

“You are our eyes and ears”: A new tool for observing parent-child interactions in large samples

Elaine Reese University of Auckland and University of Otago, New Zealand
Amy L Bird University of Auckland, New Zealand
a.bird@auckland.ac.nz
Mele Taumoepeau University of Otago, New Zealand
Joanna M Schmidt University of Waikato, New Zealand
Jatender Mohal University of Auckland, New Zealand
Cameron C Grant University of Auckland and Auckland District Health Board New Zealand
Polly E Atatoa Carr University of Auckland and University of Waikato, New Zealand
Susan M B Morton University of Auckland, New Zealand

Received September 2015 Revised May 2016)

<http://dx.doi.org/10.14301/llcs.v7i4.381>

Keywords

Parent-child interaction, observation, longitudinal research

Abstract

Differences in parent-child interactions have implications for a range of developmental outcomes that are of interest to large longitudinal cohort studies. We describe a new method for observing parent-child conversations specifically designed to be a component of a more comprehensive collection of data about child health and development. Participants were mothers and their two-year-old children who were part of the Growing Up in New Zealand study. During a series of brief, prompted parent-child conversations, observers were trained to rate mothers' warmth, use of open-ended questions, talk about emotions and 'linking' talk, children's emotional expression, and mothers' overall use of discipline. Reliability was established before and reviewed mid-way through the one-year data collection wave. We observed differences in parent-child interaction ratings as a function of socio-demographic variables, ethnicity, and child gender that were in agreement with published research. Inter-scale correlations and correlations between observer ratings and maternal self-report measures provide preliminary evidence of convergent and discriminant validity. Specifically, higher maternal self-reported affiliation and more frequent book reading were significantly correlated with observer ratings of maternal warmth, maternal language style, and children's emotional expression; and negatively correlated with observer ratings of maternal discipline. Higher maternal self-reported parenting hostility was negatively correlated with observed maternal warmth and language; and positively correlated with observed maternal discipline. This observational method is a potentially useful technique for obtaining independent measures of parent-child conversational interactions during the preschool years within large cohort studies.

Introduction

Large, transdisciplinary longitudinal studies improve our understanding of development across the lifespan (Wadsworth & DeZateux, 2013). Collaboration across disciplines in the design and development of such studies is necessary if they are to increase our understanding of child development (Lu, 2014). This type of transdisciplinary research encourages researchers to incorporate knowledge and skills from other fields, and hence requires the development of new methodologies. The capacity to span multiple domains is one of the strengths of large cohort studies, yet inevitably limits the depth to which each individual construct can be measured. We outline here a new method specifically developed for observing parent-child interactions within a large child cohort study.

Observing parent-child interactions in large cohort studies

Developmental psychologists and child development researchers have long been interested in how parents’ day-to-day behaviour with their children impacts on social, emotional, and behavioural development. The quality of parent-child interactions has been implicated in developmental outcomes ranging from psychopathology, to externalising behaviour, to educational attainment (Locke & Prinz, 2002). Parents’ reports of their own parenting behaviours can be extremely informative, but for some behaviours, unbiased observations are needed (Dunn & Kendrick, 1980). For instance, it may be difficult for parents to estimate how often they smile at their children, because this behaviour is largely outside their awareness, or highly influenced by their own emotional state, their beliefs about the child, and about expectations for that behaviour in general (Gardner, 2000; Margolin et al., 1998).

Recognising the importance and value of parent-child interaction, a few large cohort studies have successfully incorporated direct observations. The Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) has incorporated parent-child play (‘Two Bags Task’) and book-reading tasks during the preschool and kindergarten data collection waves. Over 6,000 Two Bags Tasks interactions were coded from recordings at both timepoints (Najarian, Snow, Lennon, Kinsey & Mulligan, 2010), while a subsample

(700) of the book-reading tasks were coded from recordings (Hindman, Skibbe & Foster, 2014). Having detailed coding from recordings of parent-child interactions for such large samples is ideal, but may not be financially feasible for all cohort studies.

The Home Observation Measure of Environment scale (HOME) (Bradley & Caldwell, 1984) has also been pivotal in demonstrating the feasibility and value of observing children’s home environments within large cohort studies. The HOME scale is a well-validated measure of the quality of the home environment, in particular the learning environment, environmental stimulation, and parent-child interaction. The HOME scale has been instrumental in identifying associations between aspects of the home environment and developmental outcomes for both typical and atypical populations (Totsika & Sylva, 2004). The quality of the home environment – as measured by the HOME scale – has been associated with a range of important developmental outcomes such as attachment security (NICHD Early Child Care Research Network, 2001) and obesity (Strauss & Knight, 1999). The strongest associations across time have been found with cognitive and academic outcomes (Bradley et al., 1989). The development of a short-form has enabled the HOME scale to be more widely used by large cohort studies, such as the National Longitudinal Surveys of Youth. The HOME scale is succinctly described by Mott (2004, p. 260) as: “a psychometric hybrid. It represents an attempt – in my opinion, a very successful one – to integrate a psychological assessment into a large-scale data collection.”

The huge contribution and strong psychometric properties of the HOME scale therefore begs the question: do large cohort studies *need* any other measure of the home environment? We present a task developed here for use in the *Growing Up in New Zealand* cohort when children were age two that uses a prompted task to elicit parent-child conversation, and focuses more specifically on the content of the verbal interaction, as well as the non-verbal context. Based on extensive research of adult-child book-reading (e.g., Fletcher & Reese, 2005; Haden, Reese & Fivush, 1996; Reese & Cox, 1999) and past-event conversations (e.g., Bird & Reese, 2006; Farrant &

Reese, 2000; Reese, Haden & Fivush, 1993; Reese & Newcombe, 2007), we use photo prompts to elicit parent-child conversation and observe the verbal and non-verbal quality of communication.

Why a semi-structured task?

One of the key advantages to the HOME scale is that no specific tasks are required and interviewers can code behaviours and materials as they occur in their natural environment. This has clear benefits for multi-domain home assessments, but some limitations, particularly with preschoolers. Young children may be present for only some, or none, of the home visit with parent(s) and consequently completion rates and reliability estimates are lower for children under three years compared with older children (Mott, 2004).

In contrast, semi-structured methods require parents and children to engage in a particular task designed to tap into underlying constructs more efficiently. Examples from child development research range from providing the dyad with a standard set of toys or a book to elicit play or reading style (e.g., Fuligni & Brooks-Gunn, 2013; Hindman, Skibbe & Foster, 2014), to inviting the dyad to discuss specific topics (e.g., Fivush, Berlin, Sales, Menutti-Washburn & Cassidy, 2003; cf. Margolin et al., 1998), to lab procedures such as the Strange Situation (e.g., Ainsworth, Blehar, Waters, & Wall, 1978). Shorter observations are not necessarily inferior. For example, Lovejoy, Graczyk, O'Hare and Neuman (2000) showed that, in comparison with longer observations, brief (< 10-min) observations provided larger effect sizes for differences in the positive behaviours shown by depressed and non-depressed mothers' in their interactions with their children. Both unstructured and semi-structured observational methods can provide reliable measures of parent-child interaction with concurrent long-term and predictive validity of children's development (e.g., Bird & Reese, 2006; Dunn, Brown, & Beardsall, 1991; Sroufe, Egeland, Carlson, & Collins, 2005; Taumoepeau & Ruffman, 2006). We hoped that by using a semi-structured picture discussion task we would be able to engage a higher proportion of our two-year-old cohort, while also eliciting important maternal language and parent-child interaction indicators.

Why focus in more depth on the quality of parent-child conversation?

Talking with one another is an integral part of human nature; it is also a key medium through which children develop relationships and learn about language, other people, their culture, the world, and themselves (Bowlby, 1969; Brockmeier & Carbaugh, 2001; Miller, Mintz, Hoogstra, Fug & Potts, 1992). Exposure to language through adult dialogue, narration of activities, and book reading is recognised as critical for children's language and cognitive development (Bornstein & Haynes, 1998; Huttenlocher, Haight, Bryk, Seltzer & Lyons, 1991; Huttenlocher, 1998). The preschool HOME scale measures the general home language environment well; it taps into the frequency (e.g., spontaneously talks to child at least twice) and type (e.g., parent talks with interviewers, names an object or person for child) of parental talk across the home visit.

Yet, while total language exposure is clearly important, not all communication is created equal: interactive conversations between parents and children seem particularly important. For example, Zimmerman and colleagues examined associations of total adult language exposure, adult-child interactive conversations, and total television viewing with two-four year-old children's language development. While all were associated, only adult-child interactive conversations were uniquely predictive of children's language 18 months later (Zimmerman, Gilkerson, Richards, Christakis, Xu, Gray & Yapanel, 2009). These findings highlight the importance of more detailed measurement of parent-child conversations, beyond global indicators of parental speech.

