The Association Between Parent-Reported and Observed Parenting: 
A Multi-Level Meta-Analysis

A. M. Hendriks
Vrije Universiteit Amsterdam

D. Van der Giessen, G. J. J. M. Stams, and 
G. Overbeek
University of Amsterdam

The purpose of the present meta-analysis was to investigate the strength of the association between parent-reported and observed parenting, and to investigate which specific characteristics of participants, questionnaires, or observational procedures moderate this association. A systematic search of relevant peer-reviewed articles published between January 2000 and December 2014 yielded 36 articles (N = 8,510) and 89 effect sizes. Results from a 3-level random-effects meta-analysis demonstrated a weak, yet significant, overall association of r = .17 between parent-reported and observed parenting. The magnitude of the effect size depended on questionnaire length (larger effect for more items) and the type of parenting investigated (largest effects for negative parent behaviors, next largest effects for positive parent behaviors, and smallest effect for controlling parent behaviors). In conclusion, this study shows that the strength of the association between parent-reported and observed parenting is small but significant.

Public Significance Statement
Parenting is often measured using self-reports or observations. This study has shown that the association between parenting measured through self-report and through observation is small, but significant. The association is larger for negative and positive parenting compared with controlling parenting and the association is larger for studies using questionnaires with more items.

Keywords: parenting, parent-reports, observation, meta-analysis

Parenting has long been theorized to be important for the development of children (e.g., Campbell, 1997; Caron, Weiss, Harris, & Catron, 2006; Kuppens, Grietens, Onghena, & Michiels, 2009). Parenting is usually assessed through self-report or through observation (e.g., Gardner, 2000; Locke & Prinz, 2002). However, the rationale for choosing one of these methods of measurement is not always clear. Furthermore, whether a higher score on a certain self-reported parenting measure is indicative of a higher score on the same parenting variable when observed, or vice versa, is unclear. Whereas several substantial meta-analyses have found small to medium association between multiple informants for child behavior or child psychopathology (e.g., Achenbach, Krukowski, Dumenci, & Ivanova, 2005; Achenbach, McConaughy, & Howell, 1987; DuBog, Renk, Epstein, & Phares, 2000), to the best of our knowledge the meta-analytic association between different measures of parenting behaviors is still unknown. Therefore, the purpose of the present study is to perform a meta-analysis to investigate the strength of the association between parent-reported and observed parenting.

In assessing parenting, questionnaires and observation instruments are used to measure similar constructs, however, there are differences between questionnaires and observation instruments. For instance, they are used for different purposes. Questionnaires are more often used for the assessment of thoughts, feelings, attitudes, and perceptions of behavior, whereas observational measures are mostly used to assess behavior (Gardner, 2000). In addition, parents and observers differ in their roles toward the child and on the existing knowledge they have on the range of behaviors of the child. For example, parents are more acquainted with the behavior that the child displays in day-to-day interactions compared with observers (Achenbach et al., 2005), which may cause parents and observers to perceive situations differently. Furthermore, assessment types differ in the extent to which social desirability and perceptions influence the parenting scores. Within the construct of parenting, there are elements that increase the likelihood of distortions associated with self-report (Morsbach & Prinz, 2006). For instance, when parents report their own parenting behaviors, they have a tendency to underreport their hostility (Waylen, Stallard, & Stewart-Brown, 2008) and to overreport the extent to which they show warmth and exert control (Bögels & van Melick, 2004; Schwarz, Barton-Henri, & Prazinsky, 1985). Although the different properties of questionnaires and observation

A. M. Hendriks, Department of Biological Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam; D. Van der Giessen, G. J. J. M. Stams, and G. Overbeek, Department of Child Development and Education, Research Priority Area Yield, University of Amsterdam.

Correspondence concerning this article should be addressed to A. M. Hendriks, Department of Biological Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, van den Boechorststraat 1, 1081 BT, Amsterdam, the Netherlands. E-mail: a.m.hendriks@vu.nl
instruments are known, there does not appear to be a common practice for measuring parenting.

How parents report their own parenting behaviors and how they actually behave in interactions is thought to be associated. Yet, it is still unknown how strongly the association between parent-reported and observed parenting actually is, and which factors might increase or decrease the strength of the association. Knowing the strength of the association between parent-reported and observed parenting could inform us on the extent to which self-reports or observations provide similar information on parenting. In addition, knowing which factors are associated with stronger associations between parent-reported and observed parenting is helpful in designing more precise assessments. Overall, this study could provide a first step toward establishing common or “best” practices for measuring parenting.

In this study we examine the strength of the association between parent-reported and observed parenting. With this study, we want to inform research and clinical practice on the extent to which these different measures of parenting are related and thereby provide information on whether they potentially provide different information. In addition, we examine whether the association between parent-reported and observed parenting may be moderated by characteristics of participants, questionnaires, observations, and parenting.

Moderators

Participant characteristics refer to specific parent and child characteristics that may influence the strength of the association between parent-reported and observed parenting, including gender of the parent, sample type for the parents (i.e., general or clinical sample), age of the children, and sample type for the children (i.e., children from the general population, children with a history of maltreatment, ADHD, externalizing behavior problems, or an autism spectrum disorder).

Research on parenting is mostly focused on mothers (e.g., Bögels & Phares, 2008; Phares, 1992; Smith, 2011), but there has been an increase of interest in the influence of fathers (e.g., Calzada, Eyberg, Rich, & Querido, 2004; McBride, Schoppe, &ane, 2002; Phares, 1992). If the association between parent-reported and observed parenting is moderated by whether there are more mothers in a sample, this might have implications for including mothers and fathers in future research. If the association between parent-reported and observed parenting differs for parents from general samples or from clinical samples, this should be considered when measuring parenting in different types of samples. Because as children age, disciplinary techniques and nurturing behaviors of parents change to adapt to the changing needs of their children (Locke & Prinz, 2002), child age might affect the strength of the association between parent-reported and observed parenting, and thus have implications for assessing parenting for different age groups. If whether children are from clinical or general populations samples influences the association between parent-reported and observed parenting, this should be considered when choosing an assessment method for parenting in a sample with children from the general population or a clinical sample.

Questionnaire characteristics consist of specific features that might influence the strength between parent-reported and observed parenting. For this study the length of the questionnaire and which questionnaire is administered are of interest. Psychometric theory indicates that a greater number of items in a measure often leads to increased reliability (e.g., Cronbach, 1951; Smith, McCarthy, & Anderson, 2000), whereas administering fewer questions has economic advantages. Many questionnaires exist to assess parenting, however, it could be possible that parenting scores assessed by different questionnaires are related differently to parenting scores from observations.

Observation characteristics refer to specific aspects of an observation procedure that might influence the strength of the association between parent-reported and observed parenting. Observation characteristics of interest are the duration of the observation, the location of the observation, whether the observation was performed directly or from a video, which observation instrument was used, and which type of task was employed.

There are many decisions following the choice to include an observational measure of parenting. Whether the association between parent-reported and observed parenting is moderated by duration of an observation, the location of the observation, whether the observation was performed directly or from a video, or which task to give to the parents, would indicate that scores on parent-reported and observed parenting are related differently across these variables. Similarly, as with questionnaires, the association between parent-reported and observed parenting could be different for different observational instruments.

