Video-Based Assessment and Rating of Parent-Child Interaction Within the National Educational Panel Study

Anja Sommer, Claudia Hachul and Hans-Günther Roßbach

Abstract

There is strong evidence that the learning opportunities offered in familial learning environments have a long-lasting impact on children's development and educational career. As one of only a few large-scale longitudinal studies, the National Educational Panel Study (NEPS) Starting Cohort 1—Newborns is taking up the challenge of direct assessment of parent-child interaction in familial learning environments. This article describes how this assessment was developed, comparing existing observational designs and instruments with regard to their large-scale practicability and utility for the NEPS. To gain reliable data on parent-child interaction, we apply the following procedure: (1) an overt, non-participant field observation of parent-child interaction embedded in a semi-standardized play situation, which is videotaped, and (2) an analysis of the videotaped parent-child interaction using a macroanalytic rating instrument adapted from the NICHD Study of Early Child Care and Youth Development (NICHD SECCYD). We illustrate the general practicability and reliability of this assessment with results from the first pilot study (N = 466). We point out potential pitfalls in implementing this approach by discussing the results of different in-depth analyses. Finally, we detail the resulting adaptations in the assessment and rating of parent-child interaction for the first main study (N = 3,481).

1 Parent-Child Interaction and its Importance for Child Development

The familial learning environment is of profound significance, especially in early life. Familial learning opportunities are most important at a very young age, and there is strong evidence that these opportunities have a long-lasting impact on child devel-

Hans-Peter Blossfeld et al. (eds.), *Methodological Issues of Longitudinal Surveys,*
DOI 10.1007/978-3-658-11994-2_9

opment (Belsky et al., 2007; Blomeyer, Laucht, Pfeiffer, & Reuß, 2010; NICHD Early Child Care Research Network, 2002).

Structural characteristics of the familial learning environment, such as income and parental education, are often considered when associations between family characteristics and cognitive or social development in early life are studied (Halle et al., 2009; Hillemeier, Farkas, Morgan, Martin, & Maczuga, 2009). However, not only structural characteristics, but also educational processes such as parent-child interactions seem to play a key role in children's cognitive, linguistic, and socio-emotional development. Even if structural factors are controlled, associations between a child's development and educational processes remain significant (Belsky et al., 2007; Bornstein & Tamis-LeMonda, 1989; Bromley, 2009; Leerkes, Blankson, & O'Brien, 2009; NICHD Early Child Care Research Network, 2002; Page, Wilhelm, Gamble, & Card, 2010). Accordingly, an assessment of learning environments should not only consider structural characteristics, but also educational processes (Bäumer, Preis, Roßbach, Stecher, & Klieme, 2011).

A detailed look at these educational processes is offered through the observation of parent-child interactions. In these interactions, different factors, such as activating behavior, sensitivity, and responsiveness, have been found to be related to different aspects of later child development (Blomeyer et al., 2010; Leerkes et al., 2009; NICHD Early Child Care Research Network, 1998; Page et al., 2010). Therefore, the National Educational Panel Study (NEPS) assesses processes in familial learning environments beyond parent self-reports and observes parent-child interaction in the very first years of a child's life. In order to assess these aspects (in addition to others), the Newborn Cohort of the NEPS used a nationally representative sample of 3,481 children[1] born in Germany from March to August 2012 and follows these children longitudinally (Aßmann et al., 2011). In the first three years of the child's life, three measurement points at the age of 7, 16, and 26 months are given in the longitudinal study design of the NEPS.

2 Assessment of Parent-Child Interaction in Large-Scale Studies

Diverse methodological approaches can be applied for the assessment of parent-child interaction. To justify the choice of the methodological approach used in the NEPS, we discuss different observational designs and instruments regarding these approaches' large-scale practicability and utility for the NEPS.

1 National Educational Panel Study (NEPS): Starting Cohort 1 – Newborns (SC1), doi:10.5157/NEPS:SC1:1.0.0

2.1 Observational Designs

For the classification of structured observation, Greve and Wentura (1997) distinguish different observational designs along several bipolar classifications: (1) overt vs. covert observation, (2) participant vs. non-participant observation, (3) laboratory vs. field observation, and (4) technically mediated vs. non-mediated observation. These designs differ with regard to their capability of assessing the targeted observational subject and regarding their large-scale practicability.

(1) Considering ethical correctness, an assessment of parent-child interaction has to be overt (Greve & Wentura, 1997). (2) Aiming at a standardized assessment of the interaction between parent and child, active participation of the observer is not constructive for the assessment of this dyadic situation. (3) Laboratory observation offers the opportunity to control framework better than field observation. However, with respect to the assessment of parent-child interaction, field observation in the natural home setting of the family may decrease reactive effects (Rentzsch & Schütz, 2009). (4) These days, most studies use video-mediated observation for the assessment of interactions. Because the assessment of an interaction sequence and rating parent-child interaction is separated, interviewers as well as raters are prevented from managing too many tasks simultaneously, which improves the quality of the assessment. Additionally, storage and repeatability of the data allow for consistent field monitoring and checking for quality via the possible application of several raters. Therefore, video-based observation is highly practicable for large-scale studies.

