

# Parenting Assessed by Observation versus Parent-report: Moderation by Parent Distress and Family Socioeconomic Status

Janette E. Herbers<sup>1</sup> · Elisa B. Garcia<sup>2</sup> · Jelena Obradović<sup>2</sup>

Published online: 28 August 2017  
© Springer Science+Business Media, LLC 2017

**Abstract** Assessment of parenting frequently relies on parent self-report despite well-documented discrepancies between this methodology and observational measures. In the current study, we investigated patterns of disagreement between self-reported and observational measures of positive parenting and negative parenting with 102 dyads of parents and 4–6 year-old children. Specifically, we tested whether differences in parent distress and family socioeconomic status (SES) moderated the association between self-reported and observed parenting. Results showed significant interactions with less agreement between self-report and observational measures among parents reporting higher levels of distress and among families from lower SES backgrounds. Effects were more pronounced in models of negative parenting than those of positive parenting. Findings suggest that two factors relevant to parenting quality, emotional distress and SES, predict meaningful differences in manner of parent-report relative to observational measures. We discuss implications for the extant literature and for future efforts regarding effective and efficient parenting measurement.

**Keywords** Parent–child relationship · Socioeconomic status · Observational measures

## Introduction

Developmental scientists investigating processes of parenting frequently rely upon parent-report assessments despite well-documented limitations of this methodology (Morsbach and Prinz 2006). While parent-report measures have strengths in certain instances, it is generally accepted within the field that observational measures of parenting behaviors and parent-child relationships are superior in criterion and predictive validity. However, observational measures are considerably more costly and time-consuming for both researchers and participants (Aspland and Gardner 2003; Margolin et al. 1998; Morsbach and Prinz 2006; Zaslow et al. 2006). As such, understanding precisely how these different methods of assessing parenting behaviors correspond with and differ from each other holds value for advancing the study of parenting. Among the relatively few studies that report associations between methods, correlations of observations and parent-report are inconsistent and low-to-moderate at best. These inconsistencies suggest a more complex relation between observational and parent-report measures that may involve moderation by individual parent and family factors.

Asking parents directly about their parenting has notable benefits, such as convenience, efficiency, and their experience with all interactions that make up the relationship. Parents' report of the nature and quality of their daily interactions with children has superior ecological validity compared to laboratory tasks (Gardner 2000). Self-report is most appropriate when assessing parenting attitudes and beliefs, which are inherently subjective. Utilizing self-report for parenting behaviors, however, introduces systematic sources of bias (Gardner 2000). Like other cases of self-report, parents' reports of their own parenting can be influenced by individual differences in interpretation,

---

✉ Janette E. Herbers  
janette.herbers@villanova.edu

<sup>1</sup> Department of Psychology, Villanova University, 800 E. Lancaster Ave., Villanova, PA 19004, USA

<sup>2</sup> Graduate School of Education, Stanford University, Stanford, CA, USA

memory, and perceptions of both the questionnaire content and the behavior being sampled (Lovejoy et al. 1997; Morsbach and Prinz 2006). Furthermore, social desirability can motivate parents to endorse items generally considered to represent “good parenting,” such as praising or giving affection to children, and avoid items considered “bad parenting,” such as criticizing or spanking children, regardless of their actual behavior (Lovejoy et al. 1997; Sessa et al. 2001). How the parent-report measures are constructed including the syntax, grammar, and vocabulary also can influence responding (Morsbach and Prinz 2006).

Compared to the measurement issues involved in self-reported parenting behaviors, ratings of observed parenting behaviors by trained researchers are considered more reliable. While observer biases can emerge, the measurement error is more likely to be random than that of parent-report and more practically addressed through careful development of coding schemes and criteria for inter-rater reliability (Gardner 2000; Repetti et al. 2015). Particularly when parent-child interactions are conducted under structured conditions with task demands that mimic everyday situations, such as having parents and children play together or clean up toys, raters can assess parenting behaviors according to carefully defined coding schemes. Thus all parents within a sample are scored against the same criteria. Observational measures generally show better concurrent and predictive validity in relation to other methods of assessing parenting (e.g., child report) and to child outcomes (Gonzales et al. 1996; Scott et al. 2011; Zaslow et al. 2006). Among different types of observational coding schemes, those targeting small units of behavior with limited inference, or micro-social approaches, are less susceptible to halo effects and other rater bias than more global coding schemes (Alexander et al. 1995). In a separate analysis using the current sample, we demonstrated that micro-social observational coding of the parent-child relationship outperformed global observational codes in predicting teacher reports of child behavior (Bardack et al. 2017).