Parenting refers to which type of parenting was measured (e.g., positive parenting, negative parenting, and controlling parenting). The literature distinguishes between several parenting dimensions: positive parenting (e.g., warmth, responsiveness, positive affect), negative parenting (e.g., hostility, criticism, intrusiveness), and control (e.g., commands, monitoring, discipline). This categorization follows the dominant differentiation between parenting behaviors in the literature (Baumrind, 1971; Locke & Prinz, 2002; McKee, Colletti, Rakow, Jones, & Forehand, 2008; Steinberg, Mounts, Lamborn, & Dornbusch, 1991). Measuring positive, negative, or controlling parenting could yield different associations between parent reports and observations.

The Present Study

The main aim of this study was to examine to what extent parent-reported and observed parenting were associated. In addition, we explore which participant characteristics, questionnaire characteristics, observation characteristics, and parenting moderated the association between parent-reported and observed parenting.

Method

Articles were collected through a computerized database search of PsycARTICLES, PsycINFO, ERIC, and MEDLINE searching for articles published from January 2000 through December 2014 to cover a period of the last 15 years following the reviews on measurements of parenting by Gardner (2000) and by Locke and Prinz (2002). The search terms were: report, observational, parent, child, adolescent, convergence, correspondence, convergent validity, and correlation put together in different combinations. With these search terms, 14,070 candidate articles were identified. Because an investigation of the association between self-reported and
observed parenting in most studies was not the primary goal, during the first collection of articles abstracts were scanned on whether both questionnaire and observational measures of parenting had been administered. After screening the abstracts, 290 articles were included for further analyses of eligibility. These 290 articles were then scanned to determine whether there were correlations reported either between parent-reported positive parenting and observed positive parenting, parent-reported negative parenting and observed negative parenting, or parent-reported control and observed control, and to determine whether other inclusion criteria were met. Of 33 articles that did not report correlations, authors were contacted (yielding an additional 17 articles to include). To further ensure that all relevant articles were included, lists of references of the useful articles were scanned. These processes together led to the inclusion of 36 articles (N = 8,510) with a total of 89 effect sizes, as summarized in Table 1. Because we conducted secondary analyses on already gathered data, there was no involvement of a research ethics committee.

Criteria for the Selection of Studies

Articles included in the analyses had to meet the following inclusion criteria:

1. The study was published between January 2000 and December 2014.\(^1\)
2. Children in the sample were between 1 and 18 years.\(^2\)
3. The parent who reported parenting behavior did not have any impairment that would restrict behavioral or cognitive functioning, such as autism or intellectual disabilities.
4. The study provided a measure of association between observed positive parenting, negative parenting, or control combined with assessment using a questionnaire that measured parenting behavior on a similar dimension.
5. The report on parenting behavior was filled in by the parent under scrutiny.
6. The reports were not filled in retrospectively.
7. Observations were either directly performed or videotaped.

Moderators

Participant characteristics. The percentage of mothers was coded as a continuous variable, summarizing the percentage of mothers from the sample that reported on the parenting behaviors. Sample type for the parents was coded either as a general population sample or a clinical sample. Age of the children was coded as a continuous variable. Sample type for the children was coded categorically as one of the following categories: general sample, children with maltreatment history, children with ADHD, children with externalizing behavior problems, and children with an autism spectrum disorder. There were no studies including children with internalizing problems.

Questionnaire characteristics. The number of questions to measure the relevant parenting construct was coded as a continuous variable, with the score reflecting the actual number of items in the questionnaires. Which questionnaire was administered was coded categorically with the following categories: Parenting Dimensions Inventory, Parenting Practices Interview, Alabama Parenting Questionnaire, Parenting Scale, scores of multiple questionnaires combined, or other. Each of these categories represented the questionnaires that were used in at least three studies.

Observation characteristics. Duration of the observation was coded as a continuous variable representing the duration in minutes of the concerning observation task. The location of the observation referred to whether an observation took place either at home or in a laboratory setting. Direct observation versus videotaped referred to whether the observation was performed while the interaction took place or later on from a videotape. Which coding instrument was used for the observation was coded categorically with the following categories: Coders Impression Inventory, Dyadic Parent-Child Interaction Coding System, Family Observation System, a combination of multiple observation systems, or other. These categories represented the observation instruments that were used in at least three studies. Which task was coded for the observation was one of the following categories: a play task, a combination of playing and cleaning up, a homework task, a discussion task, observations made during the interview, a compliance task, a problem solving task, or other.

Parenting. Which type of parenting was measured for the association between parent reports and observations of parenting was coded categorically. Categories consisted of positive (e.g., warm, prosocial, supportive parenting), negative (e.g., hostile, critical, harsh parenting), or controlling (e.g., commands, monitoring, discipline).

Coding the Studies

To capture the information regarding moderators, we developed a coding scheme. The moderators were coded according to this coding scheme by the first author. If studies reported associations between parent-reported and observed parenting before and after an intervention, we only included the results from before an intervention. In total, 22% of the articles were randomly selected to be double coded by the second author to check for interrater reliability. For continuous variables, we calculated intraclass correlations. The average intraclass correlation measuring correspondence between the coders was .90 on the continuous variables. The lowest intraclass correlations between the two raters were for the

\(^1\) To make sure that no relevant body of literature before 2000 was missed, we conducted an additional electronic literature search for articles published before 2000. This search yielded 5916 results. Based on the titles and abstracts, we fully scanned 31 articles; only three articles appeared to be eligible. This number of studies was minimal compared with the far greater number of studies found in the period of 2000–2015, and thus indicated that no meaningful proportion of literature was missed due to our time frame.

\(^2\) The review by Locke and Prinz (2002) focused on measures of parenting for children aged 1-18, because parenting during infancy has not yet developed into stable behavioral patterns. Specifically, Lecuyer-Maus (2000) explains that only after a year, mothers and children have practiced their interactions to such an extent that meaningful patterns can be investigated. Thus, because there are significant developmental differences between children younger and older than 1 year, we decided to not include studies that focused on parenting for children below 1 year of age.
## Table 1