During infancy most parent talk occurs in the 'here and now', but by the age of two children are engaging in discussions that extend both temporally and contextually: talk about the past, the future, other people, and the hypothetical (Snow, 1991). Examples of such decontextualized talk could include: talking about a past trip to the zoo while reading a book about animals; building a plane out of Lego and talking about a family holiday you would like to take one day; or even talking about the day at childcare while driving home in the car.

Naturalistic observations indicate that mothers of two year olds talk about up to six past-events per hour, and two and a half year olds themselves

contribute to, on average, two past-event discussions per hour (Miller & Sperry, 1988). This ‘there-and-then’ talk appears particularly important for children’s cognitive, social and emotional development. For example, parents who guide their children to consider new information by using open-ended questions and discuss emotional or evaluative content have children with more advanced autobiographical memory (Farrant & Reese, 2000; Reese, Haden & Fivush, 1993; Reese & Newcombe, 2007). When children are engaged in conversations with their parents about past events that highlight and explain the child’s internal states and emotions, they are more likely to display a coherent self-concept and higher self-esteem (Bird & Reese, 2006; Reese, Bird & Tripp, 2007; Welch-Ross, 1997).

Book-reading is another key medium through which children are exposed to varying levels of decontextualised talk. Overall, parents who describe pictures, discuss story meaning, and use the book as a link to other comments or conversations have children with more advanced language and literacy outcomes than parents who focus more on ‘reading the words’ (Haden, Reese & Fivush, 1996; Reese & Cox, 1999). Interestingly, mothers’ talk about internal states during a picture description task at age 15 months was associated with children’s social and emotional understanding at two years of age (Taumoepeau & Ruffman, 2006).

The broader context of parent-child interactions

While the specific verbal content of parent-child interactions is increasingly seen as important, the broader non-verbal context of the parent-child interaction is also critical. Parental discipline and warmth – both key constructs measured by the HOME scale – appear central. Differences in parental discipline are associated with children’s externalising behaviour difficulties and academic achievement difficulties (see Locke & Prinz, 2002, for a review). At a broad level, discipline might encompass many different practices (e.g., from verbal correction, to use of behavioural techniques such as ‘time-out’, to shouting or smacking); or be related to the degree of control in parenting styles (e.g., an authoritarian style characterised by high control and low warmth; Baumrind, 1967). Similar to the HOME scale, we focus here on specific unhelpful discipline behaviours that

relate negatively with child outcomes, such as shouting, visible hostility or physical hitting (Bradley, Corwyn, Burchinal, McAdoo & Garcia Coll, 2001).

Parental warmth – defined as “the expression of positive affect, affection, and admiration towards the child” – is theoretically and empirically related to several other key aspects of parent-child relatedness, such as parental sensitivity and responsivity, which have their origins in attachment theory (Ainsworth et al., 1978; Bowlby, 1969). Measurement of parental warmth typically includes visible positive affect and animation when talking with the child, as well as physical affection and closeness. Parental warmth is an important moderator or protective factor for a range of developmental outcomes (Rutter, 2013). For example, maternal warmth has been found to moderate the relationship between low birth weight and ADHD (Tully, Arseneault, Caspi, Moffitt & Morgan, 2004), and between peer bullying and internalising and externalising problems (Bowes, Maughan, Caspi, Moffitt & Arseneault, 2010).

From the child’s perspective, understanding and expressing emotion is a crucial developmental task, and underlies healthy functioning across multiple domains (Gross, 1998). The parent-child relationship is a fundamental context through which children develop emotional expression (Grusec, 2011): first through basic cries, and later through more complex facial, behavioural and verbal communication (Malatesta & Wilson, 1988). Children who are able to express social emotions such as empathy tend to exhibit more prosocial behaviour (Roberts & Strayer, 1996) and may elicit more positive social responses from both parents and peers (Findlay, Girardi & Coplan, 2006). Conversely, low empathy has been linked with children’s externalising behaviour and social difficulties (de Wied, Gispens-de Wied & van Boxtel, 2010); and deficits in emotional understanding and expression underlie most forms of psychopathology (Aldao, Nolen-Hoeksema & Schweizer, 2010).

Across all of these domains of parent-child interaction – parental discipline, warmth, and verbal communication – reliable differences have been observed in parent-child interactions as a function of parents’ ethnicity, educational achievement, and socioeconomic status. Parents from cultures with an independent orientation (e.g., most European

cultures) tend to engage in more distal (talking) and fewer proximal (touching, smiling) behaviours compared to parents from interdependent cultures (e.g., most non-European cultures; Bornstein et al., 1992). Parents with higher educational achievement and socioeconomic status also tend to talk more, and to be less harsh in their disciplinary practices (Bornstein & Bradley, 2014; Hart & Risley, 1995; Jansen et al., 2012).

Establishing Inter-Observer Agreement in Very Large Samples

A crucial issue for cohort studies that employ a large team of interviewers is the establishment of inter-observer agreement or reliability. Most developmental research studies involve, at most, several hundred participants. The ‘gold standard’ of reliability measurement for parent-child interaction was developed in this context: two independent observers evaluate the same behaviour from recordings of the original interaction, and inter-rater reliability is established on a subset of the sample. This gold-standard inter-rater reliability procedure is not always pragmatically possible in the context of a large cohort study.

Motivated by a desire to observe parent-child interactions for our full sample without the means to record and then code, we sought to identify other methods of establishing inter-rater reliability. We identified these from research conducted in large educational settings, where similar challenges are encountered but for a different reason – namely, the busyness of the classroom environment which results in video recordings failing to capture the behaviours of interest. To overcome this issue in the classroom setting, Coffman, Ornstein, McCall, and Curran (2008) trained all observers prior to going into the field to look for specific teacher behaviours within 30-second intervals. Inter-rater reliability was established by watching pre-prepared video clips of teacher behaviour and comparing ratings with those of an expert coder.

Similar time-sampling techniques were used in research conducted prior to the 1970s, before filming techniques had become widespread in child development research (see Lytton, 1971 for a historical review). For example, Rheingold (1960) used this method to observe maternal care of very

young infants. Observers were given a checklist of 42 maternal and infant behaviours to observe and tally during set time periods. Again, inter-observer reliability was established prior to the commencement of the study: observers were paired and asked to simultaneously rate maternal and infant behaviours on seven dyads who were not part of the larger study. This procedure was repeated several times during the study to prevent observer drift.

Development of a new observational measure within a child cohort study

We developed a new observational measure of parent-child interaction, the Parent-Child Interaction task, ideally suited for inclusion in studies with larger sample sizes. This measure was designed for the *Growing Up in New Zealand* study and completed with two-year-old children and their mothers (Morton et al., 2013). Ideally, we would have repeated the observation with fathers, but for this first step we focused only on mothers as the primary caregivers for most of the children. Given the practical limitations of recording and observing thousands of parent-child interactions, observer reliability was established prior to going into the field (Coffman et al., 2008).

We asked mothers and children to engage in a semi-structured shared discussion task. The observers were the group of interviewers employed to complete the face-to-face interview with the mother of each cohort child when the child was two years old. We used time-sampling techniques to train the observers to observe one behaviour at a time for 30-second intervals. The constructs were chosen for their prominence in the child development literature as outlined above: maternal discipline, maternal warmth, maternal verbal communication (open-ended questions, linking, and emotion talk), and children’s emotional expression.

Sample diversity was an essential design feature of the *Growing Up in New Zealand* cohort (Morton et al., 2013). We were interested in how parent-child interaction constructs differed across our diverse sample as a function of ethnicity, child gender, maternal education and socio-demographic status. In order to examine construct validity, we examined associations between *observations* of parent-child interactions and parent *self-report* of parenting

warmth and hostility, parent-child affiliation, and frequency of book reading and oral story-telling. We predicted that observed maternal warmth and language constructs would be positively associated with self-reported maternal warmth and affiliation, and negatively associated with maternal hostility. Similarly, we predicted that observed maternal discipline would be positively associated with maternal hostility, and negatively associated with observed maternal warmth, affiliation, and language variables. We also predicted that mothers who reported more frequent oral story-telling and book reading interactions with their children would be observed to use more of all aspects of verbal communication: more open-ended questions, linking talk, and emotion references.

Method

Cohort study participants

Participants were members of New Zealand’s longitudinal pre-birth cohort study: *Growing Up in New Zealand*. Analyses here refer to 5,536 two-year-old children and their mothers who completed the Parent-Child Interaction task. This represented 88% of the 6,327 children who took part in the two-year data collection wave and 81% of the original 6,853 children in the cohort (Morton et al., 2013). Women were recruited during pregnancy from a geographically defined region of New Zealand chosen for its population diversity. All pregnant women who resided within this region and who had an estimated due date between 25th April 2009 and 25th March 2010 were eligible. A multi-faceted recruitment strategy was utilised with the goal of recruiting a sample broadly generalisable to the contemporary New Zealand national birth cohort (Morton et al., 2014a). Alignment of the enrolled cohort with the national birth cohort was confirmed (Morton et al.,

2014b). Ethical approval was obtained from the Ministry of Health Ethics Committee. Written informed consent was obtained from all participating women.






