

Parenting Quality from Observational Ratings at Age 2:

Validation from Norwegian and U.S. Samples

Kristin Berg Nordahl^{1*}, Margaret Tresch Owen², Luisa Antunes Ribeiro¹³,

and Henrik Daae Zachrisson¹⁴

Nordahl, K. B., Owen, M. T., Ribeiro, L. A., & Zachrisson, H. D. (2020). Parenting quality from observational ratings at age 2: Validation from Norwegian and US samples. *Early Childhood Research Quarterly*, 53, 379-390.

¹The Norwegian Center for Child Behavioral Development, Post Office Box 7053, Essendropsgate 3, Majorstuen, 0306 Oslo, Norway

²The University of Texas at Dallas, School of Behavioral and Brain Sciences, GR41 800 W. Campbell Road, Richardson, TX 75080-3021

³Faculty of Education and Psychology, Research Center for Human Development, Universidade Católica Portuguesa, Rua Diogo Botelho 1327, 4169-005 Porto, Portugal

⁴ University of Oslo, Department of Special Needs Education, Postboks 1140, Blindern, Helga Engs hus, Sem Sælands vei 7, 0371 Oslo, Norway

* Corresponding author

Kristin Berg Nordahl: kristin.nordahl@nubu.no

This work was supported by the Research Council of Norway (grant # 237944 & 276672)

PARENTING QUALITY CONSTRUCT VALIDITY

Abstract

This study investigated the construct validity of a proposed measure of parenting quality derived from extensively used observational ratings of parenting in mother-child interaction procedures with 2-year-olds in two large samples. Data included global ratings of mother-child interaction in an unstructured free-play and a semi-structured teaching task from the Behavior Outlook Norwegian Developmental Study ($N=1,157$) and from the Three Boxes procedure used in the NICHD Study of Early Child Care and Youth Development ($N=1,364$). Confirmatory Factor Analyses, including ratings of mothers' sensitivity, detachment, intrusiveness, cognitive stimulation, positive regard and negative regard, revealed a similarly structured latent parenting construct across samples and observational procedures, supporting the structural and content validity, and the generalizability of the measure. In the Norwegian sample, better predictive validity to child outcomes at age 4 was found from the semi-structured teaching task than from the less structured free-play task. Comparable predictive validity to child outcomes was found for the U.S. sample's latent parenting quality construct. The results hold implications for the careful selection and study of observational procedures for measuring parenting quality in early childhood.

Keywords: parenting quality, construct validity, observational methods, structured interaction, measurement model

PARENTING QUALITY CONSTRUCT VALIDITY

Parenting Quality from Observational Ratings at Age 2: Validation from Norwegian and U.S. Samples

A large research literature relies on observational measures from mother-child interaction to address how parenting quality in early childhood relates to child developmental outcomes across multiple domains (e.g., Bocknek, Brophy-Herb, & Banerjee, 2009; Dyer, Owen, & Caughy, 2014; Mills-Koonce et al., 2015; NICHD Early Child Care Research Network (ECCRN), 2004). Observational ratings of parenting such as sensitive responsiveness, positive involvement, warmth, and stimulation of development, have been associated with language, cognitive, social, and behavioral child outcomes (e.g., Barnett, Shanahan, Deng, Haskett, & Cox, 2010; Pungello, Iruka, Dotterer, Mills-Koonce, & Reznick, 2009; Tamis-LeMonda, Shannon, Cabrera, & Lamb, 2004; Whiteside-Mansell, Bradley, Owen, Randolph, & Cauce, 2003). Given the important role of parenting to children's development and the heavy reliance on observational measures of parenting quality, careful systematic psychometric evaluation and validation of such measures is critical.

The aim of the present study was to use state-of-the-art psychometric methods to evaluate the construct validity of a proposed latent measure of parenting quality derived from global ratings of mothers' interactions with their 2-year-olds in two large, community-based samples in Norway and in the U.S., respectively. A number of observational rating systems exist to capture aspects of parenting similar or closely related to the above mentioned items from parent-child interaction, e.g., the Early Parenting Coding System (Winslow & Shaw, 1995), the Parent-Child Interaction System (Deater-Deckard, Pylas, & Petrill, 1997), the Parenting Assessment Battery (Wilson & Durbin, 2012), the Parent-Child Early Relational Assessment (Clark, 1999), and the Emotional Availability Scales (Biringen, Robinson, & Emde, 1998). The observational ratings examined in the present study were based on the rating system used in infancy and early childhood in the NICHD SECCYD (NICHD ECCRN,

PARENTING QUALITY CONSTRUCT VALIDITY

1999). Comprising parents' sensitivity and responsiveness, intrusiveness, detachment or disengagement, cognitive stimulation, and positive regard and negative regard for the child, these rating items have been used extensively in studies across multiple samples examining determinants of parenting, child outcomes related to parenting qualities, or controlling for experiences of parenting in studies of effects of other environmental contexts. The ratings have been used in different observational procedures, ranging from unstructured free-play situations to more structured goal-oriented tasks, and lasting from 5 minutes or less to 15 minutes or more (e.g., Maas, Vreeswijk, & van Bakel, 2013; Mills-Koonce et al., 2015; NICHD ECCRN, 2003; Towe-Goodman, Willoughby, Blair, Gustafsson, Mills-Koonce, & Cox, 2014). Various higher-order parenting quality constructs have been derived from these ratings using different approaches, including most typically a priori composites (e.g., Mills-Koonce et al., 2015; NICHD ECCRN, 1999; 2004; Peredo, Owen, Rojas, & Caughy, 2015). Nevertheless, despite widespread use of the parenting rating items examined in the current study, the literature contains little if any validation of a latent parenting quality construct or constructs derived from these ratings.

The inclusion of the two databases allows us to address whether the structure of the measurement model found in Norway is sensitive to the specifics of its particular sample of Norwegian families or is more broadly characteristic of parenting quality found from the measurement model identified in a large U.S. sample. By replicating our model and the predictions made from it in a U.S. context we strengthen the case for comparability of research findings using observational methods to measure parenting quality across these contexts. Given that Norway and the U.S. can be considered polar opposites among developed countries with regard to family friendly policies (e.g., paid parental leave and early childhood education), this may also strengthen the general case that the observational measures of parenting can be compared across sociopolitical contexts.

PARENTING QUALITY CONSTRUCT VALIDITY

The Global Rating System

The six observational parenting items utilized in the current study include the following: 1) The *sensitivity/responsiveness* rating item was based on Mary Ainsworth's observational rating item (see Mesman & Emmen, 2013) and reflects how the parent monitors and responds to the child's behavior and signals, including social cues and signs of positive as well as negative affect. Sensitive parents are tuned in to the child's activities and needs and respond appropriately to the child's behavior and expressions. 2) *Intrusiveness* reflects the parent's attempts to control the child's behavior based on the parent's agenda rather than recognizing and respecting the validity of the child's perspective. 3) *Detachment/disengagement* is the degree to which the parent lacks emotional involvement or interest in and involvement with the child and his/her activities. 4) *Cognitive stimulation* is the degree to which the parent's input serves to facilitate the child's learning. 5) *Positive regard* for the child is the expression of positive feelings toward the child, shown by warmth, physical affection, smiles, sharing laughter, enthusiasm, praise, and enjoyment. 6) Markers of *negative regard* for the child include disapproval, body tenseness, negativity when correcting, abruptness, tense facial muscles and strained expression, harshness, and non-playful teasing.

Interaction Procedures

Observational ratings of parenting with infants and toddlers are typically collected in parent-child interaction procedures set up with a standardized set of toys and instructions to ensure occurrence of behaviors of interest. Variation in the degree of induced procedural structure to the interaction tasks may produce a difference that matters to the validity of parenting measures derived from the ratings and thus to predictions thereof. A Three Boxes procedure adapted from Vandell (1979) was used at 15, 24, and 36 months in the NICHD SECCYD (see NICHD ECCRN, 1999; 2004) and has been adopted with minimal variations by others (e.g., Berlin, Brady-Smith & Brooks-Gunn, 2002; Bocknek et al., 2009). In this

PARENTING QUALITY CONSTRUCT VALIDITY

semi-structured procedure, the parent-child dyad is presented with three numbered containers, each holding an attractive toy or a book. Instructions to the parents are that the contents of each container should be shown to the child in the order indicated on the container, and the parent is in control of the amount of time spent with each (see Martin, Ryan, & Brooks-Gunn, 2007; NICHD ECCRN, 1999; Peredo et al., 2015). Other interaction procedures with more or less structure have been adopted in studies using the same parent-child rating items. For example, a minimum of structure is provided in free-play tasks, for which the parent and child are typically provided an array of attractive toys (e.g., Bakermans-Kranenburg, van Ijzendoorn, & Kroonenberg, 2004; Burchinal, Vernon-Feagans, Cox, and Key Family Life Project Investigators, 2008; Propper, Willoughby, Halpern, Carbone, & Cox, 2007) and asked to play as they choose. Tasks with either an implied or specified goal to be achieved (e.g., clean up or teaching tasks) typically involve more structure, such as assisting the child in solving a puzzle or set of puzzles, usually too difficult for the child to complete alone (e.g., Mills-Koonce et al., 2015; Pungello et al., 2009; Whiteside-Mansell et al., 2003) have also been utilized.

Although variability in task structure is found across studies, there is little testing of whether validity of the ratings or outcomes differ by task structure. The few studies that have explicitly examined effects of task differences have focused on how task variations relate to differences in parent behaviors. Results show that more structured tasks tend to yield more directive parenting behavior and higher levels of conflict than less structured tasks (e.g., Ginsburg, Grover, Cord, & Ialongo, 2006; Panfile, Laible, & Eye, 2012), while less structured free-play observations generally yield more playful and child-oriented behavior (e.g., Kwon, Bingham, Lewsader, Jeon, & Elicker, 2012). Despite fairly consistent differences in parenting behavior related to procedural structure, stability of individual differences across tasks may be expected. That is, a parent may be both more directive in a

PARENTING QUALITY CONSTRUCT VALIDITY

teaching task than in a free-play task and be relatively more directive than other parents in both situations (see Gonzalez, 1996; Maas et al., 2013). However, while there is some evidence that parents behave differentially, conditional on procedure, there is no strong evidence to suggest whether ratings of parenting behavior from different procedures have differential predictive validity to child outcomes. In other words, do procedural differences matter when it comes to predictions of child outcomes from the ratings made? In the current study, we tested if latent parenting quality constructs derived from ratings made from three slightly different mother-child interaction procedures at age 2 were predictive of comparable child outcomes approximately two years later.