### The Included Articles

<table>
<thead>
<tr>
<th>Authors</th>
<th>N</th>
<th>% mothers</th>
<th>Type sample parents</th>
<th>Age children</th>
<th>Type sample children</th>
<th>No. items</th>
<th>Quest.</th>
<th>Dur. obs.</th>
<th>Loc.</th>
<th>Direct</th>
<th>Obs.</th>
<th>Task</th>
<th>Type parenting</th>
<th>(r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asscher Hermans, and Deković (2008)</td>
<td>105</td>
<td>100 G</td>
<td>2.5 G</td>
<td>8</td>
<td>Other</td>
<td>12 H V</td>
<td>Other P</td>
<td>Pos</td>
<td>-.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asscher et al. (2008)</td>
<td>105</td>
<td>100 G</td>
<td>2.5 G</td>
<td>6</td>
<td>PDI</td>
<td>12 H V</td>
<td>Other P</td>
<td>Neg</td>
<td>.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asscher et al. (2008)</td>
<td>105</td>
<td>100 G</td>
<td>2.5 G</td>
<td>8</td>
<td>Other</td>
<td>12 H D</td>
<td>CII</td>
<td>Other Pos</td>
<td>-.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asscher et al. (2008)</td>
<td>105</td>
<td>100 G</td>
<td>2.5 G</td>
<td>6</td>
<td>PDI</td>
<td>12 H D</td>
<td>CII</td>
<td>Other Neg</td>
<td>.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bennett Sullivan, &amp; Lewis (2006)</td>
<td>139</td>
<td>100 G</td>
<td>5 M</td>
<td>5</td>
<td>Other</td>
<td>20 L V</td>
<td>Other PS</td>
<td>Neg</td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bhandari &amp; Barnett (2007)</td>
<td>78</td>
<td>96 G</td>
<td>4.3 G</td>
<td>18</td>
<td>Other</td>
<td>7 L V</td>
<td>Other P</td>
<td>Cont</td>
<td>.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blair et al. (2014)</td>
<td>1,050</td>
<td>99 G</td>
<td>2 G</td>
<td>18</td>
<td>Other</td>
<td>6 H D</td>
<td>CII</td>
<td>Other Neg</td>
<td>.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breitenstein et al. (2012)</td>
<td>532</td>
<td>90 G</td>
<td>2.8 G</td>
<td>22</td>
<td>Other</td>
<td>15 V</td>
<td>DPICS</td>
<td>Pos</td>
<td>.83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breitenstein et al. (2012)</td>
<td>532</td>
<td>90 G</td>
<td>2.8 G</td>
<td>6</td>
<td>Other</td>
<td>15 V</td>
<td>DPICS</td>
<td>Cont</td>
<td>-.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brotman et al. (2005)</td>
<td>92</td>
<td>90 G</td>
<td>4.1 EXT</td>
<td>1</td>
<td>PPI</td>
<td>10 H V</td>
<td>Multiple P</td>
<td>Neg</td>
<td>.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brotman et al. (2005)</td>
<td>92</td>
<td>90 G</td>
<td>15.6 EXT</td>
<td>13</td>
<td>PPI</td>
<td>10 H V</td>
<td>Multiple Neg</td>
<td>.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brotman et al. (2011)</td>
<td>171</td>
<td>90 G</td>
<td>15.6 G</td>
<td>12</td>
<td>PPI</td>
<td>15 H V</td>
<td>Other PC</td>
<td>Cont</td>
<td>-.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Browne, Meunier, O’Connor, and Jenkins (2012)</td>
<td>381</td>
<td>83 G</td>
<td>1.6 G</td>
<td>5</td>
<td>Other</td>
<td>15 H V</td>
<td>Other P</td>
<td>Pos</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chronis-Tuscano et al. (2008)</td>
<td>70</td>
<td>89 C</td>
<td>8.1 ADHD</td>
<td>10</td>
<td>APQ</td>
<td>5 H V</td>
<td>DPICS</td>
<td>Pos</td>
<td>.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chronis-Tuscano et al. (2008)</td>
<td>70</td>
<td>100 C</td>
<td>8.1 ADHD</td>
<td>10</td>
<td>APQ</td>
<td>10 H V</td>
<td>DPICS</td>
<td>HW Pos</td>
<td>.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feinberg Neiderhiser, Howe, and Hetherington (2001)</td>
<td>702</td>
<td>100 G</td>
<td>14.5 G</td>
<td>11</td>
<td>Other</td>
<td>10 H V</td>
<td>Other D</td>
<td>Pos</td>
<td>.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feinberg et al. (2001)</td>
<td>702</td>
<td>100 G</td>
<td>14.5 G</td>
<td>4</td>
<td>Other</td>
<td>10 H V</td>
<td>Other D</td>
<td>Neg</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feinberg et al. (2001)</td>
<td>699</td>
<td>100 G</td>
<td>12.9 G</td>
<td>11</td>
<td>Other</td>
<td>10 H V</td>
<td>Other D</td>
<td>Pos</td>
<td>.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feinberg et al. (2001)</td>
<td>694</td>
<td>0 G</td>
<td>14.5 G</td>
<td>11</td>
<td>Other</td>
<td>10 H V</td>
<td>Other D</td>
<td>Pos</td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feinberg et al. (2001)</td>
<td>694</td>
<td>0 G</td>
<td>14.5 G</td>
<td>4</td>
<td>Other</td>
<td>10 H V</td>
<td>Other D</td>
<td>Neg</td>
<td>.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feinberg et al. (2001)</td>
<td>701</td>
<td>0 G</td>
<td>12.9 G</td>
<td>11</td>
<td>Other</td>
<td>10 H V</td>
<td>Other D</td>
<td>Pos</td>
<td>-.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feinberg et al. (2001)</td>
<td>701</td>
<td>0 G</td>
<td>12.9 G</td>
<td>4</td>
<td>Other</td>
<td>10 H V</td>
<td>Other D</td>
<td>Neg</td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fossum, Morch, Hansdegård, and Drugli (2007)</td>
<td>532</td>
<td>100 G</td>
<td>6.6 EXT</td>
<td>15</td>
<td>PPI</td>
<td>15 H V</td>
<td>DPICS</td>
<td>PC Pos</td>
<td>.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fossum et al. (2007)</td>
<td>532</td>
<td>100 G</td>
<td>6.6 EXT</td>
<td>14</td>
<td>PPI</td>
<td>15 H V</td>
<td>DPICS</td>
<td>PC Neg</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friesen, Woodward, Horwood, and Fergusson (2013)</td>
<td>146</td>
<td>77 G</td>
<td>G</td>
<td>13</td>
<td>Multiple</td>
<td>H D</td>
<td>Other DI</td>
<td>Pos</td>
<td>.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friesen et al. (2013)</td>
<td>146</td>
<td>77 G</td>
<td>G</td>
<td>6</td>
<td>Multiple</td>
<td>H D</td>
<td>Other DI</td>
<td>Pos</td>
<td>.47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hahlweg, Heinrichs, Kuschel, Bertram, and Naumann (2014)</td>
<td>276</td>
<td>100 G</td>
<td>4.5 G</td>
<td>PS</td>
<td>20 H V</td>
<td>FOS PC</td>
<td>Neg</td>
<td>.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hanisch Hautmann, Ptück, Eichelberger, and Döpfner (2014)</td>
<td>155</td>
<td>100 G</td>
<td>4.2 G</td>
<td>13</td>
<td>Multiple</td>
<td>20 H V</td>
<td>CII PC</td>
<td>Pos</td>
<td>.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hawes &amp; Dadds (2006)</td>
<td>56</td>
<td>G</td>
<td>6.3 EXT</td>
<td>14</td>
<td>APQ</td>
<td>60 H V</td>
<td>iOS PC</td>
<td>Pos</td>
<td>.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hawes &amp; Dadds (2006)</td>
<td>56</td>
<td>G</td>
<td>6.3 EXT</td>
<td>14</td>
<td>APQ</td>
<td>20 H V</td>
<td>iOS PC</td>
<td>Pos</td>
<td>.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hill et al. (2008)</td>
<td>335</td>
<td>100 G</td>
<td>4.5 EXT</td>
<td>PDI</td>
<td>20 L V</td>
<td>Other C</td>
<td>Pos</td>
<td>.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hill et al. (2008)</td>
<td>335</td>
<td>100 G</td>
<td>4.5 EXT</td>
<td>PDI</td>
<td>20 L V</td>
<td>Other C</td>
<td>Cont</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hill et al. (2008)</td>
<td>335</td>
<td>100 G</td>
<td>4.5 EXT</td>
<td>Other</td>
<td>20 L V</td>
<td>Other C</td>
<td>Pos</td>
<td>.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hill et al. (2008)</td>
<td>335</td>
<td>100 G</td>
<td>4.5 EXT</td>
<td>Other</td>
<td>20 L V</td>
<td>Other C</td>
<td>Neg</td>
<td>.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hutchings et al. (2007)</td>
<td>151</td>
<td>G</td>
<td>3.9 EXT</td>
<td>PS</td>
<td>30 H V</td>
<td>DPICS PC</td>
<td>Neg</td>
<td>.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Karreman, van Tuijl, van Aken, and Deković (2008)</td>
<td>89</td>
<td>100 G</td>
<td>3 G</td>
<td>PDI</td>
<td>H V Other</td>
<td>P Pos</td>
<td>-.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Karreman et al. (2008)</td>
<td>89</td>
<td>100 G</td>
<td>3 G</td>
<td>PDI</td>
<td>H V Other</td>
<td>P Pos</td>
<td>-.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Karreman et al. (2008)</td>
<td>89</td>
<td>0 G</td>
<td>3 G</td>
<td>PDI</td>
<td>H V Other</td>
<td>P Pos</td>
<td>-.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Karreman et al. (2008)</td>
<td>89</td>
<td>0 G</td>
<td>3 G</td>
<td>PDI</td>
<td>H V Other</td>
<td>P Pos</td>
<td>-.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kohanska, Kim, and Koenig Nordling (2012)</td>
<td>186</td>
<td>100 G</td>
<td>2.5 G</td>
<td>13</td>
<td>Other</td>
<td>55 L D</td>
<td>Other C</td>
<td>Cont</td>
<td>.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*(table continues)*
### Table 1 (continued)