Data collection procedure with the cohort

The two-year computer-assisted interview was conducted face-to-face in the child’s home. Mothers were asked a range of questions across multiple domains (health, psychosocial and cognitive development, family and whanau, education, culture and identity, and neighbourhood and societal context; see Morton et al., 2013). A series of child observation tasks were completed part-way through the interview, including the Parent-Child Interaction taskⁱ.

Parent-Child Interaction task description

Mothers were given a series of five photographs in the same order. These photographs were specifically selected to elicit the parent-child behaviours of interest (see table 1). Mothers were asked to describe each picture to their child as if they were telling a story. After 30 seconds the interviewer retrieved the picture, recorded his or her rating of the target behaviour, and handed the parent the next picture. Each picture and 30-second interval corresponded to a specific construct (e.g., the picture of a mother and child walking in a forest was coded for maternal warmth), with the exception of maternal discipline, which was rated across all five pictures. If children did not engage in the task, interviewers recorded the reason from the following response options: child was asleep; child was unwell; child not in the home; child did not engage; mother refused; child has a physical injury; child has a physical disability; child has a developmental delay; the interview was interpreted; or other.

Table 1. Parent-child interaction task: pictures, constructs and coding instructions

Picture	Photo	Parent-child Interaction Construct	Focus	Behaviour Rating
Mother and child in forest		Maternal warmth	Look	<p>1 = No emotional expression Mother may be engaged with child and talking but is not smiling or laughing during the coding period.</p> <p>2 = Smile only Mother gives a smile, however slight, at any point during the coding period.</p> <p>3 = Laugh OR cuddle OR kiss Mother laughs at any point during the coding period, or cuddles / embraces the child with one or both arms, or she kisses the child.</p>
Children washing the car		Maternal questions	Listen	<p>1 = No questions</p> <p>2 = One question only</p> <p>3 = Two or more questions</p> <p>Once two open-ended questions have been counted, record response as 3 and stop coding. Note, an open-ended question requires more than a “yes” or “no” and contains a <i>What, Where, Who, How, Why or When</i>.</p>
Child reaching up for Dad		Maternal desire or emotion words	Listen	<p>1 = No desire or emotion words</p> <p>2 = One desire or emotion word</p> <p>3 = Two or more desire or emotion words</p> <p>If two desire/emotion words are counted, record response as 3 and stop coding.</p> <p>Desire or emotion words include <i>want, like, don't like, don't want, hate, happy, sad, scared, grumpy, excited</i>. Not included are words such as <i>crying, hungry or tired</i> because these are about physiological (or physical) rather than emotional states.</p>
Child crying		Children's emotional expression (empathy)	Look	<p>1 = Absence of concern or empathy</p> <p>2 = Presence of concern or empathy (child appears even mildly distressed, sad or concerned. For example, furrowed brow, raised eyebrows, downturned mouth)</p> <p>Once empathy or concern is observed, record response as 2 and stop coding. The most reliable indicator of concern/empathy is a furrowed brow, which may or may not be accompanied by a downturned mouth. Note, the target is empathy so does not include other emotional expressions (e.g., smile) or words.</p>
Child stacks blocks		Maternal linking	Listen	<p>1 = Description of picture, but no link to child's own experience/world.</p> <p>2 = Link to child's own experience/world.</p> <p>If mother makes a link to child's experience straight away, record response as 2 and stop coding. Maternal linking occurs when the mother connects or links the task to the child's own experiences or world. One example of linking includes talking about the child's own block play (<i>You built a really tall tower like that yesterday</i>). Note, children had completed a Stack and Topple task earlier in the <i>Growing Up</i> interview, so there was an opportunity for all parents to link to recent block play. Another example could include a link to a sibling or relative (<i>That boy looks like your cousin</i>).</p>
Across all 5 photos		Maternal discipline	Listen and look	<p>1 = No discipline/behavioural correction</p> <p>2 = One instance of mild behavioural correction (<i>Don't sit like that; Stop it; gentle shove; slightly stern look</i>)</p> <p>3 = Harsh discipline (<i>yelling, smacking, hard shove/pull, evil eye, cursing at child</i>) OR more than one mild behavioural correction</p> <p>If even one instance of harsh discipline occurs, record as 3 and stop coding. Discipline is defined here as any instance of correcting a child's <i>behaviour</i>, either verbal or nonverbal (i.e. <i>Don't sit like that; Stop that; or pulling/shoving child into place; or stern look</i>). Corrections of a child's response (<i>No, I don't think that's a potato; it's a kumara</i>) were not included. Note that this only relates to discipline during the interaction exercise, not to anything that occurs outside of this exercise.</p>

Observer training in administration of the Parent-Child Interaction task

Because observers would be coding in the field rather than from video recordings, inter-rater reliability needed to be established *before* the two-year interviews were conducted. Face-to-face training events were organised for the observers to establish reliabilityⁱⁱ.

Two expert coders (ER and MT) prepared video clips for reliability training from an observational study of toddlers’ emotional development with New Zealand parents (Taumoepeau & Ruffman, 2006). The mothers in the clips were either European or Pacific; all had a two-year-old child. Videos showed mothers interacting with their toddlers in a parent-child interaction taskⁱⁱⁱ.

All 33 interviewers attended a training day (there were two separate training days for interviewers in different geographical locations). The expert coders first introduced the task to interviewers using a power-point presentation. We explained the reasons for observation, and the way that we were adapting traditional methods for the *Growing Up in New Zealand* study. We told the interviewers “you are our eyes and ears” because we would not be able to directly observe the mother-child interactions. We then introduced each target dimension to the interviewers, noting the decision points (table 1) and emphasising that for each dimension, the interviewer was only going to be looking or listening for a single type of behaviour. We believe that this focus on a single behaviour at a time is one reason we were able to succeed in establishing reliability.

Interviewers then watched the videos of mothers interacting with their toddler children and practiced coding for the constructs of interest. Feedback was given and discrepancies explained. Interviewers were then shown six video clips corresponding to the six constructs of interest. This process was repeated four more times with different videos, resulting in interviewers viewing a total of 30 different video clips of parent-child interaction (five for each of the six constructs). It should be noted that the training from video may have actually required a higher threshold than in vivo, as many of the disagreements in coding were due to difficulties in hearing or seeing important cues on the video. Interviewer agreement was calculated by dividing the number of agreements

across constructs by the total number of observations. Mean interviewer agreement was 86%, range 73.3%-96.7%.

Reliability check at midpoint in data collection wave

Six months into the 12-month long data collection wave when children were aged two, measurement of the reliability of the interviewer coding was repeated^{iv}. Mothers and their preschool children were recruited from a database of primarily European families. Interviewers were shown eight video clips for each of the six constructs, resulting in a total of 48 different video clips. Mean interviewer agreement was 86%, range 73.3%-90%.

Maternal self-report measures of interactions with their child

Mothers were asked a series of questions about their interactions with, and feelings about, their children. The Time Spent with Child Scale (Davies, Harold, Goeke-Morey & Cummings, 2002) was administered as a measure of parent-child affiliation (e.g., ‘I enjoy having my child around me’, ‘I tell my child how proud I am of them when he/she is good’). Parents indicated their responses to each question on a 1-4 scale. Item responses were totalled to give an overall parental affiliation score. The Warmth/Hostility Scale (derived from the Iowa Family Interaction Rating Scale; Melby et al., 1989-1993) contained nine questions with a 1-7 response scale. Responses to five of the items were added to give a parenting warmth total, and responses to the remaining four items were added to give a parenting hostility total. Mothers were also asked to indicate on a 1-5 scale how frequently they read books with their child and how often they told stories with their child. These last two items were designed specifically for the current study and were derived from research demonstrating links between the home literacy environment and children’s language development (Payne, Whitehurst & Angell, 1994). Full questionnaires with scales and items can be found at www.growingup.co.nz.

Socio-demographic measures and ethnicity classification

Women were asked a range of standard demographic questions at the antenatal interview. Area-level

socioeconomic deprivation was measured using the NZ Index of Deprivation (Salmond, Crampton & Atkinson, 2007). Maternal education was grouped into the following categories based on highest qualifications: no formal qualifications; secondary school; trade or university qualification. Women were asked to self-prioritise their ethnicity and responses were grouped into the following categories: European, Māori, Pacific, Asian and Other.

Data Analysis

Ordinal regression models were developed to test associations between socio-demographic variables and three-level parent-child interaction variables (maternal warmth, maternal open-ended questions, maternal use of emotion language and maternal discipline). Logistic regression models were

developed to test associations between socio-demographic variables and categorical parent-child interaction variables (maternal linking language and children’s empathy expression). Spearman correlation coefficients were used to examine associations between observed parent-child interactions and self-report measures.