For these reasons, the Newborn Cohort of the NEPS applies an overt, non-participant field observation of parent-child interaction, which is videotaped. Therefore, the assessment is subdivided into the assessment of the interaction sequence and the subsequent rating of the parent-child interaction.

2.2 Observational Instruments

Level of observation

Different types of observational instruments can be applied based on the specifications of the observational design. Instruments for observational assessment can be classified as micro- and macroanalytic, differing in their level of observation.

Microanalytic instruments aim at specific aspects of interaction and focus mainly on the categorization or coding of frequency and the duration of behavior (Faßnacht, 1995; Greve & Wentura, 1997). Faßnacht (1995) distinguishes two microanalytic approaches: Event-sampling methods record every occurrence of a preselected behavioral pattern over a specific observational period. Time-sampling methods separate the stream of time into short, continuous, consecutive time sequences, often lasting for 5 to 10 seconds. Observers decide on the occurrence of predefined behavior with regard to each sequence following an all-or-nothing principle. Both microanalytic

approaches are rather time-consuming and are commonly used in small-scale studies (e. g., Bornstein & Tamis-LeMonda, 1989; Hirschmann, Kastner-Koller, Deimann, Aigner, & Svecz, 2011). In contrast, macroanalytic rating procedures have a high level of aggregation, downplaying minute contextual variability (Bornstein, Hahn, Suwalsky, & Haynes, 2011). They offer a rather global impression and capture characteristics and enduring traitlike features of individuals and are therefore commonly used for assessing intensity or behavior as a whole (Faßnacht, 1995). Time effectiveness and a broad global assessment of the targeted construct are highly important for large-scale studies. Therefore, like the majority of large-scale longitudinal studies, NEPS implements a macroanalytic instrument for the rating of the videotaped parent-child interactions. However, due to the videotaping, microanalytic approaches focusing on details of the mother-child interaction are applied later on.

Instruments

There seems to be no standard macroanalytic instrument for rating the parent-child interaction that fits different study designs and requirements. To detail the decision for the instrument used in the NEPS, we list existing instruments regarding the included constructs and aspects indicating large-scale practicability in Table 1. For this purpose, we used an overview of Wiefel et al. (2007), but for our purpose, we excluded instruments that do not aim at the age group under consideration (FIT-K98, a family- and kindergarten-interaction test, and Mahoney's Maternal Behavior Rating), or these instruments were used for psychiatric mother and baby units (BMIS, Bethlem Mother-Infant Interaction Scale). Additionally, we considered instruments used in foregoing birth-cohort studies (see Schlesiger, Lorenz, Weinert, Schneider, & Roßbach, 2011 for an overview).

Any instrument to be used in the NEPS has to meet the discussed methodological requirements concerning the observational design and the level of observation. Regarding observational design, all listed instruments are based on an overt, non-participant video-mediated observation. Additionally, all instruments offer the opportunity to rate interaction sequences that are videotaped in home settings. Regarding the level of observation, all listed instruments are classified as macroanalytic. Although macroanalytic instruments are usually time-efficient, some approaches are more time-consuming than others. Aiming towards a short duration of rating with a high-quality analysis and reliable data at the same time, time-consuming macroanalytic instruments, such as the Mannheim Rating System for Mother-Infant Face-to-Face Interaction (MBS-MKI-S) and the Nursing Child Assessment Teaching Scale (NCATS) (see Table 1), were excluded.

In addition to these aspects, large-scale practicability for the NEPS can also be discussed along two points: First, the burden of every assessment of interaction sequence should be kept as low as possible to avoid high rates of panel attrition. Because of time constraints, time spent in the home setting of the family should be kept as low as possible (Schlesiger et al., 2011). Therefore, instruments whose rating de-

pends on interactional sequences that exceed 10 minutes cannot be used (this applies to the Emotional Availability Scales, EA-III; see Table 1). Second, accessibility of the instrument has to be considered. Coding Interactive Behavior (CIB) and CARE-Index (CARE) (see Table 1) have not been published yet. EA-III and NCATS (see Table 1) are only accessible after an intensive training by the author or other licensed trainers, who are partly not located in Germany. For a large-scale study like the NEPS, rater training should instead be flexible in time and persons.

Therefore, we decided to adapt the instrument from the NICHD-SECCYD study (see Table 1) (NICHD Early Child Care Research Network, 1991). Large-scale practicability is fulfilled regarding the discussed points: The NICHD-SECCYD study uses technically mediated observation through video, and analyses are conducted using a macroanalytic rating instrument, which can easily be taught and applied. Furthermore, this method is time-effective because the instrument is designed for rating short video-sequences not exceeding 10 minutes. Additionally, the NICHD-SECCYD study has reported good-quality indicators regarding internal consistency, reliability, concurrent validity, and predictive validity, which are illustrated in the examples below.