Of course, conducting observational assessments places greater demands on participating parents and children, who engage in the interaction tasks while being video-recorded, and especially on research teams, who must train to reliability and code the data. Observational methods are not always feasible, especially with large samples. As such, the field would benefit from a better understanding of the patterns and moderators of concordance between widely used self-reports of parenting and more costly ratings of parenting behaviors observed directly.

Multi-method studies of parenting that report direct comparisons of measures are not well represented in the literature (Hurley et al. 2014; Morsbach and Prinz 2006; Winsler et al. 2005). Self-report and observational measures of parenting show inconsistent concordance, with

correlations ranging from non-significant (Dekovic et al. 1991; Sessa et al. 2001) to low-to-moderate (Hawes and Dadds 2006; Kochanzka et al. 1989; Lui et al. 2013; Scott et al. 2011; Zaslow et al. 2006). Evidence of associations between self-report and observational measures are reported more often in multi-method assessments of negative indices rather than positive indices of parenting. Associations may be weaker in positive compared to negative indices due to greater social desirability of positive parenting items that may restrict the range of measurement (Hawes and Dadds 2006; Scott et al. 2011; Zaslow et al. 2006). Validity data for self-reported parenting measures commonly consist only of comparison to other self-reported measures of parenting, parental symptoms, or child emotional and behavioral problems (Hurley et al. 2014). As such, relatively little is known regarding predictors of agreement between observational and self-reported measures of parenting.

Relevant findings can be found in the studies that have examined validity of parent report of children's behavior problems. In particular, concerns about the influence of parents' depression symptoms on their reports of children's externalizing and internalizing problems, often referred to as *depression-distortion*, have prompted studies examining these links since at least 1967. In 1992, Richters described inconsistencies in the literature considering whether parents with depressive disorders or higher scores on continuous scales of depression symptoms would report exaggerated behavior problems in their children. He also attributed some of the inconsistency to important methodological shortcomings; many studies failed to include alternative methods of assessing child behavior problems for comparison, and others that did use criterion constructs from observational methods or other reporters had poor conceptual matching of the constructs from the differing methods (Richters 1992). Richters thus recommended improving future studies by considering whether symptoms of depression correlate with disagreement between methods as the outcome of interest, and carefully matching constructs in parent-report and criterion measures.

Following his recommendations, researchers have since produced more rigorous evidence suggesting that symptoms of depression in parents do explain some discrepancies between parent-report and alternative methods of assessing child behavior problems (Briggs-Gowan et al. 1996; De Los Reyes and Kazdin 2006; Durbin and Wilson 2012; Fergusson et al. 1993; Gartstein et al. 2009; Harvey et al. 2013; Leerkes and Crockenberg 2003; Najman et al. 2001). In the majority of studies that have considered depression-distortion, internalizing symptoms of parents were associated with reporting more problems for children relative to other reporters or observer ratings.

Given evidence that parent-report can be biased by maternal depression, it is surprising that similar work

regarding parent report of parenting behavior and parent-child relationships is scarce. It is likely that the inconsistent pattern of concordance with observational and other methods of assessing parenting denotes complexity. Parent factors of distress and SES are especially relevant to parenting quality and tend to inter-correlate, and thus may moderate the strength of the concordance between parent report and observational measures of parenting. Specifically, parents who report symptoms of depression or distress may perceive their own parenting as more problematic than it actually is, similar to the distortion hypothesis regarding child behavior. Likewise, parents from different socioeconomic backgrounds may interpret questionnaire items differently, such that some items considered socially desirable or undesirable among families of higher SES might not have the same meaning for lower SES families (Hoff et al. 2002).

Few studies have considered possible sources of distortion in parenting measures. In a clinical sample of families with children diagnosed with ADHD, parents who endorsed more symptoms of depression reported their parenting as more negative than their children did (Chi and Hinshaw 2002). Also among parents of children with ADHD, those with more of their own symptoms of hyperactivity and impulsivity tended to rate their parenting more positively than observers did (Lui et al. 2013). Sessa and colleagues (2001) found that ethnicity moderated the relation between mother and observer report of parenting within a diverse community sample of preschool-aged children such that agreement with observer ratings was lower among African American mothers than Caucasian mothers. In a recent cluster analysis of observational and parent-report parenting measures for over 400 demographically diverse dyads with kindergarten-aged children, there emerged one subtype of parents whom authors labeled “self-critical” because parent-report scores were below average while observational scores were average (Heberle et al. 2015). Compared to the other four subgroups, the self-critical parents had higher mean scores on measures of depression and anxiety.

