*. Vol.6: 183-199Crittenden, P.M. & Dilalla, D.L. (1988). Compulsive Compliance: The Development of inhibitory coping strategy in infancy. Journal of Abnormal Child Psychology. 16(5):585-599

Crittenden, P.M. (1992). Children's strategies for coping with adverse home environments: An interpretation using attachment theory. International journal of Child Abuse & Neglect, 16, 329-343.

Crittenden, P.M. (2005). Using the CARE-Index for Screening, Intervention, and Research. Retrieved from: www.patcrittenden.com

Crittenden, P.M. (2006). A Dynamic maturational model of Attachment. ANZJFT. 27(2):105–115

Crittenden, P. M. (2008). Raising parents: Attachment, parenting and child safety. Devon, UK: William Pub

Crittenden, P. M. (2010). CARE-INDEX. Infants Coding Manual. Family Relations Institute, Miami, USA. Unpublished.

De Vet HCW, Terwee CB, Mokkink LB, Knol DL.(2011). Measurement in medicine. A practical guide. first ed. Cambridge: Cambridge Univ Press.

Department for Education.(2018). Statistics: children in need and child protection.

DMM publications. (2018). Available at: *https://familyrelationsinstitute.org/wp-content/uploads/DMM-Publications-2018.pdf*.

Early Intervention Foundation (EIF report).(2018). Realising-the-potential-of-early-intervention.pdf. Available online: *https://www.eif.org.uk/report/realising-the-potential-of-early-intervention*

Eyberg, S.M., Nelson, M.M., Duke, M. and Boggs, S.R. (2009). Manual for the Dyadic Parent-Child Interaction Coding System ,3rd Ed. University of Florida

Girard, J.M. & Cohn, J.F. (2016). A Primer on Observational Measurement. *Assessment*, 23(4), pp.404–413.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry*, *38*(5), 581–586.

GRADE. GRADE Handbook. (2013). Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach.

Green, H., McGinnity, Á., Meltzer, H., Ford, T., Goodman, R., et al. (2005). *Mental health of children and young people in Great Britain, 2004*. Palgrave Macmillan Basingstoke.

Gridley, N. et al., 2019. Psychometric Properties of Parent-Child (0-5 years) Interaction Outcome Measures as Used in Randomized Controlled Trials of Parent Programs: A Systematic Review. *Clinical child and family psychology review*, 22(2), pp.253–271.

Groh, A. M., Fearon, R. M. P., van IJzendoorn, M. H., Bakermans-Kranenburg, M. J., & Roisman, G. I. (2017). Attachment in the Early Life Course: Meta-Analytic Evidence for Its Role in Socioemotional Development. *Child Development Perspectives*, *11*(1), 70–76.

Guest, G., MacQueen, K., & Namey, E. (2012). Applied thematic analysis. Thousand Oaks, CA: Sage

Guest, B.G., Namey, E.E. & Mitchell Pub., M. (2017). Sampling in Qualitative Research. Available at: http://dx.doi.org/10.4135/9781506374680.

Gunnar, M. R., & Donzella, B. (2002). Social regulation of the cortisol levels in early human development. Psychoneuroendocrinology, 27, 199-220. doi:10.1016/S0306-4530(01)00045-2

Hackmann C, Green A, Notley C*, et al*.(2017). Protocol for a qualitative study exploring perspectives on the INternational CLassification of Diseases (11th revision); Using lived experience to improve mental health Diagnosis in NHS England: INCLUDE study. *BMJ Open*;**7:**e018399. doi: 10.1136/bmjopen-2017-018399

Henrica C. W. de Vet, Caroline B. Terwee, Lidwine B. Mokkink and Dirk L. Knol. (2011). Measurement in Medicine: A Practical Guide. Cambridge University Press 978-0-521-11820-0

Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from: *www.handbook.cochrane.org*

Kendall, S. et al., 2019. Acceptability and understanding of the Ages & Stages Questionnaires®, Third Edition, as part of the Healthy Child Programme 2-year health and development review in England: Parent and professional perspectives. *Child: care, health and development*, 45(2), pp.251–256.

Kenny, M., Conroy, S., Pariante, C. M., Seneviratne, G., & Pawlby, S. (2013). Mother-infant interaction in mother and baby unit patients: before and after treatment. Journal of Psychiatric Research, 47(9), 1192–1198

Krueger, R. A., & Leader, E. (2002). Designing and Conducting Focus Group Interviews. Retrieved from *https://www.eiu.edu/ihec/Krueger-FocusGroupInterviews.pdf*

Liberati, A. et al. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* , 339, p.b2700.