Measure Derivation

As noted earlier, the formation of a priori composites of all or a subset of the rating items is a commonly used method for forming a higher order parenting construct from individual ratings, but empirical justification for the aggregates is not always made explicit. Some studies report that composites were formed a priori based on theory or former studies (e.g. Bakermans-Kranenburg et al., 2004; NICHD ECCRN, 1999), while others based their composite on results from exploratory or confirmatory factor analyses (e.g., Barnett et al., 2010; Propper et al., 2007; Pungello et al., 2009; Towe-Goodman et al., 2014).

Confirmatory approaches to form latent variables from the rated parenting items have been relatively infrequent, although there are a few notable exceptions. Through confirmatory factor analyses (CFA), Whiteside-Mansell and colleagues (2003) found that a model of three parenting factors best fit the parenting ratings derived from both the 15-min Three Boxes procedure and a 5-min clean-up activity collected at age 3 years in the NICHD SECCYD. Latent variables of responsive, intrusive and harsh parenting were used in the subsequent analyses of relations to child outcomes, finding generally similar patterns for European American and African American mothers, both in the latent parenting constructs and in

PARENTING QUALITY CONSTRUCT VALIDITY

relations between the parenting constructs and child behavior outcomes. Also based on CFAs, Mills-Koonce and colleagues (2015) formed a single latent sensitive parenting variable for mothers and for fathers from ratings of parental detachment (reversed), positive regard, stimulation, and animation observed during 10-min free play observations at 6 months and a 10-min teaching task at 2 years. Consistent associations for mothers and fathers were found between this sensitive parenting latent construct and child cognitive development at age 3 (Mills-Koonce et al., 2015).

Construct Validity

Validity of a measure is not a property of the measure itself, but of the interpretation and meaning of the measure (Cronbach, 1971). Extending Cronbach's argument, Messick (1995) argued that all available information, including validity and reliability, contribute to the interpretation of the meaning of a measure within a context and for a particular purpose (Messick, 1995). Thus, construct validity is an evolving process. Messick points to six aspects of construct validity that should be conceived as a heuristic for test validation, rather than a checklist: 1) the *content* aspect (evidence of content relevance); 2) the *substantive* aspect (theoretical rationale for observed consistencies in test scores); 3) the *structural* aspect (consistency between expected and observed structure of the construct); 4) the *generalizability* aspect (across raters, populations, and settings); 5) the *external* aspect (convergent and discriminant evidence); and 6) the *consequential* aspect (valid use of the assessment). Drawing on more recent development in statistical modeling, John & Benet-Martínez (2000) expand Messick's conceptualization of structural quality using Confirmatory Factor Analysis (CFA; Brown, 2015) as a comprehensive approach for test validation. This allows for testing whether a hypothesis about the meaning of the structure in the data reflects the observed structure in the data. CFA allows for specific hypotheses testing of a theoretical measurement model, and it provides an empirical strategy to test the validity of both the

PARENTING QUALITY CONSTRUCT VALIDITY

proposed structure and the theoretical assumptions about the indicators. In the current study, we propose a latent parenting quality construct from the six globally rated parenting items and use CFA to test the structural and content validity (the latter being related to whether specific items reflect the content of the latent construct) of our measurement model. While parenting is often conceived as a multifaceted construct in theory, the ratings from parent-child observations are recurrently combined into a composite score. As a composite score implies uni-dimensionality, this notion is pertinent to test, because uni-dimensionality imposed on a multi-dimensional construct yields lack of measurement precision and obscures interpretation of estimates.

Making assumptions of the *structural* and *content* validity explicit through latent modeling enhances the potential for replications between samples and for international comparisons. Both of these venues for further research are widely accepted in the field. Replications of measurement models in different datasets also address *generalizability*. First, replication represents an important step towards identifying general versus sample specific features of the proposed measure, and thus enhances interpretation of the measure in context. For example, as in the current study, if two similar but not identical observational procedures are used in two different samples, configural invariance (i.e., similar factor structure of underlying parenting ratings) can be tested to investigate whether the same broad parenting constructs are measured. Furthermore, this approach has the advantage of making inferences of relations between parenting and other constructs (e.g., child social and emotional functioning) more comparable, and hence replicable. This also extends to doing international comparisons, an important direction of research to enhance understanding of child development in context (Lansford et al., 2016). The general cultural climate moderates parenting effects (Lansford & Bornstein, 2011). In diverse domains of development, parenting practices relate differently to children's adjustment depending on the broader

PARENTING QUALITY CONSTRUCT VALIDITY

cultural context. Thus, there may not be invariance across samples that vary in terms of culture. A careful approach to measurement is required to make valid inferences about cross-country similarities and differences in parenting, as well as predictors and outcomes of parenting. As pointed out by Lansford and colleagues (2016), establishing some level of measurement invariance is challenging, but necessary to move international comparisons beyond correlations between measures with uncertain meaning.

Finally, the *external* aspect of validity entails some of the thinking implied in the construct of predictive validity (Cronbach & Meehl, 1955). The meaning of a construct may partly be interpreted from its relation to other constructs. Specifically, the interpretation of the meaning of a parenting quality measure is, in part, a function of whether it varies in predicted ways with both family demographics and child outcomes. In line with Belsky's model of the determinants of parenting (Belsky, 1984; Belsky & Jaffee, 2006), which holds that child, parent, and contextual characteristics influence parenting, we included child gender and child global development, maternal age, education, mental distress, and single parenthood, and family economic hardship, as hypothesized predictors of parenting quality, given considerable literature indicating meaningful differences in parenting qualities associated with these characteristics of the child, parent, and family context (e.g., Berlin et al., 2002; Tamis-LeMonda et al., 2004). To investigate the predictive validity of our parenting quality measure we examined a selection of child outcomes at age 4 frequently associated with parenting quality in the literature, including effortful control (e.g., Bocknek et al., 2009; Klein et al., 2018); receptive language (e.g., Pungello et al., 2009; Vallotton, Mastergeorge, Foster, Decker, & Ayoub, 2017), and externalizing problem behaviors (e.g., Barnett et al., 2010; Garai, Forehand, & Colletti, 2009).

The Current Study

PARENTING QUALITY CONSTRUCT VALIDITY

In this study, we took an integrated approach to address multiple aspects of the *construct validity* of an observation-based measure of parenting quality by:

1) Testing the *structural* and *content* validity of a proposed measure of parenting quality within a Norwegian sample, based on global ratings of maternal sensitivity, intrusiveness, detachment, cognitive stimulation, positive regard and negative regard for the child in mother-child interactions at age 2. We also examined the *generalizability* of this construct across two differently structured interaction procedures with the same sample and by replicating the analytic approach taken using similar observational data from the NICHD Study of Early Child Care and Youth Development's interaction procedure collected at age 2.

2) Testing the *external* and *predictive* validity of the latent parenting quality construct by examining its association with demographic and child characteristics measured in infancy and with multiple child outcomes at age 4. We also examined whether these associations varied as a function of the different observational procedures used with the Norwegian sample and whether similar associations could be found in the NICHD SECCYD data collected at comparable ages.

Method

The two databases included in the study are comparably large longitudinal studies that collected observational ratings of mother-child interactions at age 2 utilizing similar sets of rating items addressing qualities of mother-child interaction, child and family predictors of individual differences in the parenting quality, and child outcomes at a subsequent age prior to entry into grade school.

Samples and Procedures

The *Behavior Outlook Norwegian Developmental Study* (BONDS) is a longitudinal study tracking 1157 children's (51.8% boys) social development from 6 months (for detailed description, see Nærde, Janson, & Ogden, 2014). Recruitment took place 2006–2008 through

PARENTING QUALITY CONSTRUCT VALIDITY

public child health clinics in five Norwegian municipalities. Parents of 1,931 eligible children (approximately 6 months old, and at least one Norwegian speaking parent) were informed about the study, 1,465 (76%) agreed to be contacted, and of these 1,159 (60%) agreed to participate. The overall retention rate was very high: 1,098 (95%) families participated at the age 2 follow-up and 1,093 (93%) at age 4. The final sample was fairly representative of the Norwegian population, although somewhat biased toward mothers with higher education, fewer immigrant parents, more firstborns, and fewer single mothers. The recruited sample included 6.8% immigrants from Western Europe or the United States and 6.4% Non-western immigrants. For the present study, family and child characteristics were measured in lab and home visits when the children were 6 and 12 months old, structured mother-child interactions were video recorded at age 2, and child outcomes were measured at 4 years. Data were obtained from observations, interviews, questionnaires, and assessments of children's cognitive, language, and social development. Mothers of 934 (85%) children (of 1,098) agreed to participate in the structured interactions at age 2. Twenty-six of the recordings were not coded due to poor technical quality or incomprehensible language. Thus, the final sample consisted of 908 mothers and their 2-year old children (50.6% boys).

The *NICHD Study of Early Child Care and Youth Development* (SECCYD) was a longitudinal study of 1,364 children (51.7% boys) recruited in 1991 from hospitals located at 10 sites across the United States. Consents for participation were obtained from mothers in home visits when the infants were 1 month of age (for more detailed information on recruitment, see NICHD ECCRN, 1999). The study's recruited sample was not nationally representative, but it was similar with respect to demographic variables such as ethnicity and household income in 1991. The recruited families included 24% ethnic-minority children (non-White or Hispanic). Family and child characteristics for the present study were measured in home and lab visits when the children were 1, 6, and 15 months, and 2 years old,

PARENTING QUALITY CONSTRUCT VALIDITY

and child outcomes were measured at 4.5 years. The data were collected through observations, interviews, questionnaires, and assessments of children's cognitive, language, and social development. For the present study, the analysis sample included the 1,172 child participants with mother-child interaction ratings collected at age 2.