Concerns regarding how accuracy of parent-report may differ by SES also have been raised but not investigated (Hoff et al. 2002). Specifically, established differences in aspects of parenting by SES, such as lower SES parents generally demonstrating less self-efficacy regarding parenting and a tendency to be more controlling and punitive, could have implications for measurement agreement (Bradley and Corwyn 2002). Lower SES parents could evaluate impacts of certain parenting behaviors differently than observational rating schemes developed by researchers, judging them more negatively or positively. Research has shown that parents of higher SES backgrounds tend to have parenting attitudes that align more closely with their

actual parenting behaviors and more closely with prevailing expert advice than do parents of lower SES (Hoff et al. 2002). These potential sources of measurement disagreement based on SES have not been tested empirically.

The sparse and inconsistent literature investigating patterns of concordance between parent-report and observational measures of parenting leaves much to be explained. As with considerations of parent-report of child behavior problems, it is likely that aspects of measurement and factors that differ between families contribute to these discrepancies. Understanding the role of factors that are relevant to parent-child relationships, particularly parents' symptoms of emotional distress and family SES, may clarify the utility and limitations of parent-report for assessing the quality of parent-child relationships.

In the current study, we address the substantial gap in understanding the patterns of concordance between self-report and observational measures of parenting using data from a community sample of kindergarten-aged children and their parents. With detailed, micro-social coding of observational data from structured laboratory parent-child interaction tasks and two parent-report questionnaires, we carefully matched observational criterion constructs of two particular aspects of parenting, negative control and positive control, to parent-reports of similar constructs. Our sample was relatively diverse in family backgrounds including a range of SES and a range of parent-reported distress, enabling us to test whether these factors known to be associated with parenting quality moderated the concordance between observational and self-report measures.

## Method

### Participants

Participants included 102 children ranging in age from 4–6 years ( $M = 5.61$ ,  $SD = 0.56$ ) and their primary caregivers, the majority of whom were mothers (93%). Families indicated their interest in participating by responding to advertisements at community centers, preschools, elementary schools and libraries in the San Francisco Bay Area. The sample represented a range of income levels and ethnic backgrounds. Children were 52% male and 64% ethnic minority (26% Hispanic or Latino, 20% Asian, 14% Multiracial/Other, and 4% Black). Seventeen percent of parents reported being single parents, and regarding educational background, 13% reported having a high school degree or less, 36% reported having an associate's or bachelor's degree, and 42% a more advanced degree. Family income levels ranged from less than \$12,000 (5.2%) to \$200,000 or

more (36.1%), reflecting the high cost of living in the Bay area.

## Procedure

Assessments involved a 3-hour protocol with concurrent individual child and parent sessions followed by a series of structured parent–child interaction tasks. Parents completed interviews and questionnaires regarding family demographics, parenting practices, and child behavior and functioning. Both parents and children participated in behavioral tasks and physiological assessments not reported in the current analyses (Obradović et al. 2016). The parent–child interaction consisted of four structured tasks intended to elicit parenting behaviors in a variety of situations. First, parents were instructed to play with their children while enforcing a rule that children could play with toys from only one shelf while not touching a set of attractive toys from another shelf. Next, parents were given a magazine and instructed to read while asking their children to clean up the toys. In a problem-solving task, parents and children discussed a problem that parents had previously identified as a salient issue for their family. Common problems for discussion were following rules and getting along with brothers and sisters. Finally, parents were instructed to help children complete a series of geometric puzzles, using *Tangoes* cards and blocks. Interactions were video-recorded for observational coding.

## Measures

Parent behavior and child behavior during the interaction tasks were evaluated separately by two raters trained to reliability by a master rater who also coded 20% of the cases for child behavior and a separate 20% for parent behavior. All parent behavior was categorized on a second-by-second basis into one of four mutually exclusive codes: positive control, negative control, following the child's lead, and disengaged. Positive control involved behaviors such as giving instructions, teaching, redirecting, praise, and limit-setting when accompanied by a positive or neutral affective tone. Negative control involved criticisms, hostility, physical interventions, or any other behaviors accompanied by a negative or harsh affective tone. Behaviors considered to be following the child's lead were active listening, reflecting, and other behaviors in which the parent was engaged with the child and positive or neutral in tone but not directing the child in any way. The disengaged code applied when the parent was distracted, withdrawn, looking away, or otherwise not interacting with the child.