<table>
<thead>
<tr>
<th>Authors</th>
<th>N</th>
<th>% mothers</th>
<th>Type sample parents</th>
<th>Age children</th>
<th>Type sample children</th>
<th>No. items</th>
<th>Quest.</th>
<th>Dur. obs.</th>
<th>Loc.</th>
<th>Direct</th>
<th>Obs.</th>
<th>Task</th>
<th>Type parenting</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linares, Montalto, Li, and Oza (2006)</td>
<td>128</td>
<td>87</td>
<td>G</td>
<td>6,2 M</td>
<td>15</td>
<td>PDI</td>
<td>20</td>
<td>H</td>
<td>D</td>
<td>Other</td>
<td>DI</td>
<td>Pos</td>
<td>.31</td>
<td></td>
</tr>
<tr>
<td>Lui, Johnston, Lee, and Lee-Flynn (2013)</td>
<td>94</td>
<td>100</td>
<td>G</td>
<td>9,8 ADHD</td>
<td>15</td>
<td>APQ</td>
<td>20</td>
<td>L</td>
<td>V</td>
<td>Other</td>
<td>P</td>
<td>Pos</td>
<td>.20</td>
<td></td>
</tr>
<tr>
<td>Metsäpelto &amp; Pulkkinen (2005)</td>
<td>54</td>
<td>100</td>
<td>G</td>
<td>10,5 G</td>
<td>10</td>
<td>Other</td>
<td>20</td>
<td>L</td>
<td>V</td>
<td>Other</td>
<td>PS</td>
<td>Pos</td>
<td>.11</td>
<td></td>
</tr>
<tr>
<td>Metsäpelto &amp; Pulkkinen (2005)</td>
<td>52</td>
<td>0</td>
<td>G</td>
<td>10,5 G</td>
<td>10</td>
<td>Other</td>
<td>20</td>
<td>L</td>
<td>V</td>
<td>Other</td>
<td>PS</td>
<td>Pos</td>
<td>.19</td>
<td></td>
</tr>
<tr>
<td>Morawski &amp; Sanders (2007)</td>
<td>126</td>
<td>100</td>
<td>G</td>
<td>2,2 G</td>
<td>10</td>
<td>PS</td>
<td>30</td>
<td>H</td>
<td>V</td>
<td>PC</td>
<td>Neg</td>
<td>.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent et al. (2014)</td>
<td>180</td>
<td>89</td>
<td>C</td>
<td>11,5 G</td>
<td>15</td>
<td>APQ</td>
<td>30</td>
<td>H</td>
<td>V</td>
<td>Other</td>
<td>DI</td>
<td>Pos</td>
<td>.22</td>
<td></td>
</tr>
<tr>
<td>Perry et al. (2013)</td>
<td>263</td>
<td>100</td>
<td>G</td>
<td>10,5 G</td>
<td>Other</td>
<td>15</td>
<td>H</td>
<td>V</td>
<td>P</td>
<td>Pos</td>
<td>.30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perry et al. (2013)</td>
<td>244</td>
<td>100</td>
<td>G</td>
<td>10,5 G</td>
<td>Other</td>
<td>15</td>
<td>H</td>
<td>V</td>
<td>P</td>
<td>Pos</td>
<td>.32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perry et al. (2013)</td>
<td>228</td>
<td>100</td>
<td>G</td>
<td>10,5 G</td>
<td>Other</td>
<td>15</td>
<td>H</td>
<td>V</td>
<td>P</td>
<td>Pos</td>
<td>.32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scott, O’Connor, et al. (2010)</td>
<td>174</td>
<td>G</td>
<td>5,7 G</td>
<td>2</td>
<td>Other</td>
<td>23</td>
<td>H</td>
<td>V</td>
<td>Other</td>
<td>PC</td>
<td>Pos</td>
<td>−.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scott, O’Connor, et al. (2010)</td>
<td>174</td>
<td>G</td>
<td>5,7 G</td>
<td>2</td>
<td>Other</td>
<td>23</td>
<td>H</td>
<td>V</td>
<td>Other</td>
<td>PC</td>
<td>Pos</td>
<td>.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scott, O’Connor, et al. (2010)</td>
<td>174</td>
<td>G</td>
<td>5,7 G</td>
<td>2</td>
<td>Other</td>
<td>23</td>
<td>H</td>
<td>V</td>
<td>Other</td>
<td>PC</td>
<td>Cont</td>
<td>.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scott, Sylva, et al. (2010)</td>
<td>112</td>
<td>G</td>
<td>5,2 EXT</td>
<td>5</td>
<td>Other</td>
<td>15</td>
<td>H</td>
<td>V</td>
<td>Other</td>
<td>PC</td>
<td>Neg</td>
<td>.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scott, Sylva, et al. (2010)</td>
<td>112</td>
<td>G</td>
<td>5,2 EXT</td>
<td>5</td>
<td>Other</td>
<td>15</td>
<td>H</td>
<td>V</td>
<td>Other</td>
<td>PC</td>
<td>Neg</td>
<td>.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sessa, Avenevoli, Steinberg, and Morris (2001)</td>
<td>84</td>
<td>100</td>
<td>G</td>
<td>5,2 G</td>
<td>8</td>
<td>Other</td>
<td>25</td>
<td>H</td>
<td>V</td>
<td>PC</td>
<td>Pos</td>
<td>.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sessa et al. (2001)</td>
<td>84</td>
<td>100</td>
<td>G</td>
<td>5,2 G</td>
<td>5</td>
<td>Other</td>
<td>25</td>
<td>H</td>
<td>V</td>
<td>PC</td>
<td>Neg</td>
<td>−.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sessa et al. (2001)</td>
<td>84</td>
<td>100</td>
<td>G</td>
<td>5,2 G</td>
<td>5</td>
<td>Other</td>
<td>25</td>
<td>H</td>
<td>V</td>
<td>PC</td>
<td>Cont</td>
<td>−.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sheeber et al. (2012)</td>
<td>70</td>