Leventhal, A., Jacobsen, T., Miller, L., & Quintana, E. (2004). Caregiving attitudes and at-risk maternal behavior among mothers with major mental illness. Psychiatric Services , 55(12), 1431–1433.

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.

McLeod, S. A. (2017). ***Attachment theory*.** Retrieved from *https://www.simplypsychology.org/attachment.html*

Mesman, J., & Emmen, R. A. G. (2013). Mary Ainsworth's legacy: a systematic review of observational instruments measuring parental sensitivity. *Attachment & Human Development*, *15*(5-6), 485–506.

Mokkink LB, Terwee CB, Patrick DL, et al.(2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient reported outcomes. J Clin Epidemiol. 63:737-745.

Mokkink, L.B., De Vet, H.C.W., Prinsen, C.A.C, Patrick, D.L., Alonso, J., Bouter, L.M., et al.(2017). COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Quality of Life Research. Retrieved from: https://www.cosmin.nl/wp-content/uploads/COSMIN_risk-of-bias-checklist_dec-2017.pdf

Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW.(2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Quality of Life Research. 19:539-549. doi: 10.1007/s11136-010-9606-8.

Mokkink LB, Prinsen CA, Patrick DL, et al. (2018). COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) - user manual. Available online: *http://www.cosmin.nl*.

Mokkink LB, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, De Vet HC, et al.(2019). COSMIN Study Design Checklist for Patient-Reported Outcome Measurement Instruments. Amsterdam, The Netherlands: COSMIN. Available from: https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing checklist_final.pdf#.

Mulrow, C.D. (1994). Rationale for Systematic Reviews. *British Medical Journal,* 309(6954), 597-599

NICE guidelines.(2015). Quality Standard 37 (QS37): Postnatal Care. Quality Statement 9: Emotional Well-being and infant attachment. [online] Available at: https://www.nice.org.uk/guidance/qs37/chapter/Quality- statement-9-Emotional-wellbeing-and-infant-attachment. Available at: https://www.nice.org.uk/guidance/ng26/chapter/Recommendations [Accessed May 16, 2019].

NHS digital. (2017). Mental Health of Children and Young People in England. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-of-children-and-young-people-in-england/2017/2017 [Accessed May 21, 2019].

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.

Potter, J and Wetherell, M. (1987). Discourse and social psychology: beyond attitudes and behaviour. Sage.

Pontoppidan, M. et al. (2017). Parent report measures of infant and toddler social-emotional development: a systematic review. *Family practice*, 34(2), pp.127–137.

Prinsen CA, Vohra S, Rose MR, et al. (2016). How to select outcome measurement instruments for outcomes included in a "Core outcome set"- a practical guideline. Trials. 17:449.

Prinsen CAC, Mokkink LB, Bouter LM, et al.(2018). COSMIN guideline for systematic reviews of patient reported outcome measures. Qual Life Res. 27:1147-1157

Rigby, J., Conroy, S., Miele-Norton, M., Pawlby, S., & Happé, F. (2016). Theory of mind as a predictor of maternal sensitivity in women with severe mental illness. Psychological Medicine, 46(9), 1853–1863.

Roggman, L. A., Cook, G. A., Innocenti, M. S., Norman, J., K, V., & Christiansen, K. (2009). PICCOLO (Parenting Interactions with Children: Checklist of Observations Linked to Outcomes) User's Guide. Logan, UT: Utah State University.

Sekhon, M., Cartwright, M., & Francis, J. J. (2017). Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. *BMC Health Services Research*, *17*(1), 88

Skovgaard, A. M., Olsen, E. M., Christiansen, E., Houmann, T., Landorph, S. L., Jorgensen, T., & The CCC 2000 study group (2008). Predictors (0–10 months) of psychopathology at age 1 ½ years—A general population study in The Copenhagen Child Cohort CCC 2000. Journal of Child Psychology & Psychiatry, 49, 553–562.

Stainton Rogers, W., Stainton Rogers, R. (1997). Does critical psychology mean the end of the world? In T. Ibanez & L. Iniguez (Eds), Critical social psychology (pp67-82). London: SAGE

Steadman, J., Pawlby, S., Mayers, A., Bucks, R. S., Gregoire, A., Miele-Norton, M., & Hogan, A. M. (2007). An exploratory study of the relationship between mother–infant interaction and maternal cognitive function in mothers with mental illness. Journal of Reproductive and Infant Psychology, 25(4), 255–269.