Mother-Child Interaction

The mother-child interaction procedure collected at age 2 differed across the two samples. For the BONDS study, five different interaction tasks were selected based on their capacity to elicit parent and child behaviors associated with later child adjustment (e.g., Aspland & Gardner, 2003; Snyder, Stoolmiller, Wilson, & Yamamoto, 2003). At age 2, 16 minutes of mother-child interaction, video-recorded during visits to study offices, consisted of 1) free-play (4 min), 2) clean-up (2 min), 3) teaching (6 min), 4) inhibition (2 min), and 5) waiting (2 min). The current global ratings were applied to the 4-min unstructured free-play and the 6-min semi-structured teaching tasks. Thus, the present analyses utilized data from these procedures. In the free-play task, mothers were asked to play with their child, as they liked with a set of attractive toys. For the semi-structured teaching task, mother-child dyads were presented with a puzzle and subsequently with a set of shape sorting blocks; mothers were instructed to help their child as much as they thought necessary spending 3 minutes with each toy. An interviewer informed the mother when it was time to switch from the puzzle to the shape sorter. The toys for the teaching task were selected to be too difficult for most 2-year-olds to manage on their own. Prior to the observations, mothers were informed that they could choose to discontinue the tasks at any time.

The SECCYD mother-child interaction procedure collected at age 2 took place in lab visits (NICHD ECCRN, 1999). In the 15-min semi-structured Three Boxes procedure, mothers were asked to show their children the contents of each container, in a fixed order as designated. Decisions about the time spent with each bag were determined by the mother

PARENTING QUALITY CONSTRUCT VALIDITY

based upon her reading of the child's interests and her own agenda. A storybook was in the first bag, a toy stove and related objects were in the second bag, and a simple toy house with a few moving parts and two figures. The box contents were chosen to stimulate parent-child book reading, play, and teaching. Although not a teaching-task per se, we argue that the instructions to show the child the contents of each box in turn, carries an implication that the interaction include teaching, and thereby bears more similarity to the BONDS teaching task than to its free-play task that included a wide array of toys laid out across the floor.

Measures

Maternal parenting quality rating items and procedures. The mother-child interaction videos from the BONDS and the SECCYD studies were rated using the same global rating system. For both samples, the parenting quality rating items included: 1) sensitivity/responsiveness, 2) detachment/disengagement, 3) intrusiveness, 4) cognitive stimulation, 5) positive regard and 6) negative regard for the child (for item descriptions, see introduction). A 5-point modification of the rating scales from 1 (*not at all characteristic*) to 5 (*highly characteristic*) (Cox, 1997; Owen et al., 2010) was used with the BONDS interaction videos; the original 4-point scales (1=*not characteristic* to 5=*highly characteristic*) were used at 2 years in the SECCYD (NICHD ECCRN, 1999). The 5-point scales have been used in a large number of studies (e.g., Owen, Caughy, Hurst, Amos, Hasanizadeh, & Mata-Otero, 2013; Mills-Koonce et al., 2015; Wright, Hill, Sharp, & Pickles, 2018). The observational rating system, in its 5-point scale version, was used in the current Norwegian study because of the sizable literature stemming from the SECCYD providing evidence of significant relations between the rating items and a multitude of child outcomes (see NICHD ECCRN, 2001, for an early summary) and the growing use of the 5-point scale versions of the rating items.

PARENTING QUALITY CONSTRUCT VALIDITY

Rating scale descriptions of each point on the scale for each parenting item describe levels based on both quantity and quality. Behaviors coded as indicators of rating items may differ slightly based on the situation. For example, during free-play, a sensitive mother who was not detached or intrusive would follow the child's lead, and allow the child to pick up and handle different toys among all those provided, at his/her own pace and according to what grabbed the child's attention. The sensitive mother might comment on the child's apparent enjoyment of the toys, and would typically engage in child directed play. In the teaching task, or with each object in the Three Boxes procedure, the sensitive, engaged, and non-intrusive mother would attempt to focus the child's attention to the presented toy or book, perhaps by comments or questions asked in relation to the specific object. If the child lost interest, the sensitive mother would take time to re-engage the child with the object at the child's own pace and in sync with the child's signals and needs.

In collaboration with the SECCYD investigator who implemented the rating of the SECCYD's videotapes, original rating item descriptions and scales from the 5-point scale adaptation (Cox, 1997; Owen et al., 2010) of the rating manual used in the SECCYD (NICHD ECCRN, 1999) were adapted to fit the BONDS' interaction tasks. For the BONDS, ratings of maternal parenting quality were made for the 4-minute free-play and for the 6-minute teaching task, separately. For the SECCYD interaction videos, a single set of ratings was made for the entire 15-minute Three Boxes procedure. Inter-rater reliability of the ratings for both studies was based on double-coding 20% of the BONDS videos and 18% of the SECCYD videos, assigned blindly. As determined by intra-class correlations (ICC), inter-rater reliability coefficients for BONDS ratings ranged from .68 to .80 ($M = .76$) in the free-play task and .72 to .80 ($M = .77$) in the teaching task, and for the SECCYD ICC's ranged from .69 to .87 ($M = .76$).

PARENTING QUALITY CONSTRUCT VALIDITY

BONDS predictors of maternal parenting quality. The following predictors were included from the BONDS study: *Child gender* (coded girl [0] and boy [1]) and *maternal age* (in years) were drawn from the 6-month interview. Parents reported on *child global development* at 1 year of age using the Norwegian version of the Ages and Stages Questionnaire (ASQ; Janson & Smith, 2003; Bricker & Squires, 1999). *Economic hardship* was reported by parents in the 1-year interview. A single item, coded dichotomously for presence or absence, asked whether the parent had experienced long-term difficulties related to family finances (paying rent, mortgages, etc.). We combined responses from mothers and fathers, so any confirming response resulted in a rating of 1 (income data was not collected in BONDS). Current *single parent* status was reported by mothers at the age 2 interview. *Maternal distress* reflected symptoms of depression and anxiety reported when the child was 6 months old based on the mean score from the 13-item version of the Hopkins Symptom Check List (Strand, Dalgard, Tambs, & Rognerud, 2003). Cronbach's alpha in current sample was .89.

SECCYD predictors of maternal parenting quality. A comparable set of predictors of maternal parenting quality from the data sets of the SECCYD were used in our analyses, including *maternal age*, *maternal education*, *single parenthood*, *economic hardship*, *maternal depressive symptoms*, *child gender*, and *child global development*. The measure of single parenthood was dichotomized from a coding of the mother's report of her marital/partnered status in the home interview at 2 years. Economic hardship was based on an income-to-needs ratio, calculated on the basis of the mother's report of household income divided by the Federal Poverty Level for the household size published in the year the age 2 data were collected. The measure of maternal depression was the total number of symptoms reported using the Center for Epidemiological Studies Depression Scale (CES-D; Radloff, 1977) at 6 months; current sample's Cronbach's alpha was .90. Child global development

PARENTING QUALITY CONSTRUCT VALIDITY

was the standardized mental development score from individual child administration of the Bayley Scales of Infant Development – Revised (Bayley, 1991) at 15 months.

BONDS child outcomes. Child outcomes were selected from measures collected in the lab visit when the child was 4 years old.

Receptive language. The child's receptive language was measured with the Norwegian version of the British Picture Vocabulary Scale II (BPVS-II; Dunn, Whetton, and Burley, 1997), which consists of 12 of the 14 original sets of pictures (Lyster, Horn, and Rygvold, 2010). In each of the 12 sets of increasing levels of difficulty, children were instructed to choose one correct picture out of four in response to a word/concept produced by the experimenter. The overall score is the sum of correct responses (verbal or nonverbal). After testing was completed, minor deviations from the correct testing procedure were found for 104 children; most errors resulted in missing data for set 1 (i.e., children who should have, but were not, tested on set 1, due to ≥ 2 incorrect answers on set 2) that were estimated for these children using their set 2 scores. Cronbach's alpha for the BPVS-II in the current sample was .81.

Externalizing behaviors (teacher and parent report). Externalizing behaviors were measured from child-care teachers and parents using the Teacher Report Form of the Child Behavior Checklist for Ages 1.5-5 (Achenbach & Rescorla, 2000) for child-care teacher reports and the Child Behavior Checklist (Achenbach & Rescorla, 2000) for parent reports. As there are no Norwegian norms, we used the raw mean for the broad band externalizing syndrome score for our analyses. Cronbach's alpha for the full externalizing score in the current sample was .94 for the teacher rating, and .87 for the parent rating.

Effortful control. A set of 4 tasks, building on work by Kochanska and colleagues (Kochanska & Knaack, 2003) were used to measure child effortful control. Five trained assessors scored children's performance using strict objective scoring criteria. Tasks

PARENTING QUALITY CONSTRUCT VALIDITY

included: 1) *Shapes* (effortful attention); the child was presented with 12 cards with small shapes of familiar objects (e.g., banana) contained within larger shapes of objects, and asked to name the smaller shape, consistent or inconsistent with the larger shape, 2) *Circle* (control of motor speed); the child was asked to draw along the line of a pre-drawn circle, first for training, next as slow as possible, and last, as fast as possible, 3) *Whisper* (control of voice); the child was presented with 12 cards with easily recognizable objects (e.g., ice cream) and asked to whisper, not shout or speak in a normal volume, the names of the shapes, and 4) *Gift delay* (inhibitory control); the child was asked to stand facing the wall without speaking or turning around while the assessor pretended to wrap a surprise gift. A sum score was calculated from the total number of correct responses across tasks.

SECCYD child outcomes. Child outcomes were selected from measures collected in lab visits when the child was 4.5 years old.

Receptive language. The child's receptive language was measured using the auditory comprehension scale of the Preschool Language Scale-3 (Zimmerman, Steiner, & Pond, 1979). This instrument is a standardized measure widely used in research and for clinical assessments; it was normed with children 2 weeks to 6 years 11 months of age. Excellent concurrent reliability and validity has been documented. The standardized scoring was used. PLS-3 standard scores have a mean of 100, standard deviation of 15.

Externalizing behaviors (teacher and parent report). Externalizing behaviors were measured using the same instruments as in BONDS: The Teacher Report Form (Achenbach & Rescorla, 2000), collected from the child's primary child-care provider, and the Child Behavior Checklist (Achenbach & Rescorla, 2000), collected from the mothers. Raw scores were used in analyses of the SECCYD child outcomes data. Alphas were not reported for these normed instruments in the SECCYD; extensive psychometric information is available in ECCRN manuals and publications.