Results

Descriptive statistics

Descriptive statistics were calculated for each parent-child interaction construct and are shown in table 2. There was a good spread of scores across the scales for all but the discipline dimension, for which most parents were scored as administering no discipline across all five photos.

Table 2. Number and proportion of sample receiving each Parent-Child Interaction construct rating

Parent-Child Interaction Constructs	Number (%) parents N = 5536
Maternal warmth	
1 = No emotional expression	425 (8)
2 = Smile only	1518 (29)
3 = Laugh OR cuddle OR kiss	3378 (63)
4 = Did not engage in task	17 (0)
Maternal (open-ended) questions	
1 = No questions	1031 (19)
2 = One question only	1172 (22)
3 = Two or more questions	3092 (58)
4 = Did not engage in task	43 (1)
Maternal emotion or desire words	
1 = No desire or emotion words	1628 (31)
2 = One desire or emotion word	1622 (30)
3 = Two or more desire or emotion words	1992 (37)
4 = Did not engage in task	96 (2)
Child’s emotional expression / empathy	
1 = Absence of concern or empathy	2995 (56)
2 = Presence of concern or empathy	2276 (43)
4 = Did not engage in task	67 (1)
Maternal linking	
1 = Description of picture, but no link to child’s own experience/world	1503 (28)
2 = Link to child’s own experience/world	3774 (71)
4 = Did not engage	61 (1)
Maternal discipline	
1 = No discipline/behavioural correction	4564 (85)
2 = One instance of mild behavioural correction	634 (12)
3 = Harsh discipline OR more than one mild behavioural correction	135 (3)
4 = Did not engage in tasks	<10 (<1)

Comparing children who completed the task with those who did not

Among the 791 children for whom the Parent-Child interaction task was not completed (12.5% of the two-year cohort), the reasons for non-completion were that the child was asleep (167, 21%), the interview was completed via phone or skype (140, 18%), there were language/translation difficulties (137, 17%), the child was not at home (125, 16%) or

did not co-operate (127, 16%), the child was unwell (18, 2%), or there was a developmental or other physical reason (10, 1%).

Participation in the task differed by ethnicity: 94% of European dyads, 85% of Māori dyads; 82% of Pacific dyads; 72% of Asian dyads; and 79% of Other ethnicity dyads participated. Mothers with a trade or university qualification (89%), and those with no formal qualifications (87%), were more like to engage

in the task than dyads where mothers whose highest qualification was secondary school (84%). Dyads living in the least deprived areas (93%) were more likely to engage in the interaction task than those living in medium deprivation (88%) and high deprivation (84%) areas.

Associations between Parent-Child Interaction ratings and socio-demographic variables

Distribution of parent-child interaction ratings by socio-demographic variables are given in table 3. Ordinal and logistic regression models were then developed to test associations. Each model tested maternal ethnicity, child gender, area deprivation and maternal education as predictors of each parent-child interaction variable (table 4).

Compared with European mothers, Māori mothers (OR = 0.79) were rated as displaying less warmth, as were Asian mothers (OR = 0.74). Asian mothers were also rated as using fewer open-ended questions (OR = 0.66), fewer emotion words (OR = 0.79), more discipline (OR = 1.82) and less linking language (OR = 0.74). Children of Asian mothers were rated as displaying less empathy (OR = 0.52). Compared with

European mothers, Pacific mothers were rated as using fewer emotion words (OR = 0.76) and using more discipline (OR = 1.62) and their children as displaying less empathy (OR = 0.82).

Compared with mothers of daughters, mothers of sons were rated as displaying more warmth (OR = 1.17), more discipline (OR = 1.51), and sons were rated as displaying less empathy (OR = 0.80).

Compared with families living in low deprivation areas, living in high deprivation was associated with mothers asking fewer open-ended questions (OR = 0.84) and with lower child empathy ratings (OR = 0.85). Similarly, medium deprivation was associated with lower use of maternal emotion words (OR = 0.85).

Compared with mothers with no formal educational qualifications, mothers with a trade or university qualification were more likely to use open-ended questions (OR = 1.52), emotion words (OR = 1.45; as were mothers with high school qualifications, OR = 1.46), and less likely to be rated as using discipline (OR = 0.69).

Table 3. Distributions of Parent-Child Interaction scores by maternal ethnicity and education, area deprivation and child gender: n(%)

	Maternal warmth			Maternal (open-ended) questions			Maternal emotion or desire words			Maternal discipline			Maternal linking		Children’s emotional expression / empathy	
	Low	Smile only	Physical affection	0	1	2+	0	1	2+	0	1 mild	Harsh or >1 mild	No links	Link to child’s world	Absence	Presence
Ethnicity																
European	254 (8)	881 (27)	2175 (66)	604 (18)	688 (21)	2010 (61)	979 (30)	991 (30)	1305 (40)	2901 (87)	336 (10)	79 (2)	897 (27)	2399(72)	1796(55)	1496 (45)
Māori	72 (10)	210 (30)	411 (59)	127 (19)	172 (25)	386 (56)	210 (31)	228 (33)	244 (36)	614 (88)	66 (9)	17 (2)	200 (29)	487 (71)	360 (53)	325 (47)
Pacific	46 (7)	195 (32)	377 (61)	131 (21)	150 (24)	332 (54)	217 (36)	196 (32)	193 (32)	481 (78)	119 (19)	18 (3)	182 (30)	427 (70)	380 (62)	233 (38)
Asian	38 (7)	179 (35)	291 (57)	129 (26)	120 (24)	256 (51)	175 (35)	148 (30)	173 (35)	405 (79)	90 (18)	16 (3)	167 (34)	331 (66)	348 (70)	149 (30)
Other	12 (7)	45 (27)	107 (65)	35 (22)	34 (21)	93 (57)	41 (26)	53 (34)	62 (40)	141 (87)	18 (11)	<10 (2)	50 (31)	111 (69)	94 (60)	63 (40)
Child gender																
Male	197 (7)	758 (28)	1789 (65)	529 (19)	599 (22)	1599 (59)	849 (32)	835 (31)	1007 (37)	2287 (83)	376 (14)	84 (3)	757 (28)	1950 (72)	1608 (59)	1102 (41)
Female	228 (9)	760 (29)	1589 (62)	502 (20)	573 (22)	1493 (58)	779 (31)	787 (31)	985 (39)	2277 (88)	258 (10)	51 (2)	746 (29)	1824 (71)	1387 (54)	1174 (46)
Area deprivation																
Low	103 (7)	426 (29)	957 (64)	265 (18)	294 (20)	923 (62)	418 (29)	437 (30)	614 (42)	1303 (88)	144 (10)	42 (3)	411 (28)	1066 (72)	800 (54)	676 (46)
Medium	158 (8)	543 (27)	1318 (65)	293 (19)	441 (22)	1182 (59)	648 (32)	604(30)	743 (37)	1755 (87)	221 (11)	48 (2)	570 (28)	1434 (72)	1128 (56)	872 (44)
High	162 (9)	542 (30)	1097 (61)	370 (21)	433 (24)	979 (55)	558 (32)	578 (33)	627(36)	1496 (83)	264 (15)	45 (2)	521 (29)	1260 (71)	1058 (59)	722 (41)
Maternal education																
No formal	39 (11)	91 (27)	206 (62)	84 (25)	86 (26)	166 (49)	133 (40)	101 (30)	99 (30)	275 (82)	52 (15)	10 (3)	101 (30)	235 (70)	192 (57)	144 (43)
High school	98 (9)	348 (30)	705 (61)	242 (21)	283 (25)	617 (54)	353 (31)	348 (31)	434 (38)	957 (83)	168 (15)	31 (3)	319 (28)	823 (72)	659 (58)	477 (42)
Trade or degree	284 (7)	1067 (28)	2453 (64)	698 (18)	796 (21)	2294 (61)	1131 (30)	1168 (31)	1447 (39)	3310 (87)	409 (11)	93 (2)	1079 (29)	2694 (71)	2124 (56)	1647 (44)

Table 4. Ordinal and logistic regressions: Differences in Parent-Child Interaction variables by maternal ethnicity and education, area deprivation and child gender