The NICHD Child Care Research Network (2005) reports an internal consistency and inter-rater-reliability for the sensitivity composite (subsuming three items; see also Section 4.2) indicated by Cronbach's alpha (α = .75.) and Pearson's correlation coefficient (r = .78, p = n. a.) (Bland, Batten, Appelbaum, Wendell, & NICHD Early Child Care Research Network, 1995). Additionally, the correlation of the sensitivity composite with a positive parenting subscale of the Home Observation for Measurement of the Environment (HOME) Inventory (r = .34, p < .0001) indicates concurrent validity (Bland, Appelbaum, Batten, Wendell, & NICHD Early Child Care Research Network, 1994). In addition, the correlation of the sensitivity composite (averaged repeated measures for 6, 15, and 24 months) with different child outcomes at 36 months signals predictive validity (school readiness: r = .37, p < .001; receptive vocabulary: r = .52, p < .001; social competence: r = .27, p < .01) (NICHD Early Child Care Research Network, 1998). For further impacts on child development, see also NICHD Early Child Care Research Network (1999; 2005).

Table 1 Overview of Observational Instruments for Rating of Parent-Child Interaction

Instrument title	Scales/domains	Estimated duration in min. (interaction/rating)	Published	Training applicable for large-scale studies	Videotaping intended	References
CARE *CARE-Index, infant version*	7 scales *Parental and child behavior:* Facial expression, vocal expression, position and body contact, expression of affect, pacing of turns, control of the activity, developmental appropriateness of the activity	3/10	–	–	✓	Crittenden (2004, 2005); Letourneau & Tryphonopoulos (2012)
CIB *Coding Interactive Behavior*	43 items, divided into 6 domains *Parental behavior:* Sensitivity and responsiveness, intrusiveness *Child behavior:* Positive affect, negative emotionality, initiation, and involvement *Dyad:* Dyadic reciprocity	5/n. a.	–	✓	✓	Feldman (1998); Feldman, Weller, Sirota, and Eidelman (2003)
EA-III *Emotional Availability Scales (3rd edition)*	6 scales *Parental behavior:* Sensitivity, structuring, nonintrusiveness, nonhostility *Child behavior:* Responsiveness to adult, involvement of adult	20/n. a.	–	–	✓	Biringen (1998); Strauß & Schuhmacher (2005)
NCATS *Nursing Child Assessment Teaching Scale*	73 items divided into 6 domains *Parental behavior:* Sensitivity to child's cues, response to child's distress, social-emotional growth-fostering behavior, cognitive growth-fostering behavior *Child behavior:* Child's clarity of cues, responsiveness to the mother	6/n. a.	✓	–	✓	Gross, Conrad, Fogg, Willis, & Garvey (1993); Harrison, Magill-Evans, & Sadoway (2001); Sumner & Spietz (1995)

Instrument title	Scales/domains	Estimated duration in min. (interaction/ rating)	Published	Training applicable for large-scale studies	Video-taping intended	References
NICHD-SECCYD *Rating instrument of mother-child interaction—semi-structured procedure six months home visit*	13 scales *Parental behavior:* Sensitivity to distress, sensitivity to nondistress, intrusiveness, detachment, stimulation of development, positive regard for the child, negative regard for the child, flatness of affect *Child behavior:* Positive mood, negative mood, activity level, sociability, sustained attention	10/10	✓	✓	✓	NICHD Early Child Care Research Network (1991)
MBS-MKI-S *Mannheim Rating system for mother-infant face-to-face interaction*	13 scales *Parental behavior:* Emotion, affectionateness, vocalization, verbal restriction, congruence, variability, reactivity/sensitivity, stimulation *Child behavior:* Emotion/facial expression, vocalization, viewing direction, reactivity, willingness to interact	10/(rating interval 1 min)	✓	✓	✓	Blomeyer et al. (2010); Esser, Scheven, Petrova, Laucht, & Schmidt (1989)

3 Assessing Parent-Child Interaction in the Early Childhood Cohort of the NEPS

After discussing the reasons for the selected design of the assessment and rating of the parent-child interaction, we now specify the form and organization of videotaping and rating the parent-child interaction in the NEPS's Newborn Cohort.