<td>100</td>
<td>C</td>
<td>4,6 G</td>
<td>Other</td>
<td>20</td>
<td>H</td>
<td>V</td>
<td>Multiple</td>
<td>PC</td>
<td>Neg</td>
<td>−.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sheeber et al. (2012)</td>
<td>67</td>
<td>100</td>
<td>C</td>
<td>4,6 G</td>
<td>Other</td>
<td>20</td>
<td>H</td>
<td>V</td>
<td>Multiple</td>
<td>PC</td>
<td>Neg</td>
<td>−.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tellegen &amp; Sanders (2014)</td>
<td>64</td>
<td>94</td>
<td>G</td>
<td>5,7 ASD</td>
<td>Other</td>
<td>30</td>
<td>H</td>
<td>V</td>
<td>FOS</td>
<td>Neg</td>
<td></td>
<td>−.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Van den Akker, Deković, M., Prinzie, and Asscher (2010)</td>
<td>96</td>
<td>100</td>
<td>G</td>
<td>2,5 G</td>
<td>8</td>
<td>Other</td>
<td>12</td>
<td>H</td>
<td>D</td>
<td>CII</td>
<td>PC</td>
<td>Pos</td>
<td>.10</td>
<td></td>
</tr>
<tr>
<td>Van den Akker et al. (2010)</td>
<td>96</td>
<td>100</td>
<td>G</td>
<td>2,5 G</td>
<td>6</td>
<td>PDI</td>
<td>12</td>
<td>H</td>
<td>D</td>
<td>CII</td>
<td>PC</td>
<td>Pos</td>
<td>.25</td>
<td></td>
</tr>
<tr>
<td>Van den Akker et al. (2010)</td>
<td>96</td>
<td>100</td>
<td>G</td>
<td>2,5 G</td>
<td>8</td>
<td>Other</td>
<td>12</td>
<td>H</td>
<td>V</td>
<td>Other</td>
<td>PC</td>
<td>Pos</td>
<td>.03</td>
<td></td>
</tr>
<tr>
<td>Van den Akker et al. (2010)</td>
<td>96</td>
<td>100</td>
<td>G</td>
<td>2,5 G</td>
<td>6</td>
<td>PDI</td>
<td>12</td>
<td>H</td>
<td>V</td>
<td>Other</td>
<td>PC</td>
<td>Neg</td>
<td>.28</td>
<td></td>
</tr>
<tr>
<td>Van den Akker et al. (2010)</td>
<td>96</td>
<td>100</td>
<td>G</td>
<td>3 G</td>
<td>8</td>
<td>Other</td>
<td>12</td>
<td>H</td>
<td>D</td>
<td>CII</td>
<td>PC</td>
<td>Pos</td>
<td>.06</td>
<td></td>
</tr>
<tr>
<td>Van den Akker et al. (2010)</td>
<td>96</td>
<td>100</td>
<td>G</td>
<td>3 G</td>
<td>6</td>
<td>PDI</td>
<td>12</td>
<td>H</td>
<td>D</td>
<td>CII</td>
<td>PC</td>
<td>Neg</td>
<td>.22</td>
<td></td>
</tr>
<tr>
<td>Van den Akker et al. (2010)</td>
<td>96</td>
<td>100</td>
<td>G</td>
<td>3 G</td>
<td>8</td>
<td>Other</td>
<td>12</td>
<td>H</td>
<td>V</td>
<td>Other</td>
<td>PC</td>
<td>Pos</td>
<td>.11</td>
<td></td>
</tr>
<tr>
<td>Van den Akker et al. (2010)</td>
<td>96</td>
<td>100</td>
<td>G</td>
<td>3 G</td>
<td>6</td>
<td>PDI</td>
<td>12</td>
<td>H</td>
<td>V</td>
<td>Other</td>
<td>PC</td>
<td>Neg</td>
<td>.31</td>
<td></td>
</tr>
<tr>
<td>Van den Akker et al. (2010)</td>
<td>96</td>
<td>100</td>
<td>G</td>
<td>3,3 G</td>
<td>8</td>
<td>Other</td>
<td>12</td>
<td>H</td>
<td>D</td>
<td>CII</td>
<td>PC</td>
<td>Pos</td>
<td>−.05</td>
<td></td>
</tr>
<tr>
<td>Van den Akker et al. (2010)</td>
<td>96</td>
<td>100</td>
<td>G</td>
<td>3,3 G</td>
<td>6</td>
<td>PDI</td>
<td>12</td>
<td>H</td>
<td>D</td>
<td>CII</td>
<td>PC</td>
<td>Neg</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>Van den Akker et al. (2010)</td>
<td>96</td>
<td>100</td>
<td>G</td>
<td>3,3 G</td>
<td>8</td>
<td>Other</td>
<td>12</td>
<td>H</td>
<td>V</td>
<td>Other</td>
<td>PC</td>
<td>Pos</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>Van den Akker et al. (2010)</td>
<td>96</td>
<td>100</td>
<td>G</td>
<td>3,5 G</td>
<td>8</td>
<td>Other</td>
<td>12</td>
<td>H</td>
<td>D</td>
<td>CII</td>
<td>PC</td>
<td>Neg</td>
<td>.02</td>
<td></td>
</tr>
<tr>
<td>Van den Akker et al. (2010)</td>
<td>96</td>
<td>100</td>
<td>G</td>
<td>3,5 G</td>
<td>6</td>
<td>PDI</td>
<td>12</td>
<td>H</td>
<td>D</td>
<td>CII</td>
<td>PC</td>
<td>Neg</td>
<td>−.04</td>
<td></td>
</tr>
<tr>
<td>Van den Akker et al. (2010)</td>
<td>96</td>
<td>100</td>
<td>G</td>
<td>3,5 G</td>
<td>6</td>
<td>PDI</td>
<td>12</td>
<td>H</td>
<td>D</td>
<td>CII</td>
<td>PC</td>
<td>Neg</td>
<td>.13</td>
<td></td>
</tr>
<tr>
<td>Van den Akker et al. (2010)</td>
<td>96</td>
<td>100</td>
<td>G</td>
<td>3,5 G</td>
<td>8</td>
<td>Other</td>
<td>12</td>
<td>H</td>
<td>V</td>
<td>Other</td>
<td>PC</td>
<td>Pos</td>
<td>−.05</td>
<td></td>
</tr>
<tr>
<td>Van den Akker et al. (2010)</td>
<td>96</td>
<td>100</td>
<td>G</td>
<td>3,5 G</td>
<td>6</td>
<td>PDI</td>
<td>12</td>
<td>H</td>
<td>V</td>
<td>Other</td>
<td>PC</td>
<td>Neg</td>
<td>.06</td>
<td></td>
</tr>
</tbody>
</table>