Similarly, all child behavior was categorized as active on-task, passive on-task, withdrawn, or defiant/dysregulated. Child behavior was considered active on-task when

the child was actively participating in task activities by speaking to the parent or leading play through actions such as manipulating materials. Passive on-task behaviors occurred when the child was appropriately engaged but not leading, such as listening to the parent with eye contact or following specific directions. Withdrawn was coded when the child was distracted, avoiding eye contact or turning away, or otherwise disengaged with both the parent and the task at hand. Finally, defiant/dysregulated behaviors included disobedience, any hostility or aggression directed towards the parent or materials, whining, and displays of negative affect. Observer accuracy was calculated based on the kappa statistic and observed base rates of behavior in the sample (Bruckner and Yoder 2006). All parent and child codes had observer accuracy scores above 90%.

### *Negative control (observed)*

Our variable for observed negative control was calculated as the proportion of time during the parent-child interaction that the parent was coded as using negative control as described previously, regardless of the corresponding child's behavior. The proportion of total interaction time parents used negative control ranged from .00 to .17, with  $M = .02$  and  $SD = .03$ .

### *Negative parenting (self-report)*

The composite for self-report of negative parenting was based upon two subscales of the Parent Practices Interview (PPI; The Incredible Years Project 2011): harsh and inconsistent discipline (15 items,  $\alpha = .73$ ) and physical punishment (6 items,  $\alpha = .70$ ). Parents responded to items such as "how often do you raise your voice (scold or yell)" and "how often does your child get away with things that you feel s/he should have been disciplined for?" using a seven-point scale ranging from *never* to *always*. Item-level scores from the two subscales were averaged to create a composite score of self-reported negative parenting (21 items,  $\alpha = .74$ ).

### *Positive control (observed)*

We calculated observed positive control to represent the proportion of each child's total time off-task that the parent was concurrently coded as using positive control. In contrast to negative control behaviors that we consider insensitive regardless of child behavior, we assert that positive control behaviors are sensitive in response to children's withdrawn or dysregulated behavior but not necessarily sensitive when a child is actively or passively on-task and constructively self-directed. Scores for proportion of time

during interaction with observed positive control ranged from .18 to .79, with  $M = .50$  and  $SD = .14$ .

*Positive parenting (self-report)*

Parents completed the Coping with Child’s Negative Emotions Scale (CCNES; Fabes et al. 1990). The CCNES includes six subscales, three of which correspond conceptually with positive control when a child is off-task: expressive encouragement (12 items,  $\alpha = .89$ ), emotion-focused reactions (12 items,  $\alpha = .77$ ), and problem-focused reactions (12 items,  $\alpha = .77$ ). Each item is rated on a seven-point scale ranging from *very unlikely* to *very likely*. We created a composite of positive parenting as an average of the three subscale scores ( $\alpha = .73$ ).

*Parent distress*

Parents’ emotional distress was assessed via self-report using the Center for Epidemiological Studies-Depression Scale (CES-D; Radloff 1977). The CES-D consists of 20 items each with a four-point response scale ranging from *rarely or none of the time* to *most or all of the time* in the past week. Sample items include “I was bothered by things that usually don’t bother me,” “I felt depressed,” and “I felt sad.” The total score was calculated as the average of responses to all 20 items ( $\alpha = .89$ ).

*Family SES*

A composite score representing family socioeconomic status consisted of the average standardized z-scores from parent responses to questionnaire items regarding total family income (raw score  $M = \$125,041$ ,  $SD = \$72,701$ ) and the primary caregivers’ education in years (raw score  $M = 16.3$ ,  $SD = 3.01$ ).

**Data Analyses**

Complete data were available for child age, gender, and the composite of family SES. Small amounts of data (2–4%) were missing for the following: self-reported positive and negative control, observed positive and negative control, and parent distress. Chained multiple imputation was implemented using the *mi impute* command in Stata 14 to generate 20 complete datasets. Bivariate correlations and hierarchical regression analyses were conducted on each of the 20 datasets, and results including coefficients, standard errors, and  $R^2$  values were combined.