Stephenson, L. A., Macdonald, A. J. D., Seneviratne, G., Waites, F., & Pawlby, S. (2018). Mother and Baby Units matter: improved outcomes for both. BJPsych Open, 4(3), 119–125.

Stewart-Brown, S.L. and McMillan, A.S. (2010). Home and community-based parenting support programmes and interventions: report of Workpackage 2of the DataPrev Project. Warwick Medical School. [Online] Available at: http://wrap.warwick.ac.uk/3239/

Streiner, D. L. (2008). Qualitative Research in Psychiatry. *Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie*, *53*(3), 135–136.

Svanberg, P. O., & Jennings, T. (2002). . *"The Sunderland Infant Program (U.K.): Reflections on the first year." The Signal 9: 1-5. World Association of Infant Mental Health*. *9*, 1–5.

Svanberg, P.O., Mennet, L. & Spieker, S., (2010). Promoting a secure attachment: A primary prevention practice model. *Clinical child psychology and psychiatry*, 15(3), pp.363–378.

Svanberg, P. O., Barlow, J., & Tigbe, W. (2013). The Parent–Infant Interaction Observation Scale: reliability and validity of a screening tool. Journal of Reproductive and Infant Psychology, 31(1). *https://doi.org/10.1080/02646838.2012.751586*

Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 60(1):34-42.

Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. (2012).Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res. 21(4):651-7.

Terwee CB, Prinsen CA, Chiarotto A, et al. (2018). COSMIN methodology for assessing the content validity of PROMs - user manual. Available online: http://www.cosmin.nl.

Tong, A., Sainsbury, P. & Craig, J., 2007. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International journal for quality in health care: journal of the International Society for Quality in Health Care / ISQua*, 19(6), pp.349–357.

Tryphonopoulos, P.D., Letourneau, N. & Ditommaso, E. (2014). Attachment and caregiver-infant interaction: a review of observational-assessment tools. *Infant mental health journal*, 35(6), pp.642–656

Tuckett, A.G. 2005: Applying thematic analysis theory to practice: a researcher's experience. Contemporary Nurse 19, 75/87.

Ward, M. J., & Carlson, E. A. (1995). Associations among adult attachment representations, maternal sensitivity, and infant-mother attachment in a sample of adolescent mothers. *Child Development*, *66*(1), 69–79.

Widdicombe, S. and Wooffitt, R. (1995). The language of youth subcultures: social identity in action. Harvester Wheatsheaf.

WHO (World Health Organisation). (2018). Mental health. Child and adolescent mental health – WHO. Retrieved from: www.who.int>mental_health>maternal-child>child_

Wright, B. et al., (2015). Clinical effectiveness and cost-effectiveness of parenting interventions for children with severe attachment problems: a systematic review and meta-analysis. *Health technology assessment* , 19(52), pp.1–348.

References of Included studies in the systematic review

Azar, R., Paquette, D., Zoccolillo, M., Baltzer, F., & Tremblay, R. E. (2007). The association of major depression, conduct disorder, and maternal overcontrol with a failure to show a cortisol buffered response in 4-month-old infants of teenage mothers. Biological Psychiatry, 62(6), 573–579.

Barlow, J., Davis, H., McIntosh, E., Jarrett, P., Mockford, C., & Stewart-Brown, S. (2007). Role of home visiting in improving parenting and health in families at risk of abuse and neglect: results of a multicentre randomised controlled trial and economic evaluation. Archives of Disease in Childhood, 92(3), 229–233.

Barlow, J., Sembi, S., & Underdown, A. (2016). Pilot RCT of the use of video interactive guidance with preterm babies. Journal of Reproductive and Infant Psychology, 34(5), 511–524.

Binda, V., Figueroa-Leigh, F., & Olhaberry, M. (2019). Antenatal and postnatal depressive symptoms: Association with quality of mother–infant interaction. Infant Behavior & Development, 57, 101386.

Borghini, A., Habersaat, S., Forcada-Guex, M., Nessi, J., Pierrehumbert, B., Ansermet, F., & Müller-Nix, C. (2014). Effects of an early intervention on maternal post-traumatic stress symptoms and the quality of mother-infant interaction: the case of preterm birth. *Infant Behavior & Development*, *37*(4), 624–631.

Cameron, J., & Shepherd, A. L. (2018). Evaluation of outcomes from an evidence-based programme for mothers and babies. Journal of Health Visiting, 6(10), 499–505.

Conroy, S., Marks, M. N., Schacht, R., Davies, H. A., & Moran, P. (2010). The impact of maternal depression and personality disorder on early infant care. Social Psychiatry and Psychiatric Epidemiology, 45(3), 285–292.