*This table continues...*
variables observation duration, number of positive items, and sample size; their values were, respectively, .75, .82, and .85. For categorical variables, we calculated Cohen’s kappas. The average of the Cohen’s kappas measuring correspondence between the coders was .92 on the categorical variables. The lowest agreement between the two raters was for the variables observation location and sample type parents; their values were, respectively, .74 and .77. For the other categorical values, the raters fully agreed.

Effect Sizes

In order to make generalizations beyond the articles, we chose to use random effect methods for the analyses, yielding a more conservative significance test for combined effects (Field, 2001). Furthermore, it enabled us to estimate the distribution of effect sizes across studies (Thompson & Higgins, 2002). To investigate whether particular covariates or moderators accounted for the heterogeneity of effect sizes between studies, meta-regression was used. To summarize the strength of the associations, Pearson’s correlation coefficient $r$, and Fisher’s $z$ were either obtained or calculated for all the studies. Fisher’s $z$ is used to correct for skewness in the sampling distribution of $r$ (Field, 2001). Analyses were performed with the Fisher’s $z$ corrected effect sizes denoted as $ES_z$. After analyses the outcomes for the significant models were transformed into correlations ($p$) for interpretation purposes.

Publication Bias

Because for most studies the associations between parent-reported and observed parenting were not the primary research interest, we assumed that there was a relatively low risk for publication bias. To check this assumption, we inspected a funnel plot using Fisher’s $z$ transformations, as displayed in Figure 1. The distribution of the effect sizes was symmetrical with the effect sizes of the studies with a larger sample size spread out more narrowly around the estimated overall effect size compared with the studies with a smaller sample size. This demonstrated that there was no bias in the correlations (Lipsy & Wilson, 2001). In addition, we tested asymmetry using Egger’s regression test (Egger, Davey Smith, Schneider, & Minder, 1997), which was not

![Figure 1](image-url)

Figure 1. Funnel plot with Fisher’s $z$ transformed correlations. On the $y$-axis are the standard errors of the studies, with smaller standard errors representing larger sample sizes. On the $x$-axis are the associations between parent-reported and observed parenting.
significant \( z = 0.88, p = .38 \) and therefore provided additional evidence that there was no publication bias present in this meta-analysis.

**Analyses**

Many of the included studies contained multiple effect sizes for different groups or conditions and therefore there was dependence between effect sizes. To account for this dependence, we applied a multilevel approach (Hedges, Tipton, & Johnson, 2010) with a level of sampling variance around the estimated population effect size (Level 1), a level of variance between effect sizes within studies (Level 2), and a level of variance between effect sizes across studies (Level 3; Van den Noortgate, López-López, Martín-Martínez, & Sánchez-Meca, 2013; Wibbelink & Assink, 2015). The estimation of the parameters was performed with a restricted maximum likelihood estimation method (Wibbelink & Assink, 2015). The analyses were performed in R using the metafor package (Viechtbauer, 2010). First, an overall effect size was estimated. Second, to decide whether the variance between effect sizes within studies and the variance between effect sizes were significant, two log-likelihood ratio tests were performed to see whether constraining the variances of level 2 and level 3 to zero significantly deteriorated model fit. Third, we investigated how the variance was distributed across the three levels using the Higgins and Thompson method (Higgins & Thompson, 2002) as proposed by Cheung (2014), showing how much of the variance was a result of sampling variability (Level 1), within-study variance (Level 2), and between-study variance (Level 3).

Next, we performed separate moderator analyses. For the categorical moderators we created dummy variables. We investigated which moderators significantly explained heterogeneity of the effect sizes with univariate analyses. These moderators were included in the final model. To establish each moderator’s unique contribution to the model, we performed a multivariate analysis with a model including all moderators that were significant in the univariate analyses. Finally, the proportion of explained variance obtained by adding the moderators was calculated using the method as described in Cheung (2014).

**Results**

**Overall Effect Size**

The estimate of the true effect size was \( .17 \), which significantly deviated from zero \( (SE_z = .03, t(88) = 5.88, p < .01, \rho = .17, 95\% CI [.11, .22]) \). This indicated that overall, there was a significant but weak association between parent-reported and observed parenting. The \( Q \) statistic was significant, \( Q(88) = 590.57, p < .01 \), indicating that there was significant heterogeneity in the effect sizes not accounted for by this main effects model. The Level 2 variance, or the variance of effect sizes within studies, was \( .01 \). Constraining this variance to zero significantly deteriorated model fit, \( \chi^2(1) = 18.09, p < .01 \), indicating that the Level 2 variance did significantly deviate from zero. The Level 3 variance, or the variance between the effect sizes, was \( .02 \). Constraining this variance to zero also significantly degraded model fit, \( \chi^2(1) = 22.49, p < .01 \), indicating that the Level 3 variance did significantly deviate from zero. Next, we calculated how the total variance was divided across the three levels. The percentage of variance at Level 1, or sampling variance of the effect sizes around the mean estimate, was 14.06\%. The percentage of variance at Level 2 was 26.25\%, indicating that around 26\% of the variance in effect sizes could be attributed to differences between effect sizes within studies. The percentage of variance at Level 3 was 65.15\%, indicating that 65\% of the total variance in associations in this study could be explained by differences between studies. Because the heterogeneity coefficient \( Q \)—containing the heterogeneity at Levels 1, 2, and 3—in our model was significant, the next step was to perform univariate moderator analyses to investigate which moderator variables significantly explained heterogeneity of effect sizes across studies.

**Participant Characteristics**

The \( Q_M \) statistics for whether the moderators were significant, and the \( Q_E \) statistics for whether there was significant residual heterogeneity are displayed in Table 2. None of the examined participant characteristics (i.e., percentage of mothers, type of sample parents, age of the children, and type of sample children) significantly moderated the association between parent-reported parenting and observed parenting.