	Maternal warmth			Maternal (open-ended) questions			Maternal emotion or desire words			Maternal discipline			Maternal linking (at least one link made vs none)	Children’s emotional expression / empathy (displayed vs not)
	Low	Smile only	Physical affection	0	1	2+	0	1	2+	0	1 mild	Harsh or >1 mild		
	OR (95% CI)			OR (95% CI)			OR (95% CI)			OR (95% CI)			OR (95% CI)	OR (95% CI)
Ethnicity														
European	1.00			1.00			1.00			1.00			1.00	1.00
Māori	0.79 (0.67, 0.94), p = .009			0.99 (0.84, 1.18)			0.96 (0.81, 1.12)			0.81 (0.62, 1.05)			0.92 (0.76, 1.12)	1.17 (0.98, 1.40)
Pacific	0.91 (0.75, 1.11)			0.91 (0.76, 1.10)			0.76 (0.64, 0.91), p = .003			1.62 (1.27, 2.07), p = .0001			0.88 (0.72, 1.09)	0.82 (0.67, 0.99), p = .0007
Asian	0.74 (0.61, 0.89), p = .001			0.66 (0.55, 0.79), p < .0001			0.79 (0.67, 0.95), p = .01			1.82 (1.43, 2.32), p < .0001			0.74 (0.61, 0.91), p = .004	0.52 (0.42, 0.64), p < .0001
Other	1.01 (0.73, 1.40)			0.89 (0.65, 1.21)			1.10 (0.81, 1.48)			1.07 (0.67, 1.70)			0.83 (0.59, 1.17)	0.83 (0.59, 1.15)
Child gender														
Male	1.17 (1.05, 1.31), p = .005			1.01 (0.91, 1.13)			0.95 (0.86, 1.05)			1.51 (1.29, 1.77), p < .0001			1.05 (0.93, 1.18)	0.80 (0.72, 0.90), p = .0002
Female	1.00			1.00			1.00			1.00			1.00	1.00
Area deprivation														
Low	1.00			1.00			1.00			1.00			1.00	1.00
Medium	1.05 (0.92, 1.21)			0.89 (0.77, 1.01)			0.85 (0.75, 0.96), p = .009			1.02 (0.84, 1.25)			0.99 (0.85, 1.15)	0.93 (0.81, 1.06)
High	0.94 (0.80, 1.10)			0.84 (0.72, 0.97), p = .02			0.91 (0.79, 1.05)			1.20 (0.96, 1.49)			0.98 (0.83, 1.16)	0.85 (0.73, 0.99), p = .003
Maternal education														
No formal	1.00			1.00			1.00			1.00			1.00	1.00
High school	0.98 (0.77, 1.26)			1.21 (0.96, 1.53)			1.46 (1.16, 1.84), p = .001			0.89 (0.64, 1.22)			1.10 (0.84, 1.44)	1.00 (0.78, 1.28)
Trade or degree	1.09 (0.87, 1.38)			1.52 (1.22, 1.90), p = .0002			1.45 (1.17, 1.80), p = .0007			0.69 (0.50, 0.93), p = .02			1.07 (0.82, 1.37)	1.07 (0.84, 1.35)

Parent-Child Interaction task inter-correlations

Because some of the Parent-Child Interaction scales contained only two ordinal response options, Spearman correlation coefficients were conducted (Cliff, 2014). Parent-Child Interaction inter-correlations are shown in table 5. Small but significant

positive correlations were present between maternal warmth, maternal verbal communication variables, and child emotional expression. Observed maternal discipline was negatively correlated with maternal verbal communication variables and child emotional expression.

Table 5. Correlations among Parent-child Interaction task constructs

	Maternal warmth	Maternal (open-ended) questions	Maternal emotion or desire words	Maternal linking	Child’s emotional expression	Maternal discipline
Maternal warmth (n = 5321)	1.00	0.09***	0.06***	-0.00	0.02	0.00
Maternal (open-ended) questions (n = 5295)		1.00	0.06***	-0.03	0.05***	-0.06***
Maternal emotion or desire words (n = 5242)			1.00	0.13***	0.09***	-0.10***
Maternal linking (n = 5277)				1.00	0.06***	-0.04***
Child’s emotional expression (n = 5271)					1.00	-0.07***
Maternal discipline (n = 5333)						1.00

*** p <.0001

Convergent and discriminant construct validity

Parent-Child Interaction scale scores were correlated with maternal self-reported enjoyment of parenting, warmth and hostility towards their child, and frequency of story-telling and book reading with their child (table 6). Maternal self-report of the frequency of oral story-telling and book-reading was positively correlated with most of the positive Parent-Child Interaction constructs, and negatively correlated with observed maternal discipline. Similarly, mothers’ self-reported affiliation with her

child was positively correlated with all of the positive Parent-Child Interaction constructs, and negatively correlated with observed maternal discipline. Maternal self-reported warmth had few significant associations with observed Parent-Child Interaction constructs. Maternal self-reported parenting hostility was negatively correlated with the positive Parent-Child Interaction scale constructs, and positively correlated with observed discipline.

Table 6. Correlations between Parent-Child Interaction task observation ratings and maternal self-report of parent-child interactions

Observer Ratings	Maternal Self-Report				
	Frequency of oral story-telling with child	Frequency of reading books with child	Self-reported parenting warmth	Self-reported parenting hostility	Parent-child affiliation (Time Spent with Child Scale)
Maternal warmth (n = 5321)	0.01	0.09 ***	0.02	-0.06 ***	0.03 *
Maternal (open-ended) questions (n = 5295)	0.07 ***	0.09 ***	-0.01	-0.04 *	0.04 **
Maternal emotion or desire words (n = 5242)	0.03 *	0.05 ***	0.03 *	-0.06 ***	0.06 ***
Maternal linking (n = 5277)	0.01	0.03 *	0.00	-0.02 *	0.03 *
Child’s emotional expression (n = 5271)	0.05 ***	0.09 ***	0.01	-0.05 **	0.06 ***
Maternal discipline (n = 5333)	-0.05 **	-0.10 ***	-0.01	0.08 ***	-0.07 ***

*** p <.0001, ** p <.001, *p <.05

Discussion

We have demonstrated the inter-observer reliability and provided preliminary evidence of the convergent and discriminant validity of a new observational tool specifically designed to measure parent-child conversations in large interdisciplinary cohort studies. Significant correlations between observed variables were in the directions predicted (e.g., maternal discipline was negatively correlated with maternal verbal communication variables and children’s empathy; maternal warmth was positively correlated with maternal open-ended questions and emotion talk). The pattern of significant correlations between observer Parent-Child Interaction ratings and mothers’ self-reported parenting variables were in the directions predicted, and provide preliminary evidence of convergent and discriminant validity. For example, observed discipline was positively

correlated with maternal reports of parenting hostility. Observer ratings of positive Parent-Child Interaction constructs were positively correlated with maternal self-reports of parenting affiliation and book-reading and story-telling with their children; and negatively correlated with maternal self-reports of parenting hostility.

The small effect sizes warrant further discussion, particularly given statistical significance is more likely to be achieved with large samples. Small correlations between different Parent-Child Interaction constructs do indicate that we are in fact measuring discrete and specific constructs, rather than observers relying on a global impression of ‘positive’, or ‘negative’ parenting.

Associations between observed and self-report parent-child interaction variables, although generally in the direction expected, were small also. We must

acknowledge the possibility that these small effect sizes reflect larger than desired error (e.g., from observer ratings), particularly given that observer reliability was established prior to going into the field. However, observer agreement ratings were good: both at the initial training and mid-stream check.

Another explanation is simply that we are tapping into related, but meaningfully different constructs in these two modes of measurement: observation and self-report. On the one hand, this demonstrates one of the key reasons for observing parent-child interaction in the first place. Self-report parenting measures are designed to tap into parents' internal experience of their relationship with their child (e.g., "I enjoy having my child around me"; "I get angry at my child"), but parents may not always be aware of some aspects of their behaviour. In the example we gave earlier, parents may not be aware of how often they smile as an indicator of parental warmth. Indeed, there was no significant correlation between parents' observed smiling behaviour and their self-reported warmth. On the other hand, this highlights a limitation whereby self-report measures have been used to validate observational measures. Ideally we would have included another observational measure of the quality of parent-child interaction (e.g., the HOME scale) against which to compare our new task, but this simply was not possible given the interview time and overall cost constraints of the two-year data collection wave.

The current findings demonstrate that observers can be reliably trained to code the quality of parent-child conversations and interactions before going into the field, and that this reliability can be maintained across the data collection period. This is a particular strength given that while observers were experienced in data collection, none had prior experience with behavioural observation of parent-child interactions. We believe this was possible because of our emphasis during training on explaining a clear rationale for why we measure parent-child interactions, and simplifying the observation task to focus on a single behaviour during each 30-second time interval (i.e., to utilise *either* what you see *or* hear, not both). Certainly with greater personnel and financial resource, inter-observer agreement could always be improved. Ideally observers would demonstrate reliability on around 15% of their sample: for our interviewers that

would equate to between thirty and forty different dyads. It was simply not possible to record this number of different interactions before going into the field: a sample of this magnitude would represent an entirely separate study. Despite this, our findings demonstrate that reliability can be achieved within the practical constraints of a large cohort study.

Parent-child conversations differed as a function of maternal ethnicity. These findings were generally consistent with past research: Western parents tend to have more elaborative discussions (of which open-ended questions is a key coding target), for example, when compared with Asian parents (Wang, 2001; Wang & Fivush, 2005; Wang, Leichtman & Davies, 2000). There is some evidence that Pacific parents (Schluter, Sundborn, Abbott & Paterson, 2007), and some groups of Asian parents (Lau, Takeuchi & Alegria, 2006) more commonly use physical discipline. The current findings highlight slightly higher levels of observed physical discipline as well as higher rates of mild behavioural corrections for Pacific and Asian mothers.