Crugnola, R, C,. Ierardi E, Albizzati A and Downing G (2016) Effectiveness of an Attachment-Based Intervention Program in Promoting Emotion Regulation and Attachment in Adolescent Mothers and their Infants: A Pilot Study. Front. Psychol. 7:195. doi: 10.3389/fpsyg.2016.00195

Emery, J., Paquette, J. & Bigras, M. (2008) Factors predicting attachment patterns in infants of adolescent mothers, Journal of Family Studies, 14:1, 65-90, DOI: 10.5172/jfs.327.14.1.65

Flykt, M., Kanninen, K., Sinkkonen, J., & Punamaki, R.-L. (2010). Maternal Depression and Dyadic Interaction: The role of Maternal Attachment Style. Infant and Child Development, 19(5), 530–550.

Fuertes, M., Beeghly, M., dos Santos, P. L., & Tronick, E. (2011). Predictors of infant positive, negative and self-direct coping during face to face still-face in a Portuguese preterm sample. Análise Psicológica, 29(4), 553–565.

Fuertes, M., Faria, A., Beeghly, M., & Lopes-dos-Santos, P. (2016). The effects of parental sensitivity and involvement in caregiving on mother-infant and father-infant attachment in a Portuguese sample. Journal of Family Psychology: JFP: Journal of the Division of Family Psychology of the American Psychological Association , 30(1), 147–156.

Fuertes, M., Faria, A., Soares, H., & Crittenden, P. (2009). Developmental and evolutionary assumptions in a study about the impact of premature birth and low income on mother–infant interaction. Acta Ethologica, 12(1), 1–11.

Fuertes, M., Lopes-dos-Santos, P., Beeghly, M., & Tronick, E. (2009). Infant Coping and Maternal Interactive Behavior Predict Attachment in a Portuguese Sample of Healthy Preterm Infants. European Psychologist, 14(4), 320–331.

Forcada-Guex, M., Borghini, A., Pierrehumbert, B., Ansermet, F., & Muller-Nix, C. (2011). Prematurity, maternal posttraumatic stress and consequences on the mother-infant relationship. Early Human Development, 87(1), 21–26.

Hofer, T., Hohenberger, A., Hauf, P., & Aschersleben, G. (2008). The link between maternal interaction style and infant action understanding. Infant Behavior & Development, 31(1), 115–126.

Hohenberger, A., Elsabbagh, M., Serres, J., de Schoenen, S., Karmiloff-Smith, A., & Aschersleben, G. (2012). Understanding goal-directed human actions and physical causality: the role of mother-infant interaction. Infant Behavior & Development, 35(4), 898–911.

Karmiloff-Smith, A., Aschersleben, G., de Schonen, S., Elsabbagh, M., Hohenberger, A., & Serres, J. (2010). Constraints on the Timing of Infant Cognitive Change: Domain-Specific or Domain-General? International Journal of Developmental Science, 4(1), 31–45.

Kenny, M., Conroy, S., Pariante, C. M., Seneviratne, G., & Pawlby, S. (2013). Mother-infant interaction in mother and baby unit patients: before and after treatment. Journal of Psychiatric Research, 47(9), 1192–1198.

Komoto, K., Hirose, T., Omori, T., Takeo, N., Okamitsu, M., Okubo, N., & Okawa, H. (2015). Effect of Early Intervention to Promote Mother - Infant Interaction and Maternal Sensitivity in Japan: A Parenting Support Program based on Infant Mental Health. Journal of Medical and Dental Sciences, 62(4), 77–89.

Kristensen, I. H., Simonsen, M., Trillingsgaard, T., & Kronborg, H. (2017). Video feedback promotes relations between infants and vulnerable first-time mothers: a quasi-experimental study. BMC Pregnancy and Childbirth, 17(1), 379.

Letourneau, N. L., Kozyrskyj, A. L., Cosic, N., Ntanda, H. N., Anis, L., Hart, M. J., Campbell, T. S., Giesbrecht, G. F., & APrON Team. (2017). Maternal sensitivity and social support protect against childhood atopic dermatitis. Allergy, Asthma, and

Clinical Immunology: Official Journal of the Canadian Society of Allergy and Clinical Immunology, 13, 26.

Neri, E., Agostini, F., Perricone, G., Morales, M. R., Biasini, A., Monti, F., & Polizzi, C. (2017). Mother- and father-infant interactions at 3 months of corrected age: The effect of severity of preterm birth. Infant Behavior & Development, 49, 97–103.