PARENTING QUALITY CONSTRUCT VALIDITY

Effortful control. The measure of the child's effortful control at this age in the SECCYD was an average of the effortful control dimension items from the Children's Behavior Questionnaire (CBQ; Rothbart, Ahadi, & Hershey, 1994) collected from the mothers. In the current sample, the effortful control dimension included items from the two scales Attentional Focusing (9 items, $\alpha = .74$) and Inhibitory Control (10 items, $\alpha = .75$). Attentional Focusing captures the child's ability to concentrate and focus on an activity; item examples include: "Has a hard time concentrating on an activity when there are distracting noises" and "When building or putting something together, becomes very involved in what she/he is doing, and works for long periods". Inhibitory Control measures the child's ability to respond readily to parent or teacher requests; example items: "Has difficulty waiting in line for something" and "Can easily stop an activity when told 'no'".

Analysis Methods

We took a latent variable approach using confirmatory factor analyses (CFA) to determine the best fitting measurement model, using Mplus, version 7.11 (Muthén & Muthén, 1998-2013) for all inferential analyses. Given that the observational ratings were performed using rating scales with anchor points, we used a Weighted Least Squares, Mean and Variance adjusted (WLSMV) estimator. This estimator does not rely on assumptions about normality. We used the default option of covariance matrix input in Mplus. For all analyses, we evaluated model fit according to conventional standards for fit indexes, with Root Mean Square Error of Approximation (RMSEA) below .06, Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) above .95 indicating very good model fit, and values below .08 and above .9, respectively, indicating reasonable fit (Hu & Bentler, 1999). For our models including covariates, we handled missing data by the standard estimation procedure in Mplus, which means including all available data to estimate the model using full information maximum likelihood.

PARENTING QUALITY CONSTRUCT VALIDITY

Results

Descriptive statistics are presented in Table 1. These include raw scale scores for observed parenting ratings, demographic, child and maternal predictors as well as child outcomes. Note that mothers' detachment score was dichotomized in the SECCYD, due to the skewed distribution at this age. Table 1 includes rates of missing data, which was about 22% for the mother-child observations, about 42% for teacher rated externalizing behaviors, 16% for mother-rated externalizing behaviors, and slightly above 20% for receptive language and effortful control. Multivariate tests of missingness predicted from our key background variables (included in Table 1) indicate that in the BONDS data, there were some differences in missing data across municipalities and birth cohorts. In addition, missingness on mother-reported externalizing behaviors was higher among mothers with lower education, while missingness on effortful control was higher among girls. Apart from this, missingness was unrelated to our covariates. In the NICHD SECCYD, missingness in observed parenting was predicted by maternal age with older mothers less likely to be missing. There were no significant predictors of missing data for language comprehension, effortful control, or maternal report of externalizing problems. Teacher reported externalizing problems were more likely missing for mothers with low education and income, and with higher levels of maternal depression. Note that under the assumption of missing data at random (MAR), bias is minimized in our prediction analyses by conditioning on these covariates.

Model Testing in BONDS Data

Across tasks. In order to identify a best fitting measurement model from the six parenting items rated for mother-child interactions in free-play (task 1) and teaching (task 2), we started with the simplest possible model. In a one factor model including all twelve parenting items; six from each task (Model 1), we tested whether one overarching *parenting quality* factor accounted for variability in all rated items collected across the two tasks. As the

PARENTING QUALITY CONSTRUCT VALIDITY

same set of ratings (e.g., sensitivity, intrusiveness) was applied to both tasks, we allowed the residuals for equivalent items across tasks to correlate. As can be seen in Table 2, this model did not fit well, although factor loadings were high and statistically significant (ranging from $-.399$ to $.990$, for details, see Table S1 in the online appendix). Notably, in this model, there was (negative) collinearity between the errors of the sensitivity and intrusiveness items in both tasks. From the literature review, it seems that the intrusiveness item has been similarly difficult to include in an overall parenting measure in some reports. In several studies intrusiveness was left out of the composite or latent measure(s) (Caughy & Owen, 2015; Mills-Koonce et al., 2015; Towe-Goodman et al., 2014), and in others the raw intrusiveness scores were used as a variable by itself (Berlin et al., 2002; Peredo et al., 2015). Based on our poor model fit and the collinearity problems, and on previous studies, we removed the intrusiveness item from the fitting of a model across both tasks, and refitted the model, including the remaining ten items (Model 2). This solution improved the model fit slightly (see Table 2), although the fit was still poor (see Table S1 for factor loadings). Thus, we concluded that including ratings from the two differently structured tasks in the same model did not fit the data, and we turned to analysis of the parenting ratings separately for each task.

Within task. In the next models, we tested a single factor *parenting quality* model for each task (task 1 separate from task 2). In the first model (Model 3), all the six parenting items from task 1 and the six items for task 2 were included, and the two latent variables from items rated in each task were allowed to correlate. Some model fit indices improved, CFI = $.93$ and TLI = $.90$, but the Index RMSEA was unsatisfactory ($.12$) (see Table S1 for factor loadings). Subsequently, we tested the same solution but removed intrusiveness, as discussed above, from the sets of items rated in each of the two tasks (Model 4), analyzing the fit of two single *parenting quality* factors without the intrusiveness item. We obtained good model fit (CFI = $.99$, TLI = $.98$; RMSEA = $.06$) (see Table 2) and factor loadings were high and

PARENTING QUALITY CONSTRUCT VALIDITY

statistically significant (.38 to .89, see Table S1). This model indicates that both for the free play task and the teaching task there is a well-fitting one-dimensional latent *parenting quality* factor including 5 of the 6 rated dimensions, when rating items are modeled separately for the two tasks.

Since our one factor model did not fit well when including items from both tasks in a single factor, while Model 4 described above fitted the data well, we next compared constrained and unconstrained factor loadings for similar items across tasks to check whether we measure the same construct with the two tasks (i.e., measurement invariance). For example, sensitivity in task 1 was set to be equal to sensitivity in task 2 and so on. The constrained model was significantly worse than the unconstrained model (DIFF test = 7.97, $p < .05$) so we maintained all parameters free. This means that *Parenting Quality* should be interpreted as being slightly different constructs when measured with the free-play task versus the teaching task.

Replication of Model Fitting in SECCYD Data

We then replicated our model fitting using data from the SECCYD. The SECCYD includes only one set of ratings for the entire Three Boxes procedure, and our replication was therefore restricted to fitting the main measurement model, not task-specificity of the model.

Initially, we tested a single factor *parenting quality* model including all 6 parenting dimensions. As can be seen in Table 2, this first model (Model 1) had good fit indices, CFI = .95 and TLI = .92, but the Index RMSEA was unsatisfactory (.17). Next, we tried the same solution but with intrusiveness removed (Model 2), as discussed above. We obtained adequate, though not ideal model fit (CFI = .99, TLI = .98; RMSEA = .09) and factor loadings were high and statistically significant (.70 to .92, see Table S2 in appendix). Hence, for both BONDS and SECCYD data the best fitting measurement model for *parenting quality* involved a one-factor solution including sensitivity/responsiveness, detachment (reversed),

PARENTING QUALITY CONSTRUCT VALIDITY

cognitive stimulation, positive regard, and negative regard (reversed) for the child, and excluding intrusiveness.

BONDS Measurement Model with Predictors and Outcomes

Returning to the BONDS data, we expanded the validity check of our latent parenting measure by using our best fitting measurement model (Model 4) from the prior analyses to test whether selected maternal and child characteristics predicted the level of *parenting quality* and whether parenting quality predicted a number of child outcomes at age 4. Specifically, we fitted structural equation models for three measurement models with structural parts including observed scores for predictors and outcomes; one including the free play-task (task 1) and the teaching task (task 2) in the same model, and a second comprised tests of models for task 1 and task 2 separately. Because the patterns of associations are slightly different when the two tasks are modeled in combined versus separate models, we provide results for both modeling options. Results for the tasks modeled separately also enhances comparison with the NICHD SECCYD sample (see below). The analyses of outcomes were controlled for the predictors of parenting in the models.

As can be seen in the upper part of Table 3, child gender appears as a consistent predictor of *parenting quality* in both tasks and across models, with higher parenting quality exhibited in interactions of mothers with girls compared to mothers with boys. Moreover, better overall development at age 1, was a good predictor of higher *parenting quality* during the teaching task, but not so for the free-play task. Less educated mothers and those without a partner displayed lower general *parenting quality* levels than more educated mothers and those living with a partner. Older mothers had higher parenting quality ratings in the teaching task. Economic hardship and maternal distress had no consistent significant predictive value for *parenting quality* across tasks or models.

PARENTING QUALITY CONSTRUCT VALIDITY

In order to test the predictive validity of our measurement model, we included in the models child outcomes measured at age 4 hypothesized to be associated with *parenting quality*: effortful control, language development, and externalizing behaviors (both parent- and teacher-rated). As can be seen in the lower part of Table 3, *parenting quality* showed predictive validity for all outcomes, particularly when measured in the teaching task. More specifically, when the two tasks were included in the same model, the predictive value of *parenting quality* in the free-play task was mostly non-significant, except for language skills, for which the coefficient was negative and significant and thus in an unexpected direction. In other words, *parenting quality* during the free-play task was generally not significantly predictive of the outcome variables over and above *parenting quality* during the teaching task. Conversely, *parenting quality* during the teaching task remained a significant predictor of all the outcome variables over and above the effect of *parenting quality* during the free-play task. See fit indexes for full models with predictors and outcomes in Table 4.

We then tested a similar model separately for each task. As expected, better *parenting quality* was predictive of higher levels of effortful control and receptive language, and of lower levels of externalizing behavior as reported by parents and teachers. Moreover, these associations were significant for both the free-play task and the teaching task with only one exception; externalizing behavior rated by parents was significantly predicted by *parenting quality* in the teaching task, but not in the free-play task. Notably, the coefficients for effortful control and language were about twice the size in the model with parenting ratings from the teaching task only, compared to results obtained in the model using the free-play task ratings, while the coefficients for externalizing problems were more similar across the models tested.

Replication of Prediction Model in SECCYD Data

PARENTING QUALITY CONSTRUCT VALIDITY

We used our best fitting measurement model from the NICHD SECCYD (Model 2) to test a full model with maternal and child predictors and child outcomes chosen for their equivalence to those in the BONDS data. As can be seen in Table 5, and closely similar to the results of the previous analyses with BONDS, all associations were significant and in the expected direction. Note that maternal distress and economic hardship were significant predictors of lower *parenting quality* in the U.S. sample, whereas they did not quite reach significance as predictors in the Norwegian sample. In addition, we had to remove single parenthood from the analyses in the U.S. sample as our model failed to converge with that predictor included. Despite these minor distinctions, taken together, our results lend support to the robustness of our findings across different tasks and different socio-cultural contexts, and adds to the generalizability of our latent measure of *parenting quality*.