3.1 Assessment of Interaction Sequences

Videotaping of the interaction between the parent (primarily the mother) and his or her child is embedded in a personal interview in the home setting of the family. Parent-child interaction takes place in a semi-standardized play situation. Standardization covers (1) place, (2) play material, and (3) frame of the play situation but does not include strict instruction for interaction. Therefore, the parent is asked to interact with the child as usual. (1) The play situation is carried out on a blanket on the floor, which only serves as a visual localization of the play situation to support the interviewer (for the focus of the camera). (2) Play material included in the NEPS toy set had to meet different criteria regarding type and quality. Considering their type, toys were selected that aimed towards a specific goal of action outcome at different levels of difficulty. As Heckhausen and Heckhausen (2010) point out, some goals of action are more difficult because they demand higher levels of the child's activity regulation than others: Sudden-discrete effects get attention easily and are therefore attractive goals of action for very young children (e. g., squeezing a toy). Continuous effects, which are in conjunction with the action (e. g., the rattle of a car moved back and forth), demand a higher level of self-regulation, whereas stateful goals of a chain of activities are highly demanding because they appear only at the end of an activity (e. g., a finished tower of stacking cups). We selected the number and type of play materials aligned to children's age along this classification of effects (see Table 2). Additionally, we completed this compilation with toys evoking symbolic play and joint-attention episodes. Moreover, the quality of the toys has to be given: First, they had to be age-appropriate (resistant to saliva, not have small parts that can be swallowed); second, they had to offer a seal of quality; and third, they had to be easy to clean with disinfectant wipes because interviewers used the same toys for different households. As in the NICHD-SECCYD study, the framing of the play situation is adapted to the changing requirements of young children throughout their development. The frame of the play situation differs slightly from Wave 1 to Waves 2 and 3. In Wave 1, mothers were asked to play with their infants with five toys of their own for 3 minutes, then for another 5 minutes with toys from the NEPS toy set (see Table 2). In Waves 2 and 3, the observation procedure followed a three-bag procedure in which mothers were asked to play with their children for 10 minutes with toys divided into three bags in a set order (NICHD Early Child Care Research Network, 2005).

Table 2 Play Material for Parent-Child Interaction

	Age of child in months	Sudden-discrete effects	Continuous effects in conjunction with the action	Stateful goal of chain of activities	Symbolic play	Joint attention
Wave 1	*7*	Rattle, squeaking book	Duckling Ball	Stacking cups		
Wave 2	*16*	Squeaking animal	Rattling car	Stacking cups, sorting box	Plates, spoons	Book
Wave 3	*26*	Xylophone	Rattling car	Puzzle	Plates, spoons, animals	Book

The administration of the assessment is conducted by female interviewers in order to provide easier access to the homes of mothers and their 7-month-olds. Interviewer training was provided over several days, focusing on the requirements of the target group and correct assessment.

3.2 Rating of Parent-Child Interaction

Based on the videotaped interaction sequences, the rating of parent-child interaction is conducted by trained coders. Videos of parent-child interaction are delivered to the NEPS and stored in a special room in which access is strictly regulated according to NEPS data-protection standards.

As already described in Section 2, a macroanalytic rating instrument of the NICHD-SECCYD study was chosen for rating the parent-child interaction, which is shown in Table 1. The instrument covers parental and filial interaction style, which can be rated on a 4-point scale ranging from not-at-all characteristic to highly characteristic, supplemented by one missing category. We translated the English version into German, added additional examples for different scale points, and tested the instrument in a feasibility study (n = 20). As in the NICHD-SECCYD study, the raters in the NEPS rate all items after viewing five minutes (or 10 minutes for Waves 2 and 3) of videotaped parent-child interaction. Because of great demands of a highly inferent rating instrument, raters in the NEPS were trained extensively during a 50-hour rater training.

4 Results of Pilot Study Wave 1

In order to provide high-quality data, assessment of videos of interaction sequences in the field and the rating of parent-child interactions are tested extensively at the outset of every main study with the help of diverse pre-studies. For illustration, we concentrate on results of the first pilot study, which took place from October 2011 to January 2012. This pilot study led to different adjustments for the assessment and rating of parent-child interaction in the first main study, with a field time lasting from August 2012 to February 2013.

4.1 Results of Assessment of Interaction Sequences

466 interviews could be realized. The acceptance for participation in videotaped parent-child interaction was very high: 422 participants gave their written consent to be videotaped (90 %); after completion of a videotaped test for competencies, videotaping for parent-child interaction began in 376 parent-infant dyads (80 %) and was completed in 360 cases (77 %). Finally, 170 cases could be analyzed regarding aspects of parent-child interaction (190 interaction sequences were discarded due to different assessment faults, which partly occurred in the same cases). Misframed videos (141 cases) and/or an unfavorable camera setup and location of the play situation (75 cases) constituted the main assessment faults. In most cases, this resulted in videos in which the head or face of the mother or child was not visible for a significant amount of time. Thus, a valid analysis of interactional behavior that also covers facial expressions could no longer be given. Other types of faults in the assessment included an incorrect execution of the play situation, for example, when the mother and child played on a table or a couch instead of on a blanket on the floor (17 cases); when the relevant interactional sequence lay significantly below time limit needed for valid analysis (13 cases); and when technical faults occurred, such as the failure to record sound (7 cases). The reasons for the types of faults were twofold: First, interviewers had to adapt to differing framing conditions. In some cases, home settings were too small for administering the standard setup of the play situation and camera. Second, differences in the performance of the interviewers were noticeable. Presumably based on their previous technical knowledge, some interviewers administered the assessment of their cases nearly free of fault, whereas single interviewers failed at the assessment in a majority of their cases.