Neuhauser, A. (2018). Predictors of maternal sensitivity in at-risk families. Early Child Development and Care, 188(2), 126–142.

O'Shaughnessy, R., Nelki, J., Chiumento, A., Hassan, A., & Rahman, A. (2012). Sweet Mother: evaluation of a pilot mental health service for asylum-seeking mothers and babies. Journal of Public Mental Health, 11(4), 214–228.

Paquette, D., Zoccolillo, M., Bigras, M., Labelle, M, E., Azar, R. & Emery, J. (2004). A Comparative Study of Parental Sensitivity Between Three Groups of Adolescent Mothers, Journal of Emotional Abuse, 4:1, 85-104, DOI: 10.1300/J135v04n01_05

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.

Pillhofer, M., Spangler, G., Bovenschen, I., Kuenster, A. K., Gabler, S., Fallon, B., Fegert, J. M., & Ziegenhain, U. (2015). Pilot study of a program delivered within the regular service system in Germany: effect of a short-term attachment-based intervention on maternal sensitivity in mothers at risk for child abuse and neglect. Child Abuse & Neglect, 42, 163–173.

Rigby, J., Conroy, S., Miele-Norton, M., Pawlby, S., & Happé, F. (2016). Theory of mind as a predictor of maternal sensitivity in women with severe mental illness. Psychological Medicine, 46(9), 1853–1863.

Sidor, A., Kunz, E., Schweyer, D., Eickhorst, A., & Cierpka, M. (2011). Links between maternal postpartum depressive symptoms, maternal distress, infant gender and sensitivity in a high-risk population. Child and Adolescent Psychiatry and Mental Health, 5(1), 7.

Sidor, A., Fischer, C., & Cierpka, M. (2015). The effects of the early prevention program "Nobody Slips Through the Net": A longitudinal controlled study in an at-risk sample. Mental Health & Prevention, 3(3), 103–116.

Soares, H., Barbieri-Figueiredo, M., Pereira, S., Silva, M., & Fuertes, M. (2018). Parents attending to nurse visits and birth age contribute to infant development: A study about the determinants of infant development. Early Human Development, 122, 15–21.

Steadman, J., Pawlby, S., Mayers, A., Bucks, R. S., Gregoire, A., Miele-Norton, M., & Hogan, A. M. (2007). An exploratory study of the relationship between mother–infant interaction and maternal cognitive function in mothers with mental illness. Journal of Reproductive and Infant Psychology, 25(4), 255–269.

Stephenson, L. A., Macdonald, A. J. D., Seneviratne, G., Waites, F., & Pawlby, S. (2018). Mother and Baby Units matter: improved outcomes for both. BJPsych Open, 4(3), 119–125.

Svanberg, P. O., Mennet, L., & Spieker, S. (2010). Promoting a secure attachment: A primary prevention practice model. Clinical Child Psychology and Psychiatry, 15(3), 363–378.

Tester-Jones, M., Karl, A., Watkins, E. and O'Mahen, H. (2017), Rumination in dysphoric mothers negatively affects mother–infant interactions. J Child Psychol Psychiatr, 58: 38-45. doi:10.1111/jcpp.12633

Udry-Jørgensen, L., Pierrehumbert, B., Borghini, A., Habersaat, S., Forcada-Guex, M., Ansermet, F., & Muller-Nix, C. (2011). Quality of attachment, perinatal risk, and mother-infant interaction in a high-risk premature sample. Infant Mental Health Journal, 32(3), 305–318.

Udry-Jørgensen, L., Tissot, H., Frascarolo, F., Despland, J.-N., & Favez, N. (2016). Are parents doing better when they are together? A study on the association between parental sensitivity and family-level processes. Early Child Development and Care, 186(6), 915–926.

Ward, M. J., & Carlson, E. A. (1995). Associations among adult attachment representations, maternal sensitivity, and infant-mother attachment in a sample of adolescent mothers. Child Development, 66(1), 69–79.

Underdown, A., Norwood, R., & Barlow, J. (2013). A realist evaluation of the processes and outcomes of infant massage programs: Evaluation of Infant Massage Programs. Infant Mental Health Journal, 34(6), 483–495.

Zwönitzer, A., Ziegenhain, U., Bovenschen, I., Bressem, K., Pillhofer, M., Fegert, J. M., Spangler, G., Gerlach, J., Gabler, S., & Künster, A. K. (2015). Effects of early intervention in children at risk: Short-term and long-term findings from an attachment-based intervention program. Mental Health & Prevention, 3(3), 98–102.