We must consider the possibility, however, that ratings for ethnic minority dyads may be partly impacted by the nature of the task. None of the pictures specifically included Māori, Pacific or Asian children. Our data for these populations may also be impacted by higher rates of non-participation, particularly for Asian families. It may also be that the nature of the task is less culturally relevant. For example, the use of prompting pictures may be less appropriate or necessary within Māori culture with its strong tradition of oral story-telling (Reese, Hayne, & MacDonald, 2008).

Gender differences were also consistent with the existing research pointing to differences in how sons and daughters are socialised. Girls are consistently found to display greater empathy than boys (see Chaplan & Aldao, 2013, for a recent review). This is thought to be due to a combination of genetically-based temperament differences and socialisation influences that encourage girls to be more socially driven and express positive rather than negative emotions (Chaplan & Aldao, 2013; Zahn-Waxler, 2000). The finding of higher maternal discipline ratings with boys is also consistent with the existing literature. Interestingly, mothers were also rated as displaying greater warmth with sons than daughters.

While this is somewhat surprising, it does concur with theory and research that emphasises warmth and discipline as two distinct parenting constructs. It also highlights the need to consider mediating and moderating factors – something that large longitudinal studies are ideally placed to do. In particular, further research is needed to examine interactions between child gender, child temperament, self-reported parenting practices, and observations of parent-child interactions.

Findings related to maternal warmth warrant further discussion. While maternal warmth is theoretically (Ainsworth et al., 1978; Bowlby, 1969) and empirically (NICHD Early Child Care Research Network, 2001) related to attachment security, these are distinct constructs. A secure attachment bond is characterised by responsive, sensitive care-giving which allows a child to use their caregiver as a secure base from which to explore, and to return to seek comfort. In contrast, insecure attachment is more likely to be associated with inconsistent, rejecting or avoidant caregiving, and children in turn display avoidant, anxious, or mixed patterns of exploration and responding (Ainsworth et al., 1978). Attachment security is a dyadic relationship (not a specific maternal behaviour) that encompasses cognitive internal working models as well as observed behaviours. Valid measurement of differences in attachment security is based on observations of young children under mild stress, for example, during separation (Ainsworth et al., 1978). This is very different to the observation of maternal warmth during a typical parent-child interaction used here. Similarly, our measure of child empathy is just one example of emotional expression, which may not necessarily reflect children’s responding across more salient situations (e.g., if a sibling or parent hurts themselves). This highlights the importance of considering observed behaviours in combination with parental self-report.

Our findings also highlight areas where, in hindsight, a more differentiated coding scheme may

be warranted. For example, more than half of our mothers used two or more open-ended questions. There may be variability in the complexity and quality of these questions that could be further separated. For example, a mother asking ‘what’s happening here?’ and ‘what else can you see?’ could receive the same rating as a mother who asked ‘why do you think they are washing the car?’ and ‘what sort of reward will they get for being so helpful?’ Any future researchers using this task may want to consider modifying in this way.

Observational methods, of course, are not without criticism. The very presence of an observer may create bias, combined with the somewhat artificial nature of engaging parents and children in a semi-structured interaction task (Gardner, 2000). For example, research indicates that parents talk more and laugh less with their toddlers when they think they are being videotaped (Field & Ignatoff, 1981). However, one could argue that longitudinal cohorts may be less likely to display this type of reactivity: by the time our families were visited at age two, they had already met with *Growing Up in New Zealand* researchers on two occasions face-to-face in their own homes, and had completed at least two telephone interviews.

In conclusion, we believe that we have created a promising tool for the direct observation of parent-child behaviour in very large samples. Valid measurement requires multiple informants and multiple forms of measurement (Dunn & Kendrick, 1980). It is our opinion that including behavioural observations of parent-child interactions, as well as parent report, will only strengthen our understanding of their contribution to children’s emotional, behavioural and social development across time. We look forward to extensions and refinements of the tool and training procedures as other researchers adapt it for their specific purposes.

Acknowledgements

We acknowledge the children and families who are part of the *Growing Up in New Zealand* study and contribute their immensely valuable time and knowledge. We also acknowledge all members of the *Growing Up in New Zealand* team, as well as our Kaitiaki and Scientific Advisory Groups.

Growing Up in New Zealand has received ethical approval from the Northern Y Regional Ethics Committee for data collection with participants and for specific data linkages. Reference NTY/08/06/055.

We acknowledge the key support and funding of New Zealand’s Social Policy Evaluation and Research Unit (Superu) and the Ministry of Social Development. Alongside the Ministry of Social Development and Superu (previously the Families Commission), the Ministry of Health and The University of Auckland (with Auckland UniServices Limited) have contributed the most significant funding and support to the cohort. Other agencies have also contributed funding including the Ministries of Education, Justice, Science and Innovation, Women's Affairs, and Pacific Island Affairs; the Departments of Labour, and Corrections; Te Puni Kōkiri (Ministry of Māori Affairs); New Zealand Police; Sport and Recreation New Zealand; Housing New Zealand; and the Mental Health Commission. Treasury and the Health Research Council also provided support in the development phase of the study, and the Office of Ethnic Affairs, Statistics New Zealand and the Children’s Commission provided consultation.

Elaine Reese contributed to the conception and design of the study, developed the data collection instruments, and completed the first and final drafts of the manuscript.

Amy Bird analysed and interpreted the data, and completed the first and final drafts of the manuscript.

Mele Taumoepeau developed the data collection instruments, revised the manuscript, and approved the final manuscript as submitted.

Joanna Schmidt assisted with the data collection instruments, revised the manuscript, and approved the final manuscript as submitted.

Jatender Mohal assisted with the data analysis approach, revised the manuscript, and approved the final manuscript as submitted.

Cameron Grant contributed to the conception and design of the study, revised the manuscript, and approved the final manuscript as submitted.

Polly Atatoa Carr revised the manuscript and approved the final manuscript as submitted.

Susan Morton conceived and designed the study, revised the manuscript, and approved the final manuscript as submitted.

References

- Ainsworth, M. D. S., Blehar, M. C., Waters, E. & Wall, S. (1978). *Patterns of Attachment: A Psychological Study of the Strange Situation*. Hillsdale, N.J.: Erlbaum, 1978. <http://dx.doi.org/10.1016/j.cpr.2009.11.004>
- Aldao, A., Nolen-Hoeksema, S. & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review*, 30(2), 217-237.
- Baumrind, D. (1967). Child care practices anteceding three patterns of preschool behavior. *Genetic Psychology Monographs*, 75(1), 43-88
- Bird, A., & Reese, E. (2006). Emotional reminiscing and the development of an autobiographical self. *Developmental Psychology*, 42(4), 613. <http://dx.doi.org/10.1037/0012-1649.42.4.613>
- Bornstein, M. H. & Bradley, R. H. (Eds.). (2014). *Socioeconomic Status, Parenting, and Child Development*. Routledge.
- Bornstein, M. H. & Haynes, O. M. (1998). Vocabulary competence in early childhood: Measurement, latent construct, and predictive validity. *Child Development*, 69(3), 654-671. <http://dx.doi.org/10.1111/j.1467-8624.1998.tb06235.x>