4.2 Results of Rating the Parent-Child Interaction

Rating the parent-child interaction was based on the 170 analyzable videotaped play situations. Rating was conducted by two raters (47 % of the videos were coded by

Table 3 Rater Agreement on Composites Level

	Sensitivity composite[a]	Detachment composite[b]
Joint probability in %	52	71
κ	.39***	.41***
ICC[c]	.76***	.53***
r	.76***	.55**

a Includes items: sensitivity to nondistress, positive regard for the child, intrusiveness (reversed score).

b Includes items: detachment, flatness of affect (recoded from 4-point to binary scale).

c Two-way random, nonadjusted; n = 31.

+: p < .1; *: p < .05; **:p < .01; ***:p < .001

Rater 1; 53 % by Rater 2). To check for inter-rater agreement, a double rating of 18 % of the cases was established (n = 31). To provide comparability with the NICHD-SECCYD study, we calculated the same composites of items: a sensitivity composite and a detachment composite (see Table 3). Although we constructed the scale for rating parent-child interactions so as it possesses equal intervals, the scale is in a conservative sense on an ordinal level of measurement. Therefore, we report parametric as well as non-parametric statistics in Table 3. Joint probability and Cohens Kappa indicate a rather poor agreement, with values of 52 % and 71 % as well as $\kappa = .39$ and $\kappa = .41$ (both $p < .001$), respectively[2] (for details, see Table 3). Intra-Class-Correlation (ICC) and Pearson's *r,* show moderate to good agreement, with values of ICC = .76 and ICC = .53 (both $p < .001$) and $r = .76$, $p < .001$ and $r = .55$, $p < .01$, respectively. Results of Pearson's r are comparable with the findings of NICHD-SECCYD, which reports $r = .78$ for the sensitivity composite and $r = .69$ for the detachment composite (Bland et al., 1995).

To enhance the quality of the rating instrument beyond the given results, the identification of the possible clarification of items was necessary. Therefore, a more precise look at rater agreement was required. As Uebersax (2010) points out, disagreement should be treated as a construct that can be subdivided into different components. Accordingly, an index reporting the different components simultaneously in one numerical value, such as the ICC, is not useful for identifying steps to improve agreement. Components of disagreement are mainly based on two different sources: differences between the raters in their trait definition or their definition of specific rating levels (Uebersax, 2010). As a consequence, item-level analysis is conducted for different components of disagreement regarding rater association, rating distribution, and rater bias.

2 All calculations were conducted with SPSS IBM Statistics 19.

For an indication of rater association, we conducted a simple Pearson correlation at the item level. While a majority of items show values between $r = .68$ and $r = 1$ (for p-values, see Table 4), indicating a good agreement, rater agreement for two items is rather poor, with values below $r = .20$. In addition to items with clear trait definition, evidence for a different interpretation of basic constructs or differences in the weight of trait factors is given for single items.

Disagreement can also be based on raters' differences in the definition of rating categories. A test for marginal homogeneity is used for the examination. Marginal homogeneity reflects the similarity of frequencies with which two raters use various rating categories (Uebersax, 2010). Therefore, we included all rated cases and compared marginal frequencies using a Pearson chi-square test. The significance of a single Pearson chi-square test indicates that the rater and distribution are significantly related, which implies differences in frequencies in the use of each rating category. The significance of Pearson's chi-square test is evident for the majority of items, although it only indicates a moderate significance (see Table 5). Therefore, the definition of rating levels should be clarified for the majority of items.

Finally, differences in the interpretation of the calibration of the rating scale could result in disagreement. In addition to other methods, we also tested the tendency to make generally higher or lower ratings with a t-test. Rater bias is displayed by results indicating significant differences between the means of the raters. The majority of items are not biased; only two items display a significant t-test at a 5 %-level (see Table 6).

5 Adaptations and Consequences for Upcoming Waves

Having pointed out potential pitfalls in the implementation of the assessment and rating of parent-child interaction in the first pilot study of NEPS Starting Cohort 1—Early Childhood, we now detail the resulting adaptations regarding the assessment of interaction sequences and the rating of parent-child interaction for the first main study.