**Questionnaire Characteristics**

In Table 2, the \( Q_M \) statistics for whether the moderators were significant, and the \( Q_E \) statistics for whether there was significant residual heterogeneity are displayed. Which questionnaire was used was not a significant moderator. Questionnaire length was a significant moderator, indicating that the amount of items had a significant overall effect on the strength of the association between parent-reported and observed parenting. The amount of residual heterogeneity was significant. The coefficients for this model are displayed in Table 3. The intercept for zero items did not significantly deviate from zero. The slope for more items was positive and significant, indicating that the association between parent-reported increased significantly per increase in questionnaire length.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>All the ( Q_M ) Statistics for Whether the Moderators Have a Significant Effect, and the ( Q_E ) Statistics for Whether There Is Residual Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moderator</td>
<td>Number of studies</td>
</tr>
<tr>
<td>% mothers</td>
<td>32</td>
</tr>
<tr>
<td>Type sample parents</td>
<td>36</td>
</tr>
<tr>
<td>Age children</td>
<td>35</td>
</tr>
<tr>
<td>Type sample children</td>
<td>36</td>
</tr>
<tr>
<td>Questionnaire length</td>
<td>27</td>
</tr>
<tr>
<td>Which questionnaire</td>
<td>36</td>
</tr>
<tr>
<td>Duration observation</td>
<td>31</td>
</tr>
<tr>
<td>Location</td>
<td>35</td>
</tr>
<tr>
<td>Direct vs. video</td>
<td>35</td>
</tr>
<tr>
<td>Which observation instrument</td>
<td>32</td>
</tr>
<tr>
<td>Type of task</td>
<td>34</td>
</tr>
<tr>
<td>Type of parenting</td>
<td>36</td>
</tr>
</tbody>
</table>
significant, observed parenting. The amount of residual heterogeneity was
strength of the association between parent-reported parenting and
questionnaire and which type of parenting were included on the
p 4.21, indicating that there was no significant association between parent-
from zero, indicating that in studies that assessed controlling
parenting, there was a small to medium association between
positive and negative parenting were positive and significant,
heterogeneity of the model was significant. The coefficients for the
moderators in this final model was significant, (i.e., type of parenting and type of questionnaire used). The effect
moderators that emerged as significant in the univariate analyses
were questionnaire length and which type of parenting was re-
observation was scored directly or from a videotape, and which
task was used were not found to significantly moderate the asso-
ciation between parent reported and observed parenting.

Observation Characteristics

In Table 2, the Q_M statistics for whether the moderators were
significant, and the Q_K statistics for whether there was significant
residual heterogeneity are displayed. Type of observation instru-
ment used, duration and location of the observation, whether the
observation was scored directly or from a videotape, and which
task was used were not found to significantly moderate the associa-
tion between parent reported and observed parenting.

Parenting

Which type of parenting was reported and observed was a
significant moderator, indicating that there was an overall signifi-
cant effect of the type of parenting. The amount of residual
heterogeneity of the model was significant. The coefficients for the
moderator are displayed in Table 4. The mean effect sizes for
positive and negative parenting were positive and significant,
indicating that for studies reporting either positive or negative
parenting, there was a small to medium association between
parent-reported and observed parenting. On the contrary, the mean
effect size for controlling parenting did not significantly deviate
from zero, indicating that in studies that assessed controlling
parenting there was no significant association between parent-
reported and observed parenting. In short, for studies that assessed
positive or negative parenting behavior, there was a small to
medium significant association between parent-reported and ob-
served parenting.

Model With Significant Moderators

The moderators that were significant in the univariate analyses
were questionnaire length and which type of parenting was re-
ported and observed. Next, we analyzed a model including the
moderators that emerged as significant in the univariate analyses
(i.e., type of parenting and type of questionnaire used). The effect
of the moderators in this final model was significant, Q_M (3) =
4.21, p = .01, indicating that there was an effect of which ques-
tionnaire and which type of parenting were included on the
strength of the association between parent-reported parenting and
observed parenting. The amount of residual heterogeneity was
significant, Q_K (70) = 379.66, p < .01, indicating that there was
heterogeneity of the effect sizes not accounted for by the model.
Figure 2 displays a forest plot of the final model in which the
residual heterogeneity is visualized. The mean effect sizes for this
model are presented in Table 5. In combination with the intercept
for questionnaire length, the mean effect size for negative was
small, but significant. For positive, negative, and controlling par-
tening, the slope for items was significant, indicating that for all
types of parenting, the association between parent-reported and
observed parenting increased significantly with increase in ques-
tionnaire length.

The Level 2 variance of the model with the significant moder-
ators was .01. Constraining this variance to zero significantly
deteriorated model fit, \( \chi^2(1) = 13.78, p < .01 \), indicating that
the Level 2 variance significantly deviated from zero. The Level 3
variance was .02. Constraining this variance to zero significantly
degraded model fit, \( \chi^2(1) = 15.66, p < .01 \), indicating that the
Level 3 variance significantly deviated from zero. Next, we cal-
culated how the total variance was divided across the three levels.
The percentage of variance at Level 1, or the sampling variance
around the mean estimate, was 15.20%. The percentage of var-
ance at Level 2 was 30.47%, indicating that around 30% of the
variance could be assigned to differences between effect sizes
within studies. The percentage of variance at Level 3 was 61.44%,
indicating that 61% of the total variance could be explained by
differences between studies. The increase in the proportion of
Level 2 variance of the model with the moderators compared with
the model without moderators yielded a negative proportion. Ac-
ccording to Cheung (2014), negative values are truncated to zero,
indicating that adding the moderators did not explain more of the
variance between studies. However, calculating the proportion of
Level 3 variance explained by the moderators was .13%, indicating
that adding the moderators explained about 38% of the variance
within studies.

Discussion

The purpose of the present meta-analysis was to investigate the
strength of the association between parent-reported and observed
parenting, and to investigate which variables would moderate this
association. Overall, we found an estimate of the true effect size of
.17, indicating that parent-reported parenting was weakly associ-
ated with observed parenting. Moderator analyses demonstrated
that this effect size was dependent on questionnaire length, and on
the type of parenting measured. Specifically, effect sizes were
stronger (i.e., more positive) for studies that used questionnaires
with more items, and effect sizes were stronger for studies that
assessed negative parenting, compared with studies that assessed
controlling parenting.

The small positive significant association indicates that if a
parent scores higher on a parent-reported assessment of parenting,
this parent will probably also score higher on an observational

---

### Table 3

<table>
<thead>
<tr>
<th>Parameter estimates</th>
<th>ES</th>
<th>SE</th>
<th>t</th>
<th>95% CI</th>
<th>p</th>
<th>( \rho )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (items = 0)</td>
<td>.09</td>
<td>.05</td>
<td>1.71</td>
<td>[.02, .19]</td>
<td>.09</td>
<td>.09</td>
</tr>
<tr>
<td>Slope</td>
<td>.01</td>
<td>.00</td>
<td>2.11</td>
<td>[.00, .02]</td>
<td>.04</td>
<td>.01</td>
</tr>
</tbody>
</table>

*Note.* Number of studies = 28. ES = effect size; SE = standard error; CI = confidence interval.