- Bornstein, M. H., Tal, J., Rahn, C., Galperin, C. Z., Pêcheux, M. G., Lamour, M., Toda, S., Azuma, H., Ogino, M. & Tamis-LeMonda, C. S. (1992). Functional analysis of the contents of maternal speech to infants of 5 and 13 months in four cultures: Argentina, France, Japan, and the United States. *Developmental Psychology*, 28(4), 593-603. <http://dx.doi.org/10.1037/0012-1649.28.4.593>
- Bowes, L., Maughan, B., Caspi, A., Moffitt, T. E. & Arseneault, L. (2010). Families promote emotional and behavioural resilience to bullying: evidence of an environmental effect. *Journal of Child Psychology and Psychiatry*, 51, 809-817. <http://dx.doi.org/10.1111/j.1469-7610.2010.02216.x>
- Bowlby, J. (1969). *Attachment and Loss: Vol. 1. Attachment*. London: Hogarth Press.
- Bradley, R. H. & Caldwell, B. M. (1984). The HOME Inventory and family demographics. *Developmental Psychology*, 20(2), 315. <http://dx.doi.org/10.1037/0012-1649.20.2.315>
- Bradley, R. H., Caldwell, B. M., Rock, S. L., Ramey, C. T., Barnard, K. E., Gray, C., Hammond, M. A, Mitchell, S., Gottfried, A. W., Siegel, L., & Johnson, D. L. (1989). Home environment and cognitive development in the first 3 years of life: A collaborative study involving six sites and three ethnic groups in North America. *Developmental Psychology*, 25(2), 217. <http://dx.doi.org/10.1037/0012-1649.25.2.217>
- Bradley, R. H., Corwyn, R. F., Burchinal, M., McAdoo, H. P., & García Coll, C. (2001). The home environments of children in the United States Part II: Relations with behavioral development through age thirteen. *Child Development*, 72(6), 1868-1886. <http://dx.doi.org/10.1111/1467-8624.t01-1-00383>
- Brockmeier, J. & Carbaugh, D. A. (Eds.) (2001). *Narrative and Identity: Studies in Autobiography, Self and Culture* (Vol. 1). John Benjamins Publishing. <http://dx.doi.org/10.1075/sin.1>
- Brown, J. R. & Dunn, J. (1996). Continuities in emotion understanding from three to six years. *Child Development*, 67(3), 789-802. <http://dx.doi.org/10.2307/1131861>
- Chaplan, T. M. & Aldao, A. (2013). Gender Differences in Emotion Expression in Children: A Meta-Analytic Review. *Psychological Bulletin*, 139(4), 735-765. <http://dx.doi.org/10.1037/a0030737>
- Cliff, N. (2014). *Ordinal Methods for Behavioural Data Analysis*. Psychology Press.
- Coffman, J. L., Ornstein, P. A., McCall, L. E. & Curran, P. J. (2008). Linking teachers' memory-relevant language and the development of children's memory skills. *Developmental Psychology*, 44(6), 1640. <http://dx.doi.org/10.1037/a0013859>
- Davies, P. T., Harold, G. T., Goeke-Morey, M. C. & Cummings, E. M. (2002). Child emotional security and interparental conflict. *Monographs of the Society for Research on Child Development*, 270, 67(3).
- de Wied, M., Gispen-de Wied, C. & van Boxtel, A. (2010). Empathy dysfunction in children and adolescents with disruptive behavior disorders. *European Journal of Pharmacology*, 626, 97-103. <http://dx.doi.org/10.1016/j.ejphar.2009.10.016>
- Dunn, J., Brown, J. & Beardsall, L. (1991). Family talk about feeling states and children's later understanding of others' emotions. *Developmental Psychology*, 27, 448-455. <http://dx.doi.org/10.1037/0012-1649.27.3.448>
- Dunn, J. & Kendrick, C. (1980). Studying temperament and parent-child interaction: comparison of interview and direct observation. *Developmental Medicine & Child Neurology*, 22(4), 484-496. <http://dx.doi.org/10.1111/j.1469-8749.1980.tb04353.x>
- Farrant, K. & Reese, E. (2000). Maternal style and children's participation in reminiscing: stepping stones in children's autobiographical memory development. *Journal of Cognition and Development*, 1, 193-225. <http://dx.doi.org/10.1207/S15327647JCD010203>
- Field, T. & Ignatoff, E. (1981). Videotaping effects on the behaviors of low income mothers and their infants during floor-play interactions. *Journal of Applied Developmental Psychology*, 2, 227-235. [http://dx.doi.org/10.1016/0193-3973\(81\)90003-4](http://dx.doi.org/10.1016/0193-3973(81)90003-4)
- Findlay, L. C., Girardi, A. & Coplan, R. J. (2006). Links between empathy, social behavior, and social understanding in early childhood. *Early Childhood Research Quarterly*, 21(3), 347-359. <http://dx.doi.org/10.1016/j.ecresq.2006.07.009>
- Fivush, R., Berlin, L., Sales, J. M., Menutti-Washburn, J. & Cassidy, J. (2003). Functions of parent-child

- reminiscing about emotionally negative events. *Memory*, 11, 179–192.
<http://dx.doi.org/10.1080/741938209>
- Fletcher, K. L. & Reese, E. (2005). Picture book reading with young children: A conceptual framework. *Developmental Review*, 25(1), 64-103. <http://dx.doi.org/10.1016/j.dr.2004.08.009>
- Fulgini, A. S., & Brooks-Gunn, J. (2013). Mother–child interactions in Early Head Start: Age and ethnic differences in low-income dyads. *Parenting*, 13(1), 1-26. <http://dx.doi.org/10.1080/15295192.2013.732422>
- Gardner, F. (2000). Methodological issues in the direct observation of parent-child interaction: do observational findings reflect the natural behavior of participants? *Clinical Child and Family Psychology Review*, 3, 185-198. <http://dx.doi.org/10.1023/A:1009503409699>
- Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology*, 2, 271-299. <http://dx.doi.org/10.1037/1089-2680.2.3.271>
- Grusec, J. E. (2011). Socialization processes in the family: Social and emotional development. *Annual Review of Psychology*, 62, 243-269. <http://dx.doi.org/10.1146/annurev.psych.121208.131650>
- Haden, C. A., Reese, E. & Fivush, R. (1996). Mothers' extra-textual comments during storybook reading: Stylistic differences over time and across texts. *Discourse Processes*, 21(2), 135-169.
<http://dx.doi.org/10.1080/01638539609544953>
- Hart, B. & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H Brookes Publishing.
- Hindman, A. H., Skibbe, L. E. & Foster, T. D. (2014). Exploring the variety of parental talk during shared book reading and its contributions to preschool language and literacy: evidence from the Early Childhood Longitudinal Study-Birth Cohort. *Reading and Writing*, 27, 287-313. <http://dx.doi.org/10.1007/s11145-013-9445-4>
- Huttenlocher, J. (1998). Language input and language growth. *Preventive Medicine*, 27(2), 195-199.
<http://dx.doi.org/10.1006/pmed.1998.0301>
- Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, 27(2), 236. <http://dx.doi.org/10.1037/0012-1649.27.2.236>
- Jansen, P. W., Raat, H., Mackenbach, J. P., Hofman, A., Jaddoe, V. W. V., Bakermans-Kranenburg, M. J., Van IJzendoorn, M. H., Verhulst, F. C. & Tiemeier, H. (2012). Early determinants of maternal and paternal harsh discipline: The Generation R study. *Family Relations*, 61(2), 253-270.
<http://dx.doi.org/10.1111/j.1741-3729.2011.00691.x>
- Lau, A. S., Takeuchi, D. T. & Alegria, M. (2006). Parent-to-child aggression among Asian American parents: culture, context, and vulnerability. *Journal of Marriage and Family*, 68(5), 1261-1275.
<http://dx.doi.org/10.1111/j.1741-3737.2006.00327.x>
- Locke, L. M. & Prinz, R. J. (2002). Measurement of parental discipline and nurturance: A review. *Clinical Psychology Review*, 22, 895 – 929. [http://dx.doi.org/10.1016/S0272-7358\(02\)00133-2](http://dx.doi.org/10.1016/S0272-7358(02)00133-2)
- Lovejoy, M. C., Graczyk, P. A., O’Hare, E. & Neuman, G. (2000). Maternal depression and parenting behavior: A meta-analytic review. *Clinical Psychology Review*, 20, 561-592. [http://dx.doi.org/10.1016/S0272-7358\(98\)00100-7](http://dx.doi.org/10.1016/S0272-7358(98)00100-7)
- Lu, M. C. (2014). Improving maternal and child health across the life course: where do we go from here? *Maternal and Child Health Journal*, 18(2), 339-343.
- Lytton, H. (1971). Observation studies of parent-child interaction: A methodological review. *Child Development*, 651-684. <http://dx.doi.org/10.2307/1127439>
- Malatesta, C. Z. & Wilson, A. (1988). Emotion cognition interaction in personality development: A discrete emotions, functionalist analysis. *British Journal of Social Psychology*, 27(1), 91-112.
<http://dx.doi.org/10.1111/j.2044-8309.1988.tb00807.x>