First of all, the quality of the videos was enhanced. Adaptions covered interviewer training, the selection of interviewers, and supporting material for interviewers. Interviewer training for video-based assessment was expanded extensively, and an additional hands-on training was established. Furthermore, as interviewers differed in their number of faults in data collection, the selection of interviewers is now conducted based on a test-assessment. Additionally, the interviewer manual has been adapted, and a short pictorial instruction is now also handed out to support the assessment process. The implementation of these adjustments was also conducted for the pilot study assessment in Wave 2: Here, only two interaction sequences were distorted due to faults in assessment ($n = 64$).

Table 4 Rater Association on Item-Level; Pearson's r

	M1	M2	M3	M4	M5	M6	M7	M8	C1	C2	C3	C4	C5
r	1.00	.47**	.74**	.70***	.48**	.68***	–	.48**	.20	.74***	.74***	.43*	.08

Parental behavior: sensitivity to distress (M1), sensitivity to nondistress (M2), intrusiveness (M3), detachment (M4), stimulation of development (M5), positive regard for the child (M6), negative regard for the child (M7), flatness of affect (M8); child behavior: positive mood (C1), negative mood (C2), activity level (C3), sociability (C4), sustained attention (C5); n = 31; for M1, n = 2 (M1 can only be rated if child displays distress); +: p < .1; *: p < .05; **:p < .01; ***:p < .001

Table 5 Rating Distribution on Item-Level, Chi-Square Test

	M1	M2	M3	M4	M5	M6	M7	M8	C1	C2	C3	C4	C5
X^2	3.75	6.01$^+$	8.66*	6.17*	7.08$^+$	10.46**	2.33	8.06$^+$	8.35$^+$	9.97*	2.39	11.00**	3.84
df	3	2	2	2	3	3	1	3	3	3	2	3	2

Parental behavior: sensitivity to distress (M1), sensitivity to nondistress (M2), intrusiveness (M3), detachment (M4), stimulation of development (M5), positive regard for the child (M6), negative regard for the child (M7), flatness of affect (M8); child behavior: positive mood (C1), negative mood (C2), activity level (C3), sociability (C4), sustained attention (C5); differences in degrees of freedom result from single scale levels not being used. n = 170; for M1 n = 10. (M1 can only be rated if child displays distress); $^+$: p < .1; *: p < .05; **: p < .01

Table 6 Rater Bias on Item Level; t-Test

		M1	M2	M3	M4	M5	M6	M7	M8	C1	C2	C3	C4	C5
Rater 1	*M*	2.50	3.16	1.42	1.48	2.58	3.23	1.00	1.84	2.10	1.35	2.29	2.71	2.52
	SD	2.12	0.69	0.67	0.68	0.72	0.67	0.00	0.74	0.54	0.80	0.53	0.69	0.57
Rater 2	*M*	3.00	3.16	1.26	1.26	2.45	3.10	1.00	2.23	2.06	1.16	2.23	2.77	2.58
	SD	1.41	0.64	0.51	0.58	0.57	0.54	0.00	0.76	0.25	0.52	0.43	0.56	0.50
	t	−1.00$^+$	0	1.98$^+$	2.53*	1.10	1.44	–	−2.83**	0.33	1.99	1.00	−0,53	−0.49
	df	1	30	30	30	30	30	30	30	30	30	30	30	30

Parental behavior: sensitivity to distress (M1), sensitivity to nondistress (M2), intrusiveness (M3), detachment (M4), stimulation of development (M5), positive regard for the child (M6), negative regard for the child (M7), flatness of affect (M8); child behavior: positive mood (C1), negative mood (C2), activity level (C3), sociability (C4), sustained attention (C5); *n* = 31; for M1 n = 2 (M1 can only be rated if child displays distress); $^+$: *p* < .1; *: *p* < .05; **:*p* < .01

The rating of parent-child interaction was modified regarding both the instrument itself and rating processes. The rating manual of the instrument was restructured. While the definition of trait and construct remained unmodified, the structure of each item and item-level description were unified. Additionally, we accommodated the fact that the trait is more continuous than discrete and expanded the rating scale from four to five levels, thereby providing a detailed description and example for each item level. These adaptions resulted in an adapted version of the rating instrument from the NICHD-SECCYD study, which is used for rating videotaped interactions in the first main study (Sommer & Mann, 2015).

Second, the rating process itself was adjusted: To avoid observer drift, the duration of the period of rating is kept as low as possible, and regular refreshment-trainings during the rating are conducted in addition to the rater training.

For the implementation of a video-based assessment and the rating of parent-child interaction in large-scale studies, different challenges had to be faced. After testing the assessment and rating in first pilot study in Wave 1 and identifying potential pitfalls, we made different adjustments for the main study in Wave 1 and subsequent waves. By assessing and rating parent-child interaction in the first main study (N = 3,481), the NEPS will gather substantiate information about educationally relevant processes in familial learning environments. The data were released in 2015 in a Scientific Use File.