Discussion

The purpose of the present study was to test the construct validity of a proposed latent measure of parenting quality from widely used observational ratings of mothers' interactions with their 2-year-olds in two large community samples in Norway and the U.S. Careful model testing suggested that a one-factor parenting quality latent construct, including mothers' sensitivity, detachment, cognitive stimulation, positive regard, and negative regard, but excluding intrusiveness, was the best fitting model across samples. The two samples differed in procedures used to observe interactions, and in rating scales, which allowed for testing of the structural and content validity aspects of our global parenting quality measure and finding similarities in results across the samples. Furthermore, in the Norwegian BONDS data, greater predictive validity from the observational ratings was obtained from ratings of interactions in the more structured observational procedure (the teaching task). Adding support to the generalizability and the predictive validity of our measure, associations

PARENTING QUALITY CONSTRUCT VALIDITY

between age 2 parenting quality in the SECCYD Three Boxes procedure and age 4.5 child outcomes resembled the results from the BONDS teaching task.

Following the most comprehensive statistical approach for testing the *structure* in the observed parenting items (John & Benet-Martínez, 2000), and building our approach on previous work in the field (Mills-Koonce et al., 2015; Whiteside-Mansell et al., 2003), our hypothesized parenting quality measure including all six of the rated items did not fit our data well. Our CFAs showed that 5 of the 6 ratings items used in the SECCYD coding procedures, and adapted to our Norwegian BONDS data, reflected variability in one underlying parenting quality dimension. Specifically, these parenting quality rating items were sensitivity, detachment (reversed), cognitive stimulation, positive regard, and negative regard (reversed). Following the logic of the CFA (e.g., Brown, 2015), the variability in these ratings is causally determined by the single unobserved construct of parenting quality. The sixth rating item, intrusiveness, did not fit the model, and following the same logic does not therefore reflect the same parenting construct. Had we followed a conventional approach to summarizing the ratings into an overall score through summing the item scores, as has previously been done in the SECCYD and elsewhere (e.g., Bocknek et al., 2009; NICHD ECCRN, 2003; Towe-Goodman et al., 2014), this would have gone undetected, with the consequence of this providing a potentially biased measure. In addition to testing the *structural* and *content* validity aspects of our latent measure, the analyses show that while some rating items may go together from a theoretical point of view, this need to be tested empirically.

That the current results differ from Whiteside-Mansell et al. (2003), who found that a model of three parenting factors fit their data best, may reflect the diversity in derivation of measures (combining rating items from a Three Boxes procedure and a clean-up task vs. separate ratings from a free-play and a teaching task) and child age (3 vs. 2 years). The measurement model of sensitive parenting in Mills-Koonce et al. (2015) included only a

PARENTING QUALITY CONSTRUCT VALIDITY

subset of the rated parenting items, leaving out sensitivity, intrusiveness, and positive regard in their CFA, and including animation, which was not among the current study's rating items). The lack of an explicit justification for their choice of items makes a comparison of results difficult.

Adding to the *construct* validity of our measure, in particular to the *generalizability* aspect, we were able to fit this one-factor measurement model across two tasks in the Norwegian data, and in the SECCYD data. In technical terms, this is called configural invariance; the same number of factors (one) explains variability in the same number of comparable ratings. Based on this, it seems that regardless of type of task (free play or teaching task or Three Boxes procedure), country (Norway or the U.S.), duration of observed sequence (4 and 6 minutes in Norway, 15 in the U.S.) and rating scale (5-point in Norway, 4-point in the U.S.), a parenting quality construct is reflected in these 5 ratings. Of course, this conclusion is preliminary, as it is true only across the conditions included in this study. Of note, however, and as noted further below, is the similarity of these findings with results of a single latent parenting factor identified through CFA in a sample of low-income, Spanish-speaking Mexican-American mothers observed in a Three Boxes procedure with their 2½ year old children (Peredo et al., 2015). Somewhat different results emerged, however, when single-factor and two-factor models were tested using data from both the African American and the Hispanic mother-child dyads in the larger sample from which the Spanish-speaking dyads were drawn (see Caughy, Mills, Owen, Dyer, & Oshri, 2017).

The seemingly robust latent measure of maternal parenting quality across tasks and samples, in which mothers' sensitive, responsive, stimulating, and positive engagement appeared to draw in the same direction, whereas the ratings of mothers' intrusiveness did not fit in, raises interesting questions regarding this particular aspect of parenting, and concerns the *content* aspect of validity. Former studies have also found intrusiveness to protrude from

PARENTING QUALITY CONSTRUCT VALIDITY

the other rated items. In line with the current study, Peredo and colleagues (2015) found that a CFA of the same rating items at child age 2.5 years indicated that a single parenting factor including sensitivity, cognitive stimulation, positive regard, negative regard, and detachment, and excluding intrusiveness, fit their data best. These results with a Hispanic Spanish-speaking sample in the U.S. lend further support to the generalizability of the parenting quality measure in question. Furthermore, three of the reviewed studies omitted the intrusiveness rating item and used a composite of other items in their analytic measures of parenting quality (Caughy & Owen, 2015; Mills-Koonce et al., 2015; Towe-Goodman et al., 2014). The rationale for omitting intrusiveness from the composite measure of parenting or from the analyses all together is seldom made explicit.

According to descriptions in the rating manual (Owen et al., 2010), higher levels of parental sensitivity, cognitive stimulation and positive regard are all associated with higher degrees of positive engagement, while detachment and negative regard are indicated by a lack of positive involvement. These parenting items are easily placed along a positive-to-negative dimension, and our parenting quality measure may thus reflect the level of warm, sensitive, supportive and engaged parenting. Intrusiveness may be more complex and represent a different feature. Whereas sensitive and responsive parenting behaviors are almost exclusively associated with positive child outcomes, relations between parental intrusiveness and child outcomes have no such clear patterns and the findings are contradictory (Ispa et al., 2004; McFadden & Tamis-LeMonda, 2013). Even if intrusive interactions are clearly adult-centered and involve imposing the parent's agenda on the child, the affective context in which intrusiveness occurs may be a critical determinant of its impact, and parental warmth may influence the meaning children attribute to the firm control (Dyer et al., 2014; Ispa et al., 2004; McFadden & Tamis-LeMonda, 2013).

PARENTING QUALITY CONSTRUCT VALIDITY

Furthermore, even if the factor structure in our one-factor model proved identical in the two differently structured interaction tasks in the BONDS sample, the factor loadings were not the same across the two tasks, meaning we found evidence for configural invariance across tasks, but not for metric invariance. The substantive interpretation of this lack of evidence for metric invariance is that the ratings do not have the same meaning when applied to the two different tasks. Rather, it seems like the two tasks elicit slightly different parenting behaviors. This finding is not surprising, given that the free-play and the teaching tasks were set up to elicit a variety of behaviors associated with children's social development and adjustment based on empirical evidence (e.g., Aspland & Gardner, 2003; Snyder et al., 2003). A free-play task is assumed to be suitable as a warm-up in front of the camera, given that there is no implied pressure to reach a specific goal, and the parents and children generally find the task pleasant. Contrary, and although the parent introductory is told that there is no demand to complete the assignment (i.e., get all puzzle pieces in place or all the blocks placed correctly), the teaching task used in the present study involves a mild, implied, pressure to achieve. With intent, both the puzzle and the sorting blocks that were provided are somewhat difficult to complete for most 2-year-olds, and consequently, the task may elicit variation in parenting behaviors as our results suggest.

Further evidence for this difference in parenting behaviors between tasks is provided by our analyses of predictive associations with child outcomes. When we allowed the two latent models from the free-play and teaching tasks to compete in predicting effortful control, receptive language, and externalizing behavior at age 4, only the teaching task parenting quality was consistently associated with the outcomes. When we tested the associations between the latent models and the outcomes separately, the coefficients from the teaching task were about twice as large as those from the free play task. The coefficients for the teaching task predicting child outcomes were essentially similar regardless of whether free-

PARENTING QUALITY CONSTRUCT VALIDITY

play was controlled for or not. Thus, more structured and goal-oriented tasks seem to elicit parenting behaviors more predictive of child outcomes. Perhaps with more structure, mothers may feel a greater demand to succeed, which again may elicit more active engagement from mothers than in a free-play task. Furthermore, more active maternal engagement may be expressed in positive and supportive ways, but also in negative and critical ways, and this, as shown by Panfile and colleagues (2012) may lead to more mother-initiated parent-child conflicts. Negativity and conflict versus positive involvement may be an important distinction in parenting quality between the structured and the unstructured tasks, as supported by earlier studies (e.g. Ginsburg et al., 2006; Kwon et al., 2012; Panfile et al., 2012). The current study expands our understanding of these differences in behavior by showing that maternal parenting quality in a more structured procedure, even when the structure is “semi” in nature, is more predictive of child outcome in multiple domains two years later. Taken together, this suggests that the teaching task elicits more of the variability in parenting quality which is essential for predicting future child development than does the free-play task. Moreover, the free-play task adds little in terms of predictive value over the predictive value of the teaching task by itself. If, on the basis of these findings we were to recommend a task for assessing the quality of parent-child interaction, we would recommend a more structured procedure, rather than purely free play between parent and child.