- Margolin, G., Oliver, P. H., Gordis, E. B., O’Hearn, H. G., Medina, A. M., Ghosh, C. M. & Morland, L. (1998). The nuts and bolt of behavioural observation of marital and family interaction. *Clinical Child and Family Psychology Review*, 1(4), 195-213. <http://dx.doi.org/10.1023/A:1022608117322>
- Melby, J. N., Conger, R. D., Book, R., Rueter, M., Lucy, L., Repinski, D., Ahrens, K., Black, D., Brown, D., Huck, S., Mutchler, L., Rogers, S., Ross, J. & Stavros, T. (1989-1993). *The Iowa Family Interaction Rating Scales (editions 1-4). Unpublished coding manual*. Iowa State University, Institute for Social and Behavioral Research, Ames.
- Miller, P. J., Mintz, J., Hoogstra, L., Fung, H. & Potts, R. (1992). The narrated self: Young children's construction of self in relation to others in conversational stories of personal experience. *Merrill-Palmer Quarterly (1982-)*, 45-67.
- Miller, P. J., & Sperry, L. L. (1988). Early talk about the past: The origins of conversational stories of personal experience. *Journal of Child Language*, 15(02), 293-315. <http://dx.doi.org/10.1017/S0305000900012381>
- Morton, S. M. B., Atatoa Carr, P. E., Grant, C. C., Robinson, E. M., Bandara, D. K., Bird, A., Ivory, V. C., Kingi, T. K., Liang, R., Marks, E. J., Perese, L. M., Peterson, E. R., Pryor, J. E., Reese, E., Schmidt, J. M., Waldie, K. E. & Wall, C. (2013). Cohort Profile: *Growing Up in New Zealand*. *International Journal of Epidemiology*, 42(1), 65-75. <http://dx.doi.org/10.1093/ije/dyr206>
- Morton, S. M. B., Grant, C. C., Atatoa Carr, P. E., Robinson, E. M., Kinloch, J. M., Fleming, C. J., Kingi, T. R., Perese, L. M. & Liang, R. (2014a). How do you recruit and retain a prebirth cohort? Lessons learnt from *Growing Up in New Zealand*. *Evaluation and the Health Professions*, 37(4), 411-433. <http://dx.doi.org/10.1177/0163278712462717>
- Morton, S. M. B., Ramke, J., Kinloch, J., Grant, C. C., Atatoa Carr, P. E., Leeson, H., Lee A. C. & Robinson, E. (2014b). *Growing Up in New Zealand* cohort alignment with all New Zealand births. *Australian and New Zealand Journal of Public Health*, 39(1), 82-87. <http://dx.doi.org/10.1111/1753-6405.12220>
- Mott, F. L. (2004). The utility of the HOME-SF Scale for child development research in a Large national longitudinal survey: The National Longitudinal Survey of Youth 1979 Cohort. *Parenting*, 4(2-3), 259-270. <http://dx.doi.org/10.1080/15295192.2004.9681273>
- Najarian, M., Snow, K., Lennon, J., Kinsey, S. & Mulligan, G. (2010). Early Childhood Longitudinal Study, Birth Cohort (ECLS-B). *Preschool-Kindergarten 2007 Psychometric Report (NCES 2010-009)*. National Center for Education Statistics, Institute of Education Sciences, U. S. Department of Education, Washington, DC.
- NICHD Early Child Care Research Network. (2001). Child-care and family predictors of preschool attachment and stability from infancy. *Developmental Psychology*, 37(6), 847-862. <http://dx.doi.org/10.1037/0012-1649.37.6.847>
- Payne, A. C., Whitehurst, G. J. & Angell, A. L. (1994). The role of home literacy environment in the development of language abilities in preschool children from low income families. *Early Childhood Research Quarterly*, 9, 427-440. [http://dx.doi.org/10.1016/0885-2006\(94\)90018-3](http://dx.doi.org/10.1016/0885-2006(94)90018-3)
- Reese, E., Bird, A. & Tripp, G. (2007). Children's self-esteem and moral self: Links to parent–child conversations regarding emotion. *Social Development*, 16(3), 460-478. <http://dx.doi.org/10.1111/j.1467-9507.2007.00393.x>
- Reese, E. & Cox, A. (1999). Quality of adult book reading affects children's emergent literacy. *Developmental Psychology*, 35(1), 20-28. <http://dx.doi.org/10.1037/0012-1649.35.1.20>
- Reese, E., Haden, C. A. & Fivush, R. (1993). Mother-child conversations about the past: Relationships of style and memory over time. *Cognitive Development*, 8(4), 403-430. [http://dx.doi.org/10.1016/S0885-2014\(05\)80002-4](http://dx.doi.org/10.1016/S0885-2014(05)80002-4)
- Reese, E., Hayne, H. & MacDonald, S. (2008). Looking back to the future: Māori and Pakeha mother–child birth stories. *Child Development*, 79(1), 114-125. <http://dx.doi.org/10.1111/j.1467-8624.2007.01114.x>
- Reese, E. & Newcombe, R. (2007). Training mothers in elaborative reminiscing enhances children’s autobiographical memory and narrative. *Child Development*, 78(4), 1153-1170. <http://dx.doi.org/10.1111/j.1467-8624.2007.01058.x>

- Rheingold, H. L. (1960). The measurement of maternal care. *Child Development*, 31(3), 565-575.
<http://dx.doi.org/10.2307/1126050>
- Roberts, W. & Strayer, J. (1996). Empathy, emotional expressiveness, and prosocial behavior. *Child Development*, 449-470. <http://dx.doi.org/10.2307/1131826>
- Rutter, M. (2013). Annual research review: Resilience – clinical implications. *The Journal of Child Psychology and Psychiatry*, 54, 474-487. <http://dx.doi.org/10.1111/j.1469-7610.2012.02615.x>
- Salmond, C., Crampton, P. & Atkinson, J. (2007). *NZDep2006 Index of Deprivation*. Wellington: Department of Public Health, University of Otago.
- Schluter, P. J., Sundborn, G., Abbott, M. & Paterson, J. (2007). Smacking—are we being too heavy-handed? Findings from the Pacific Islands Families Study. *New Zealand Medical Journal*, 14.
- Snow, C. E. (1991). The theoretical basis for relationships between language and literacy in development. *Journal of Research in Childhood Education*, 6(1), 5-10. <http://dx.doi.org/10.1080/02568549109594817>
- Sroufe, L. A., Egeland, B., Carlson, E. A. & Collins, W. A. (2005). *The development of the person: The Minnesota Study of Risk and Adaptation from Birth to Adulthood*. New York: Guilford Press.
- Strauss, R. S. & Knight, J. (1999). Influence of the home environment on the development of obesity in children. *Pediatrics*, 103(6), e85-e85. <http://dx.doi.org/10.1542/peds.103.6.e85>
- Taumoepeau, M. & Ruffman, T. (2006). Mother and infant talk about mental states relates to desire language and emotion understanding. *Child Development*, 77(2), 465-481. <http://dx.doi.org/10.1111/j.1467-8624.2006.00882.x>
- Totsika, V. & Sylva, K. (2004). The Home Observation for Measurement of the Environment revisited. *Child and Adolescent Mental Health*, 9(1), 25-35. <http://dx.doi.org/10.1046/j.1475-357X.2003.00073.x>
- Tully, L. A., Arseneault, L., Caspi, A., Moffitt, T. E. & Morgan, J. (2004). Does maternal warmth moderate the effects of birth weight on twins' Attention-Deficit/Hyperactivity Disorder (ADHD) symptoms and IQ? *Journal of Consulting and Clinical Psychology*, 72, 218-226.
- Wadsworth, M. & DeZateux, C. (2013). The growing inter-disciplinary field of longitudinal studies beginning in early life: this journal's response. *Longitudinal and Life Course Studies*, 4(1), 1-3.
- Wang, Q. (2001). “Did you have fun?”: American and Chinese mother–child conversations about shared emotional experiences. *Cognitive Development*, 16(2), 693-715. [http://dx.doi.org/10.1016/S0885-2014\(01\)00055-7](http://dx.doi.org/10.1016/S0885-2014(01)00055-7)
- Wang, Q. & Fivush, R. (2005). Mother–child conversations of emotionally salient events: exploring the functions of emotional reminiscing in European-American and Chinese families. *Social Development*, 14(3), 473-495. <http://dx.doi.org/10.1111/j.1467-9507.2005.00312.x>
- Wang, Q., Leichtman, M. D. & Davies, K. I. (2000). Sharing memories and telling stories: American and Chinese mothers and their 3-year-olds. *Memory*, 8(3), 159-177. <http://dx.doi.org/10.1080/096582100387588>
- Welch-Ross, M. K. (1997). Mother–child participation in conversation about the past: Relationships to preschoolers' theory of mind. *Developmental Psychology*, 33(4), 618-629.
<http://dx.doi.org/10.1037/0012-1649.33.4.618>
- Zahn-Waxler, C. (2000). The development of empathy, guilt, and internalization of distress: Implications for gender differences in internalizing and externalizing problems. *Anxiety, Depression, and Emotion*, 222-265. <http://dx.doi.org/10.1093/acprof:oso/9780195133585.003.0011>
- Zimmerman, F. J., Gilkerson, J., Richards, J. A., Christakis, D. A., Xu, D., Gray, S. & Yapanel, U. (2009). Teaching by listening: The importance of adult-child conversations to language development. *Pediatrics*, 124(1), 342-349.

Endnotes

- ⁱ There was occasional variability depending on child availability and cooperation, but the Parent-Child Interaction task always followed the consent process and mother interview, and was typically administered part-way through the other child observations (it always followed the Stack and Topple motor play task, and was usually before the child’s height and weight measurements).
- ⁱⁱ Observers were also trained to establish coding reliability for other child observation tasks not included in this paper (e.g., Stack and Topple).
- ⁱⁱⁱ The initial training videos also used a mother-child picture description task, although the dyads in these videos used different prompting pictures to those used in the main study. The mid-stream reliability training videos, however, were specifically developed for this project and used the same pictures as the main study.
- ^{iv} Different clips were used to the initial training. The parent-child interaction task was the same as that used for the main study.