References

Aßmann, C., Steinhauer, H. W., Kiesl, H., Koch, S., Schönberger, B., Müller-Kuller, A., … Blossfeld, H.-P. (2011). Sampling designs of the National Educational Panel Study: Challenges and solutions. In H.-P. Blossfeld, H.-G. Roßbach, & J. von Maurice (Eds.), *Zeitschrift für Erziehungswissenschaft, 14. Education as a lifelong process: The German National Educational Panel Study (NEPS)* (pp. 51–65). Wiesbaden: VS Verlag für Sozialwissenschaften.

Bäumer, T., Preis, N., Roßbach, H.-G., Stecher, L., & Klieme, E. (2011). Education processes in life-course-specific learning environments. In H.-P. Blossfeld, H.-G. Roßbach, & J. von Maurice (Eds.), *Zeitschrift für Erziehungswissenschaft, 14. Education as a lifelong process: The German National Educational Panel Study (NEPS)* (pp. 87–101). Wiesbaden: VS Verlag für Sozialwissenschaften.

Belsky, J., Vandell, D., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., Owen, M., & The NICHD Early Child Care Research Network. (2007). Are there long-term effects of early child care? *Child Development, 78*(2), 681–701.

Biringen, Z. (1998). *Emotional Availability Scales* (3rd ed.). Unpublished manual, Center for family studies, Colorado State University, Fort Collins.

Bland, S., Appelbaum, M., Batten, D. A., Wendell, C., & NICHD Early Child Care Research Network. (1994). *Correlations between HOME-HOME and structured interac-*

tion variables (Child Care Data Report—8b). Nashville: Peabody College, Vanderbilt University.

Bland, S., Batten, D. A., Appelbaum, M., Wendell, C., & NICHD Early Child Care Research Network. (1995). *Mother-child-interaction—Mother variables. Six months reliabilities* (Child care data addendum—8c). Nashville: Peabody College, Vanderbilt University.

Blomeyer, D., Laucht, M., Pfeiffer, F., & Reuß, K. (2010). *Mutter-Kind-Interaktion im Säuglingsalter, Familienumgebung und Entwicklung früher kognitiver und nicht-kognitiver Fähigkeiten: Eine prospektive Studie.* Mannheim: Zentrum für Europäische Wirtschaftsforschung.

Bornstein M. H., Hahn C.-S., Suwalsky J. T., & Haynes, O. M. (2011). Maternal and infant behavior and context associations with mutual emotion availability. *Infant Mental Health Journal, 32*(1), 70–94.

Bornstein, M. H., & Tamis-LeMonda, C. S. (1989). Maternal responsiveness and cognitive development in children. In M. H. Bornstein (Ed.), *Maternal responsiveness: Characteristics and consequences. New directions for child development* (pp. 49–61). San Francisco, CA: Jossey-Bass.

Bromley, C. (2009). *Growing up in Scotland: The impact of children's early activities on cognitive development.* Edinburgh: Scottish Government.

Crittenden, P. M. (2004). *CARE-Index: Coding manual.* Unpublished manual.

Crittenden, P. M. (2005). Der CARE-Index als Hilfsmittel für Früherkennung, Intervention und Forschung. *Frühförderung Interdisziplinär, 24*(3), 99–106.

Esser, G., Scheven, A., Petrova, A., Laucht, M., & Schmidt, M. H. (1989). Mannheimer Beurteilungsskala zur Erfassung der Mutter-Kind-Interaktion im Säuglingsalter (MBS-MKI-S). *Zeitschrift für Kinder- und Jugendpsychiatrie, 17,* 185–193.

Faßnacht, G. (1995). *Systematische Verhaltensbeobachtung. Eine Einführung in die Methodologie und Praxis.* München: Ernst Reinhardt Verlag.

Feldman, R. (1998). *Coding interactive behavior manual.* Unpublished manual, Bar-Ilan University, Ramat-Gan, Israel.

Feldman, R., Weller, A., Sirota, L., & Eidelman, A. I. (2003). Testing a family intervention hypothesis: The contribution of mother—infant skin-to-skin contact (kangaroo care) to family interaction, proximity, and touch. *Journal of Family Psychology, 17*(1), 94–107.

Greve, W., & Wentura, D. (1997). *Wissenschaftliche Beobachtung. Eine Einführung.* Weinheim: Beltz Psychologie Verlags Union.

Gross, D., Conrad, B., Fogg, L., Willis, L., & Garvey, C. (1993). What does the NCATS measure. *Nursing Research, 42*(5), 260–265.

Halle, T., Forry, N., Hair, E., Perper, K., Wnadner, L., Wessel, J., & Vick, J. (2009). *Disparities in early language development: Lessons from the Early Childhood Longitudinal Study—Birth Cohort (ECLS-B).* Washington, DC: Child Trends.

Harrison, M., Magill-Evans, J., & Sadoway, D. (2001). Scores on the Nursing Child Assessment Teaching Scale for father-toddler dyads. *Public Health Nursing, 18*(2), 94–100.