The current study has notable strengths, through the inclusion of two unusually large samples in this research area, a cross-national comparison, and a statistical approach allowing for careful testing of competing measurement models. These factors all contribute to strengthen the construct validation of our parenting quality measure. There are, however, limitations to be considered for improvements in future studies. While we draw preliminary conclusions about the advantage of a teaching task over a free-play task, our data do not include a manipulation of the order of these tasks. Thus, we cannot rule out that we in fact

PARENTING QUALITY CONSTRUCT VALIDITY

observe a sequencing effect; that is, that interactions are different and more informative of individual differences when parents and children become familiar with the tasks. Future research may consider manipulating the order of structured and unstructured tasks to investigate this matter further. Moreover, in our cross-national comparison, we rely on somewhat different measures both on the predictor side (among our measures of demographics, child- and maternal characteristics), and on the child outcome side, even though we have included measures of the same broad constructs (for details, see method section). Despite this, and supporting the predictive validity of our parenting quality construct, individual differences in parenting quality were related in similar fashions to predictors and to outcomes that were indicators of similar constructs. That economic hardship did not significantly predict parenting quality and maternal distress did not quite reach significance in the BONDS sample may reflect cultural differences between the two countries, since fewer mothers in Norway experience economic hardship and mental distress than in the U.S., possibly due to the greater prevalence of financial and family leave supports in Norway. Moreover, the measures in the two studies were not exactly the same, which may have affected the results. In addition, results of analyses of missingness among the predictors and child outcome variables in the two samples, reveal some limitations to the generality of the findings derived from these two large samples of families from Norway and the United States. Finally, this study addresses mothers' parenting only; thus, future studies may add to the construct validity by testing the parenting quality latent measure derived in samples of fathers' interaction with their toddlers.

Transparency of methods is a central tenet of science, which helps to clarify the nature of the relations between parenting and child outcomes. The current study adds to the literature by explicitly describing the procedures involved in deriving an overarching measure of parenting quality from observed mother-child interactions at age 2. The set up and

PARENTING QUALITY CONSTRUCT VALIDITY

structure of the interaction tasks, the parenting items rated, and the statistical approach applied to the data to establish a valid and robust measurement model have been displayed in detail to make possible replications across other samples. The present study contributes to ongoing efforts in the field to establish reliable and valid measures of constructs also through international comparison of measurement models. In conclusion, it is salient to note that single studies do not evidence construct validity of a measure. Construct validation is a continuous ongoing process which is expanded by each new study (Messick, 1995). Measures derived from observational ratings of parent-child interaction should be justified and described in detail, psychometric evaluations made explicit, and measure derivation made transparent. Thus, future research should not only strive to replicate or modify our measurement model, but also to expand the evidence for the validity of the parent quality construct to include theoretically meaningful relations to other related constructs not included in the current study.

References

- Achenbach, T. M., & Rescorla, L. (2000). *Manual for the ASEBA preschool forms & profiles: child behavior checklist for ages 1 1/2-5, language development survey, caregiver - teacher report form : an integrated system of multi-informant assessment.*
- Aspland, H., & Gardner, F. (2003). Observational measures of parent-child interaction: An introductory review. *Child and Adolescent Mental Health, 8*, 136-143. doi: 10.1111/1475-3588.00061
- Bakermans-Kranenburg, M. J., van Ijzendoorn, M. H., & Kroonenberg, P. H. (2004). Differences in attachment security between African-American and white children: ethnicity or socio-economic status? *Infant Behavior & Development, 27*, 417-433. doi:10.1016/j.infbeh.2004.02.002
- Barnett, M. A., Shanahan, L., Deng, M., Haskett, M. E., & Cox, M. J. (2010). Independent and interactive contributions of parenting behaviors and beliefs in the prediction of early childhood behavior problems. *Parenting: Science and Practice, 10*, 43-59. doi: 10.1080/15295190903014604
- Bayley, N. (1991). *Bayley Scales of Infant Development: Second Edition (Standardization Version)*. New York: The Psychological Corporation.
- Belsky, J. (1984). The determinants of parenting: A process model. *Child Development, 55*, 83-96.
- Belsky, J., & Jaffee, S. R. (2006). *The multiple determinants of parenting*. Developmental psychopathology: Vol. 3. Risk, disorder, and adaptation (pp. 38–85). Hoboken.
- Berlin, L. J., Brady-Smith, C., & Brooks-Gunn, J. (2002). Links between childbearing age and observed maternal behaviors with 14-month-olds in the Early Head Start Research and Evaluation Project. *Infant Mental Health Journal, 23*, 104-129. doi: 10.1002/imhj.10007

PARENTING QUALITY CONSTRUCT VALIDITY

- Biringen, Z., Robinson, J. L., & Emde, R. N. (1998). Emotional availability scales (EA). *Unpublished coding manual. Colorado, USA: Department of Human Development and Family Studies, Colorado State University.*
- Bocknek, E. L., Brophy-Herb, H. E., & Banerjee, M. (2009). Effects of parental supportiveness on toddlers' emotion regulation over the first three years of life in a low-income African American sample. *Infant Mental Health Journal, 30*, 452-476. doi: 10.1002/imhj.20224
- Bricker, D., & Squires, J. (1999). *Ages and Stages Questionnaire: A parent-completed child monitoring system* (2nd ed.). Baltimore: Paul Brookes.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. New York: Guilford Press.
- Burchinal, M., Vernon-Feagans, L., Cox, M., & Key Family Life Project Investigators. (2008). Cumulative social risk, parenting, and infant development in rural low-income communities. *Parenting: Science and Practice, 8*, 41-69. doi: 10.1080/15295190701830672
- Caughy, M. O., Mills, B., Owen, M. T., Dyer, N., & Oshri, A. (2017). Ethnic differences in mothering qualities and relations to academic achievement. *Journal of Family Psychology, 31*, 855-866. doi: 10.1037/fam0000334
- Caughy, M. O. B., & Owen, M. T. (2015). Cultural socialization and school readiness of African American and Latino preschoolers. *Cultural Diversity and Ethnic Minority Psychology, 21*, 391-399. doi: 10.1037/a0037928
- Clark, R. (1999). The parent-child early relational assessment: A factorial validity study. *Educational and psychological measurement, 59*(5), 821-846.
- Cox, M. J. (1997). *Qualitative ratings for parent-child interaction at 24-36 months of age*. Unpublished manuscript, University of North Carolina, Chapel Hill.

PARENTING QUALITY CONSTRUCT VALIDITY

- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed.). Washington, D. C.: American Council on Education.
- Cronbach, L.J., & Meehl, P.E. (1955). Construct validity for psychological tests. *Psychological Bulletin*, 52, 281-302.
- Deater-Deckard, K., Pylas, M., & Petrill, S. A. (1997). The Parent-child interaction system (PARCHISY). London: *Institute of Psychiatry*.
- Dunn, L.M., Whetton, C.W., & Burley, J. (1997). *The British Picture Vocabulary Scales, 2nd edition*. Windsor, UK: NFER Nelson.
- Dyer, N., Owen, M. T., & Caughy, M. O. (2014). Ethnic differences in profiles of mother-child interactions and relations to emerging school readiness in African American and Latin American children. *Parenting: Science and Practice*, 14, 175-194. doi: 10.1080/15295192.2014.972756
- Garai, E. P., forehand, R. L., & Colletti, C. J. M. (2009). The relation of maternal sensitivity to children's internalizing and externalizing problems with the context of maternal depressive symptoms. *Behavior Modification*, 33, 550-582.
- Ginsburg, G. S., Grover, R. L., Cord, J. J., Ialongo, N. (2006). Observational measures of parenting in anxious and nonanxious mothers: Does type of task matter? *Journal of Clinical Child and Adolescent Psychology*, 35, 323-328. doi: 10.1207/s15374424jccp3502_16
- Gonzalez, M. M. (1996). Tasks and activities. A parent-child interaction analysis. *Learning and Instruction*, 6, 287-306. Doi: 10.1016/S0959-4752(96)00017-5
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6, 1-55. doi: 10.1080/10705519909540118

PARENTING QUALITY CONSTRUCT VALIDITY

- Ispa, J. M., Fine, M. A., Halgunseth, L. C., Harper, S., Robinson, J.A., Boyce, L., Brooks-Gunn, J., Brady-Smith, C. (2004). Maternal intrusiveness, maternal warmth, and mother–toddler relationship outcomes: Variations across low-income ethnic and acculturation groups. *Child Development, 75*, 6, 1613 – 1631. doi: 10.1111/j.1467-8624.2004.00806.x
- Janson, H. & Smith, L. (2003). *Norsk Manual Supplement til Ages and Stages Questionnaires* (Norwegian Manual Supplement for the Ages and Stages Questionnaire). Oslo, Norway: Regional Center for Child and Adolescent Mental Health, Health Region East/South.
- John, O. P., & Benet-Martinez, V. (2000). Measurement: Reliability, construct validation, and scale construction. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 339-369). Cambridge: Cambridge University Press.
- Klein, M. R., Lengua, L. J., Thompson, S. F., Moran, L., Ruberry, E. J., Kiff, C., & Zalewski, M. (2018). Bidirectional relations between temperament and parenting predicting preschool-age children’s adjustment, *Journal of Clinical Child & Adolescent Psychology, 47*: sup1, S113-S126. doi: 10.1080/15374416.2016.1169537
- Kochanska, G., & Knaack, A. (2003). Effortful control as a personality characteristic of young children: Antecedents, correlates, and consequences. *Journal of Personality, 76*, 1087-1112. doi: 10.1111/1467-6494.7106008
- Kwon, K-A., Bingham, G., Lewsader, J., Jeon, H-J, Elicker, J. (2013). Structured task versus free play: The influence of social context on parenting quality, toddlers’ engagement with parents and play behaviors, and parent–toddler language use. *Child Youth Care Forum, 42*, 207–224. :207–224. doi: 10.1007/s10566-013-9198-x

PARENTING QUALITY CONSTRUCT VALIDITY

- Lansford, J. E., & Bornstein, M. H. (2011). Parenting attributions and attitudes in diverse cultural contexts: Introduction to the special issue. *Parenting, 11*, 87-101.
- Lansford, J. E., Bornstein, M. H., Deater-Deckard, K., Dodge, K. A., Al-Hassan, S. M., Bacchini, D., ... Zelli, A. (2016). How international research on parenting advances understanding of child development. *Child Development Perspectives, 10*, 202-207. doi: 10.1111/cdep.12186
- Lyster, S.-A. H., Horn, E., & Rygvold, A-L. (2010). Ordforråd og ordforrådsutvikling hos norske barn og unge. Resultater fra en utprøving av British Picture Vocabulary Scale, Second Edition (BPVS II). *Spesialpedagogikk, 09*, 35–43.
- Maas, A. J., Vreeswijk, C. M., van Bakel H. J. (2013). Effect of situation on mother–infant interaction. *Infant Behavior & Development 36*, 42– 49. doi: 10.1016/j.infbeh.2012.10.006
- Martin, A., Ryan, R. M., & Brooks-Gunn, J. (2007). The joint influence of mother and father parenting on child cognitive outcomes at age 5. *Early Childhood Research Quarterly, 22*, 423–439. doi:10.1016/j.ecresq.2007.07.001
- McFadden, K. E., Tamis-LeMonda, C. S. (2013). Maternal responsiveness, intrusiveness, and negativity during play with infants: Contextual associations and infant cognitive status in a low-income sample. *Infant Mental Health Journal, 34*, 80-92. doi: 10.1002/imhj.21376
- Mesman, J., & Emmen, R. A. (2013). Mary Ainsworth's legacy: A systematic review of observational instruments measuring parental sensitivity. *Attachment & Human Development, 15*, 485-506. doi: 10.1080/14616734.2013.820900
- Messick, S. (1995). Validity of psychological assessment. Validation of inferences from persons' responses and performances. A scientific inquiry into score meaning. *American Psychologist, 50*, 741-749. doi.org/10.1037/0003-066X.50.9.741

PARENTING QUALITY CONSTRUCT VALIDITY

Mills-Koonce, W. R., Willoughby, M. T., Zvara, B., Barnett, M., Gustafsson, H., Cox, M. J.