---

### Table 4

<table>
<thead>
<tr>
<th>Parameter estimates</th>
<th>ES</th>
<th>SE</th>
<th>t</th>
<th>95% CI</th>
<th>p</th>
<th>( \rho )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>.15</td>
<td>.03</td>
<td>5.03</td>
<td>[.09, .21]</td>
<td>&lt;.01</td>
<td>.15</td>
</tr>
<tr>
<td>Negative</td>
<td>.21</td>
<td>.03</td>
<td>6.76</td>
<td>[.16, .29]</td>
<td>&lt;.01</td>
<td>.22</td>
</tr>
<tr>
<td>Controlling</td>
<td>.09</td>
<td>.05</td>
<td>1.69</td>
<td>[.02, .19]</td>
<td>.09</td>
<td>.09</td>
</tr>
<tr>
<td>Slope pos-neg</td>
<td>.06</td>
<td>.03</td>
<td>1.92</td>
<td>[.00, .11]</td>
<td>.06</td>
<td>.05</td>
</tr>
<tr>
<td>Slope pos-cont</td>
<td>-.08</td>
<td>.05</td>
<td>-1.42</td>
<td>[-.18, .03]</td>
<td>.16</td>
<td>-.08</td>
</tr>
<tr>
<td>Slope neg-cont</td>
<td>-.13</td>
<td>.06</td>
<td>-2.31</td>
<td>[-.24, -.02]</td>
<td>.02</td>
<td>-.13</td>
</tr>
</tbody>
</table>

*Note.* Number of studies = 37. ES = effect size; SE = standard error; CI = confidence interval.
assessment, and vice versa. Furthermore, the association provides support for the assumption that parent reports and observations of parenting measure similar underlying constructs (Gardner, 2000). However, the association is weak, indicating that there are more variables explaining variance in parent-reported or observed parenting. For instance, the weak association may indicate that parent reports and observations have different functions (Gardner, 2000), and that they can still assess the same underlying construct but different aspects of this construct (i.e., perceptions and attitudes vs. visible behaviors). Moreover, as in similar research on children (Achenbach et al., 1987, 2005), parents and observers are exposed to different situations when reporting on the subject and construct of interest, which may cause parents and observers to perceive behaviors differently. In addition, the weak association could indicate that scores from parent reports and observations are influenced differently by social desirability. Within self-reports, positive biases can alter how parents report on their own behavior (e.g., Bögels & van Melick, 2004; Morsbach & Prinz, 2006; Table 5)

### Coefficients for the Final Model

<table>
<thead>
<tr>
<th>Parameter estimates</th>
<th>ES</th>
<th>SE</th>
<th>t</th>
<th>95% CI</th>
<th>p</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean positive (items = 0)</td>
<td>.05</td>
<td>.06</td>
<td>.78</td>
<td>[-.07, .17]</td>
<td>.44</td>
<td>.05</td>
</tr>
<tr>
<td>Slope items positive</td>
<td>.01</td>
<td>.00</td>
<td>2.45</td>
<td>[.00, .02]</td>
<td>.02</td>
<td>.01</td>
</tr>
<tr>
<td>Mean negative (items = 0)</td>
<td>.14</td>
<td>.05</td>
<td>2.52</td>
<td>[.03, .24]</td>
<td>.01</td>
<td>.14</td>
</tr>
<tr>
<td>Slope items negative</td>
<td>.01</td>
<td>.00</td>
<td>2.45</td>
<td>[.00, .02]</td>
<td>.02</td>
<td>.01</td>
</tr>
<tr>
<td>Mean controlling (items = 0)</td>
<td>-.00</td>
<td>.08</td>
<td>-.05</td>
<td>[-.16, .15]</td>
<td>.36</td>
<td>-.00</td>
</tr>
<tr>
<td>Slope items controlling</td>
<td>.01</td>
<td>.00</td>
<td>2.45</td>
<td>[.00, .02]</td>
<td>.02</td>
<td>.01</td>
</tr>
</tbody>
</table>

**Note.** ES = effect size; SE = standard error; CI = confidence interval.

Figure 2. Forest plot of the final model. The gray areas represent the values of the association based on the values on the moderators, the black stripes represent the observed associations.

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
observations of these parenting behaviors. For example, parents' self-report of social desirability could be similar for parent reports and observers to perceive the situations less differently compared to other informants, causing parents to rate their parenting behaviors differently from observers. Similarly, it could be that for more salient parenting constructs there is less agreement between parent-reported and observed parenting. An explanation could be that negative behaviors are more salient for parents and observers (e.g., Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Lovejoy, Weis, O'Hare, & Rubin, 2001; Locke & Prinz, 2002), until now the mean strength of the association remained unknown. Moreover, the knowledge that the association is stronger for certain questionnaires and parenting behaviors may provide some insight into the mechanisms that underlie the association between parent-reported and observed parenting. As we have previously described, the association between parent-reported and observed parenting was stronger for more salient types of parenting behaviors. Differences between purposes of measurement instruments, informants, and influences of social desirability may become less important when assessing certain types of parenting, which could indicate that these behaviors are detected more accurately and stably. A second strength of the present study was the use of a three-level analysis model that corrected for the dependency between correlations reported within and across studies. By taking the dependency of effect sizes into account, the reliability for a Type I error (i.e., a false positive finding) improves (Van den Noortgate et al., 2013).

Several limitations of the present study also warrant mentioning. First, a weakness of using meta-analytical techniques is that it uses summary statistics, and therefore nuances might get lost that would have been detected using raw data. Second, the outcomes do not explain what the association between parent-reported and observed parenting actually entails. The strength of the association does indicate that agreement between parents and observers on parenting is weak; this needs to be taken into account when researchers or clinicians want to assess parenting. Nevertheless, this study has provided a starting point for future research by exploring the strength of the association between parent-reported and observed parenting based on 15 years of research.

Future Recommendations and Implications

Although this study provides insight into the strength of the association between parent-reported and observed parenting, it does not—not did it intend to—provide insight in “which measure is best to use.” A suggestion for future research to gain more insight in which measure of parenting is best to use is to investigate which methods are most strongly associated with outcome variables. For instance, in research on child anxiety, several meta-analyses have shown that certain parenting dimensions are more strongly associated with child anxiety disorders (McLeod, Wood, & Weisz, 2007; Möller, Nikolić, Majdandžić, & Bögels, 2016; Van der Brugge, Stams, & Bögels, 2008), and that observed parenting is more strongly associated with child anxiety disorders than parent reports of parenting (McLeod et al., 2007). Although that would not provide insight into which measure provides the most robust and reliable estimates, it may provide some indication of the associations that are most strongly supported by the data.
The most accurate representation of parenting, it would provide insight in which measure is best to use for research on certain outcome variables.

This study takes a first step to investigate the strength of the association between parent-reported and observed parenting, and which moderators influence this strength. One main conclusion of our study is that except for which type of parenting was assessed and questionnaire length, none of the other participant, questionnaire, or observation characteristics that we examined moderated the main effect size. This indicates that the significant-but-small association between parent-reported and observed parenting is a robust finding. The significant portion of unexplained heterogeneity between study findings emphasizes the need for future research in this domain. Awaiting these future endeavors, what has become clear on the basis of our current findings is that the association between parent-reported and observed parenting is relatively weak and varies across types of parenting and parenting questionnaires. Researchers and clinicians should be aware of this weak association when choosing a method to assess parenting.

References

References marked with an asterisk were included in the analyses.


PARENT-REPORTED AND OBSERVED PARENTING


Received April 15, 2016
Revision received March 7, 2017
Accepted April 12, 2017