Heckhausen, J., & Heckhausen, H. (2010). Motivation und Entwicklung. In J. Heckhausen, & H. Heckhausen (Eds.), *Motivation und Handeln* (pp. 427–488). Berlin: Springer.

Hillemeier, M. M., Farkas, G., Morgan, P. L., Martin, M. A., & Maczuga, S. A. (2009). Disparities in the prevalence of cognitive delay: How early do they appear? *Paediatric and Perinatal Epidemiology, 23*(3), 186–198.

Hirschmann, N., Kastner-Koller, U., Deimann, P., Aigner, N., & Svecz, T. (2011). INTAKT: A new instrument for assessing the quality of mother-child interactions. *Psychological Test and Assessment Modeling, 53*(3), 295–311.

Leerkes, E. M., Blankson, A. N., & O'Brien, M. (2009). Differential effects of maternal sensitivity to infant distress and nondistress on social-emotional functioning. *Child Development, 80*(3), 762–775.

Letourneau, N., & Tryphonopoulos, P. (2012). Der CARE-Index: Ein Instrument zur Erfassung der Beziehungsqualität zwischen Bezugsperson und Kind ab der Geburt. In M. Stokowy, & N. Sahhar (Eds.), *Bindung und Gefahr. Das Dynamische Reifungsmodell der Bindung und Anpassung* (pp. 19–32). Gießen: Psychosozial-Verlag.

NICHD Early Child Care Research Network. (1991). *NICHD study of early child care: Volume II: 5 month manual, 6 month manuals, time use manuals*. Unpublished manuscript, NICHD Study of Early Child Care.

NICHD Early Child Care Research Network. (1998). Relations between family predictors and child outcomes: Are they weaker for children in child care. *Developmental Psychology, 34*(5), 1119–1128.

NICHD Early Child Care Research Network. (1999). Child care and mother—child interaction in the first 3 years of life. *Developmental Psychology, 35*(6), 1399–1413.

NICHD Early Child Care Research Network. (2002). Child care structure—Process—Outcome: Direct and indirect effects of child care quality on young children's development. *Psychological Science, 13*(3), 199–206.

NICHD Early Child Care Research Network. (2005). *Child care and child development. Results from the NICHD Study of Early Child Care and Youth Development*. New York: Guilford.

Page, M., Wilhelm, M. S., Gamble, W. C., & Card, N. A. (2010). A comparison of maternal sensitivity and verbal stimulation as unique predictors of infant social-emotional and cognitive development. *Infant Behavior and Development, 33*(1), 101–110.

Rentzsch, K., & Schütz, A. (2009). *Psychologische Diagnostik. Grundlagen und Anwendungsperspektiven*. Stuttgart: Kohlhammer.

Schlesiger C., Lorenz, J., Weinert, S., Schneider, T., & Roßbach, H.-G. (2011). From birth to early child care. In H.-P. Blossfeld, H.-G. Roßbach, & J. von Maurice (Eds.), *Zeitschrift für Erziehungswissenschaft, 14. Education as a lifelong process: The German National Educational Panel Study (NEPS)* (pp. 187–202). Wiesbaden: VS Verlag für Sozialwissenschaften.

Sommer, A., & Mann, D. (2015). *Qualität elterlichen Interaktionsverhaltens. Erfassung von Eltern-Kind-Interaktionen mithilfe eines makroanalytischen Ratinginstruments im Na-*

tionalen Bildungspanel. (NEPS Working Paper No. 56). Bamberg: University of Bamberg, National Educational Panel Study.

Strauß, B., & Schuhmacher, J. (Eds.). (2005). *Klinische Interviews und Ratingskalen.* Göttingen: Hogrefe.

Sumner, G., & Spietz, A. L. (1995). *NCAST caregiver/parent-child interaction teaching manual* (2nd ed.). Seattle, WA: NCAST Publications, University of Washington.

Uebersax, J. (2010). *Statistical methods for rater and diagnostic agreement.* Retrieved from http://john-uebersax.com/stat/agree.htm

Wiefel, A., Titze, K., Kuntze, L., Winter, M., Seither, C., Witte, B., … Lehmkuhl, U. (2007). Diagnostik und Klassifikation von Verhaltensauffälligkeiten bei Säuglingen und Kleinkindern von 0–5 Jahren. *Praxis der Kinderpsychologie und Kinderpsychiatrie, 56*(1), 59–81.

About the authors

C. Hachul
University of Bamberg, Bamberg.

H.-G. Roßbach
Leibniz Institute for Educational Trajectories (LifBi), Bamberg.
Chair of Early Childhood Education, University of Bamberg, Bamberg.

A. Sommer
Chair of Early Childhood Education, University of Bamberg, Bamberg.
e-mail: anja.sommer@uni-bamberg.de