(2015). Mothers' and fathers' sensitivity and children's cognitive development in low-income, rural families. *Journal of Applied Developmental Psychology, 38*, 1–10. doi: 10.1016/j.appdev.2015.01.001

Muthén, L. K., & Muthén, B. O. (1998-2013). *Mplus (version) 7.11* [Computer Software].

Los Angeles, CA: Muthén & Muthén.

NICHD Early Child Care Research Network. (1999). Child care and mother-child interaction in the first three years of life. *Developmental Psychology, 35*, 1399-1413. doi:

10.1037/0012-1649.35.6.1399

NICHD Early Child Care Research Network. (2001). Parenting and family influences when children are in child care: Results from the NICHD Study of Early Child Care. In J.G.

Borkowski, S. Ramey, & Bristol-Power, M. (Eds.), *Parenting and the child's world: Influences on intellectual, academic, and social-emotional development* (pp. 99-124).

Mahwah, NJ: Erlbaum.

NICHD Early Child Care Research Network. (2003). Early child care and mother-child interaction from 36 months through first grade. *Infant Behavior and Development, 26*,

345-370. doi: 10.1016/S0163-6383(03)00035-3

NICHD Early Child Care Research Network. (2004). Fathers' and mothers' parenting

behavior and beliefs as predictors of children's social adjustment in the transition to school. *Journal of Family Psychology, 18*, 628-638. doi: 10.1037/0893-3200.18.4.628

Nærde, A., Janson, H., & Ogden, T. (2014). *BONDS (The Behavior Outlook Norwegian*

Developmental Study): A Prospective Longitudinal Study of Early Development of Social Competence and Behavior Problems. (Report). Oslo: Norsk Senter for studier

av problematferd og innovativ praksis. ISBN: 978-82-93406-00-6.

PARENTING QUALITY CONSTRUCT VALIDITY

- Owen, M. T., Amos, M. L., Bondurant, L., Hasanizadeh, N., Hurst, J. R., & Villa, A. (2010). *Qualitative ratings for parent-child interactions: Ages 2–4 years*. The University of Texas at Dallas.
- Owen, M. T., Caughy, M. O., Hurst, J. M., Amos, M., Hasanizadeh, N., & Mata-Otero, A. (2013). Contributions of fathering and mothering to emerging self regulation in low-income ethnic minority preschoolers. *Early Child Development and Care, 183*, 464-482. PMID: 23940412
- Panfile, T. M., Laible, D. J., Eye, J. L. (2012). Conflict frequency within mother–child dyads across contexts: Links with attachment security and gender. *Early Childhood Research Quarterly, 27*, 147– 155. doi:10.1016/j.ecresq.2011.06.001
- Peredo, T. N., Owen, M. T., Rojas, R., & Caughy, M. O. (2015). Child vocabulary, maternal behavior, and inhibitory control development among Spanish-speaking children. *Early Education and Development, 26*, 749-769. doi: 10.1080/10409289.2015.1009319
- Propper, C., Willoughby, M., Halpern, C. T., Carbone, M. A., & Cox, M. (2007). Parenting quality, DRD4, and the prediction of externalizing and internalizing behaviors in early childhood. *Developmental Psychobiology, 49*, 619-632. doi: 10.1002/dev.20249
- Pungello, E. P., Iruka, I. U., Dotterer, A. M., Mills-Koonce, R., & Reznick, J. S. (2009). The effects of socioeconomic status, race, and parenting on language development in early childhood. *Developmental Psychology, 45*, 544-557. doi: 10.1037/a0013917
- Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385-401.
- Rothbart, M. K., Ahadi, S. A., & Hershey, K. L. (1994). Temperament and social behavior in childhood. *Merrill-Palmer Quarterly, 40*, 21-39.

PARENTING QUALITY CONSTRUCT VALIDITY

- Snyder, J., Stoolmiller, M., Wilson, M., & Yamamoto, M. (2003). Child anger regulation, parental responses to children's anger displays, and early child antisocial behavior. *Social Development, 12*, 335-360. doi: 10.1111/1467-9507.00237
- Strand, B. H., Dalgard, O. S., Tambs, K., & Rognerud, K. (2003). Measuring the mental health status of the Norwegian population: A comparison of the instruments SCL-25, SCL-10, SCL-5 and MHI-5 (SF-36). *Nordic Journal of Psychiatry, 57*, 113-118. doi: 10.1080/08039480310000932
- Tamis-LeMonda, C. S., Shannon, J. D., Cabrera, N., & Lamb, M. (2004). Fathers and mothers at play with their 2- and 3-year-olds: Contributions to language and cognitive development. *Child Development, 75*, 1806-1820. doi: 10.1111/j.1467-8624.2004.00818.x
- Towe-Goodman, N. R., Willoughby, M., Blair, C., Gustafsson, H. C., Mills-Koonce, W. R., & Cox, M. J. (2014). Fathers' sensitive parenting and the development of early executive functioning. *Journal of Family Psychology, 28*, 867-876. doi: 10.1037/a0038128
- Vandell, D. L., (1979). Effects of a playgroup experience on mother-son and father-son interaction. *Developmental Psychology, 15*, 379-385. doi: 10.1037/0012-1649.15.4.379
- Vallotton, C. D., Mastergeorge, A., Foster, T., Decker, K. B., & Ayoub, C. Parenting supports for early vocabulary development: Specific effects of sensitivity and stimulation through infancy. *Infancy, 22*, 78-107.
- Whiteside-Mansell, L., Bradley, R. H., Owen, M. T., Randolph, S. M., Cauce, A. M. (2003). Parenting and children's behavior at 36 months: Equivalence between African-American and European-American mother-child dyads. *Parenting: Science and Practice, 3*, 197-234. doi: 10.1207/S15327922PAR0303_02

PARENTING QUALITY CONSTRUCT VALIDITY

- Wilson, S., & Durbin, C. E. (2012). The Laboratory Parenting Assessment Battery: Development and preliminary validation of an observational parenting rating system. *Psychological Assessment, 24*, 823. doi: 10.1037/a0028352
- Winslow, E. B., & Shaw, D. S. (1995). Early parenting coding system. *Unpublished manuscript, University of Pittsburgh.*
- Wright, N., Hill, J., Sharp, H., & Pickles, A. (2018) Maternal sensitivity to distress, attachment and the development of callous-unemotional traits in young children. *Journal of Child Psychology and Psychiatry, 59*, 790-800. DOI: 10.1111/jcpp.12867
- Zimmerman, I. L., Steiner, V.G. & Pond, R.E. (1979). *Pond Preschool Language Scale*. San Antonio, TX: The Psychological Corporation.

PARENTING QUALITY CONSTRUCT VALIDITY

Tables and Figures

Table 1

Descriptives of Sample and Study Variables for the BONDS (N=1,157) and for the NICHD SECCYD Sample (N=1,364)

Variable	BONDS			NICHD SECCYD		
	% Missing	% or <i>M</i> (range)	<i>SD</i>	% Missing	% or <i>M</i> (range)	<i>SD</i>
Child gender (boy)	0.00	51.80	-	0.00	51.69	-
Global development	4.93	213.51(45-300)	49.99	13.49	108.58(63-150)	14.07
Economic hardship	5.27	11.86	-	0.73	3.24(0-19.3) ICNR	2.79
Maternal age	0.61	30.78(19-42)	4.86	0.00	28.11(18-46)	5.63
Single parent	8.56	7.66	-	12.24	13.78	-
Maternal education	5.19	14.34(9-18)	2.56	0.00	14.60(7-21)	2.45
Maternal distress	5.51	1.34(1.0-3.8)	.37	6.30	8.97(0-52)	8.34
		Free play task / Teaching task			Three Boxes procedure	
Sensitivity	21.66/22.09	3.47(1-5)/3.47(1-5)	.76/.79	14.08	3.01(1-4)	.74
Detachment	21.66/22.09	1.83(1-5)/1.52(1-5)	.80/.73	14.08	.31(0-1)	.46
Cognitive stimulation	21.66/22.09	2.87(1-5)/3.12(1-5)	.67/.76	14.08	2.71(1-4)	.74
Positive regard	21.66/22.09	3.35(1-5)/3.63(1-5)	.69/.64	14.08	2.82(1-4)	.70
Negative regard	21.66/22.09	1.03(1-3)/1.22(1-4)	.20/.52	14.08	1.25(1-4)	.57
Intrusiveness	21.66/22.09	1.70(1-4)/2.05(1-5)	.75/.87	14.08	1.48(1-4)	.72
		Child Outcomes at 48 mos.			Child Outcomes at 54 mos.	
Effortful control	21.43	2.10(0-4)	1.08	21.99	4.48(0-7)	3.01
Receptive language	23.34	40.30(10-75)	11.14	21.99	98.34(50-139)	19.92
Externalizing (teacher)	42.18	2.01(1-4.6)	.53	43.70	9.96(0-67)	11.93
Externalizing (parent)	15.90	1.26(1-2.5)	.29	22.21	10.05(0-43)	6.74

Note. BONDS = Behavior Outlook Norwegian Developmental Study; NICHD SECCYD = National Institute of Child Health and Human Development Study of Early Child Care and Youth Development; ICNR + Income-to-needs ratio

PARENTING QUALITY CONSTRUCT VALIDITY

Table 2

Fit Indexes for Factor Models of Parenting Quality in both samples

	df	Chi sq.	RMSEA	CFI	TLI
<i>BONDS Data (n=908)</i>					
Across tasks					
Model 1: One-factor with 12 items	48	1332.524*	0.17	0.85	0.80
Model 2: One-factor with 10 items	30	659.609*	0.15	0.91	0.87
Within task					
Model 3: One-factor with 6 items + 6 items	47	650.550*	0.12	0.93	0.90
Model 4: One-factor with 5 items + 5 items ^a	29	119.667*	0.06	0.99	0.98
<i>SECCYD Data (n=1,172)</i>					
Model 1: One-factor with 6 items	22	316.557***	0.17	0.95	0.92
Model 2: One-factor with 5 items ^b	18	50.136***	0.09	0.99	0.98

Note: ^a Best fitting model (BONDS); * $p < .05$; ^b Best fitting model *** $p < .001$ (SECCYD)

PARENTING QUALITY CONSTRUCT VALIDITY

Table 3

Associations of Parenting Quality in BONDS Tasks with Predictors and Outcomes (n=908).

Coefficients (standard errors) are standardized.

	Task 1 and 2 in same model		Task 1 and 2 separate models	
	PQ task 1	PQ task 2	PQ task 1	PQ task 2
Family predictors				
Maternal education	0.17(0.04)***	0.20(0.04)***	0.18(0.04)***	0.19(0.04)***
Maternal distress (6 mos)	-0.05(0.04)	-.06(0.04)	-0.08(0.04)*	-0.05(0.04)
Economic hardship (1 yr)	-0.02(0.04)	-0.06(0.04)	-0.04(0.04)	-0.06(0.04)
Single parenthood (2 yrs)	-0.11(0.04)**	-0.09(0.04)*	-0.12(0.04)***	-0.09(0.04)*
Maternal age	0.06(0.04)	0.14(0.04)***	0.06(0.04)	0.14(0.04)***
Child predictors				
Gender (boy)	-0.09(0.04)**	-0.14(.04)***	-0.12(0.04)***	-0.14(0.04)***
Global developm (1 yr)	0.04(0.04)	0.17(0.04)***	0.07(0.04)	0.17(0.04)***
Child outcomes (4 yrs)				
Effortful control	-0.05(0.05)	0.26(0.05)***	0.14(0.04)***	0.23(0.04)***
Receptive language	-0.10(0.05)*	0.31(0.05)***	0.12(0.04)**	0.24(0.04)***
Externalizing (teacher)	0.02(0.06)	-0.26(0.06)***	-0.18(.05)***	-0.25(.04)***
Externalizing (parent)	0.04(0.05)	-.12(.05)*	-0.07(.04)	-0.10(.04)**

*Note: *p<.05, **p<.01, *** p<.001; PQ = parenting quality*

PARENTING QUALITY CONSTRUCT VALIDITY

Table 4

Fit Indexes for BONDS Models with Predictors and Outcomes (n=1,157)

	df	Chi sq.	RMSEA	CFI	TLI
Task 1 task 2 in one model	145	527.25***	0.05	0.95	0.93
Task 1 only	77	390.76***	0.06	0.93	0.88
Task 2 only	77	367.64***	0.06	0.92	0.88

*Note: ***p<.001*

Table 5

*Associations of Parenting Quality in SECCYD with Predictors and Outcomes**(n=1,364). Coefficients (standard errors) are standardized.*

	Parenting Quality
Family predictors	
Maternal education	0.21(0.04)***
Maternal distress (6 mos)	-0.13(0.04) ***
Income-to-needs (average 6 mos, 1 & 2 yrs)	0.15(0.03) ***
Single parenthood (2 yrs)	-
Maternal age	0.17(0.04) ***
Child predictors	
Gender (boy)	-0.05(0.03) ***
Global development (15 mos)	0.21(0.03) ***
Child outcomes (4.5 yrs)	
Effortful control	0.30(0.04) ***
Receptive language (PLS)	0.50(0.024) ***
Externalizing (teacher)	-0.27(0.04) ***
Externalizing (parent)	-0.19(0.03) ***

*Note: *** p<.001, Fit indexes for the full model: Chi sq (69) = 615.29***, RMSEA=.076, CFI=922, TLI=0.882. The model did not converge with single parenthood included.*

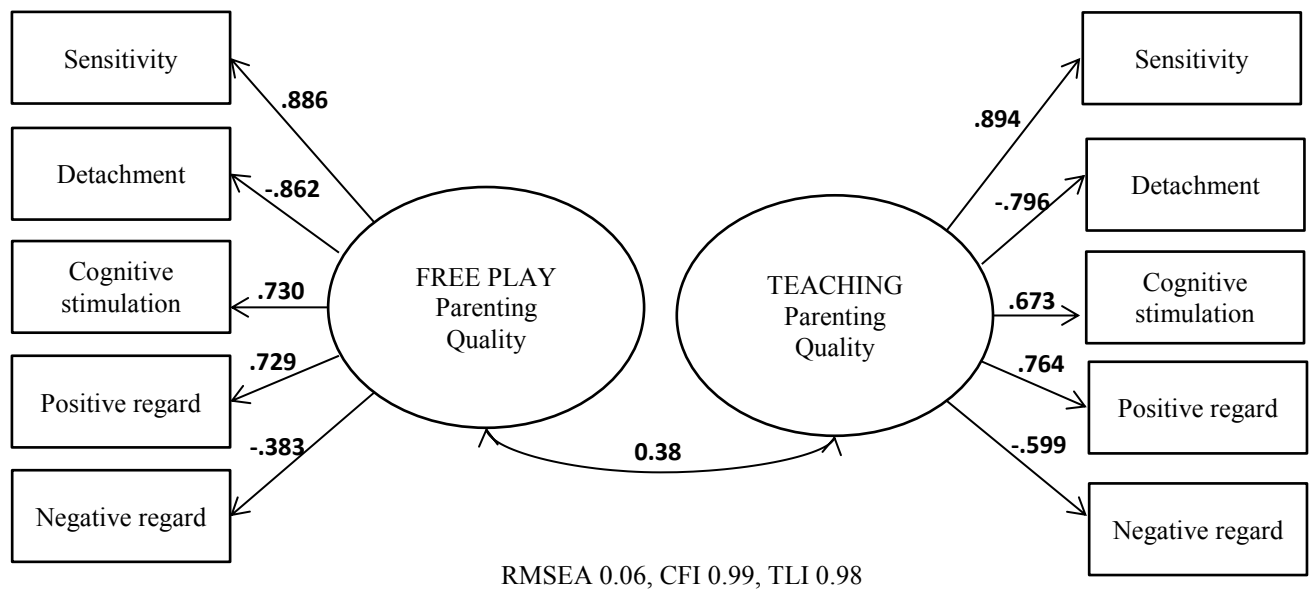


Figure 1. BONDS measurement model of latent parenting quality from free play and teaching tasks

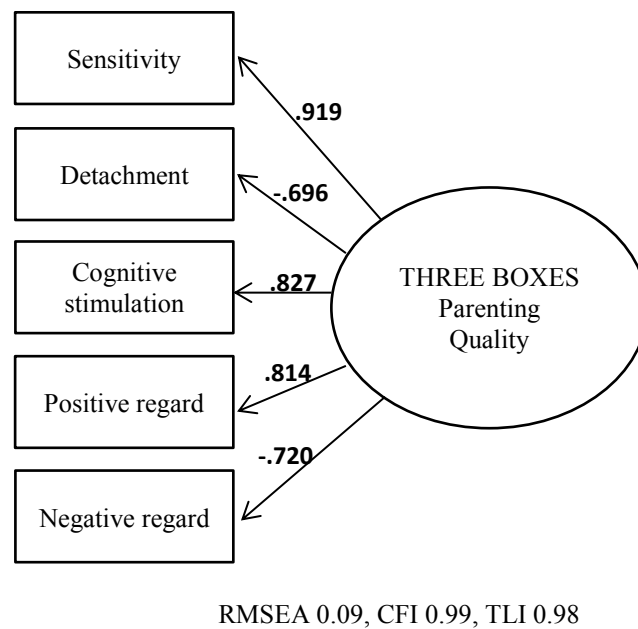


Figure 2. NICHD SECCYD measurement model of latent parenting quality from the Three Boxes procedure

PARENTING QUALITY CONSTRUCT VALIDITY

Supporting Information for Online Appendix

Manuscript: Parenting Quality from Observational Ratings at Age 2: Validation from Norwegian and U.S. Samples.

Table S1

Factor Loadings for BONDS' Preliminary Models as Part of the Fitting Process (n=908)

	Model 1	Model 2	Model 3	Model 4 ^a
Free-play task				
Sensitivity	.920	.847	.964	.886
Intrusiveness	-.383	-	-.432	-
Detachment	-.752	-.820	-.811	-.862
Cognitive stimulation	.631	.673	.704	.730
Positive regard	.630	.678	.696	.729
Negative regard	-.399	-.347	-.449	-.383
Teaching task				
Sensitivity	.990	.833	1.039	.894
Intrusiveness	-.582	-	-.630	-
Detachment	-.663	-.744	-.706	-.796
Cognitive stimulation	.574	.609	.630	.673
Positive regard	.667	.706	.723	.764
Negative regard	-.573	-.533	-.633	-.599

Note: ^aBest fitting model

Table S2

Factor Loadings for SECCYD Preliminary Models as Part of the Fitting Process (n=1,172)

	Model 1	Model 2 ^a
Three Boxes procedure		
Sensitivity	.941	.919
Intrusiveness	-.677	-
Detachment	-.663	-.696
Cognitive stimulation	.798	.827
Positive regard	.793	.814
Negative regard	-.800	-.720

Note: ^aBest fitting model

Credit Author Statement

Kristin Berg Nordahl: Conceptualization, Methodology, Investigation, Resources, Writing - Original Draft, Writing - Review & Editing. **Margaret Tresch Owen:** Conceptualization, Methodology, Resources, Writing - Original Draft, Writing - Review & Editing. **Luisa Antunes Ribeiro:** Conceptualization, Methodology, Formal analysis, Writing - Review & Editing. **Henrik Daae Zachrisson:** Conceptualization, Methodology, Writing - Review & Editing, Project administration, Funding acquisition.