**Results**

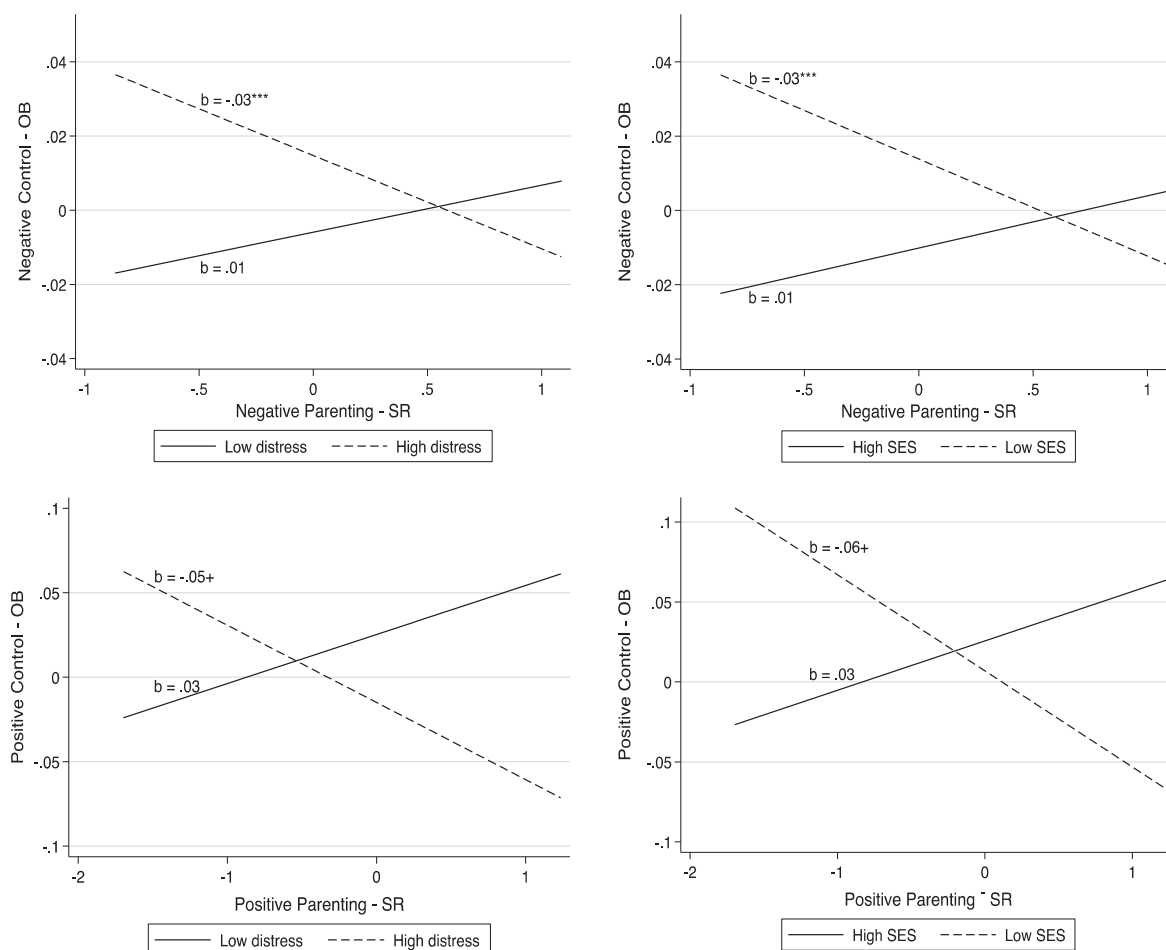
**Associations between Self-Reported and Observational Parenting**

Bivariate correlations of all study variables are presented in Table 1. The association between self-reported negative parenting and observed negative control was not significant, nor were the associations between self-reported and observational measures of positive parenting. Parent distress correlated significantly with negative parenting based on self-report ( $r = .29$ ,  $p < .01$ ) as well as negative control based on observation ( $r = .39$ ,  $p < .001$ ). Correlations between parent distress and each method for assessing positive parenting were not significant. Household SES was negatively correlated with self-reported positive parenting ( $r = -.27$ ,  $p < .01$ ) and observed negative control ( $r = -.49$ ,  $p < .001$ ), but not with self-reported negative parenting ( $r = -.03$ ,  $p = .73$ ) or observed positive control ( $r = .15$ ,  $p = .13$ ). While child gender was associated with observed positive control, neither age nor gender influenced the regression results described next.

**Table 1** Bivariate correlations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. Negative control—OB	–						
2. Negative parenting—SR	–.08	–					
3. Positive control—OB	.18	–.24*	–				
4. Positive parenting—SR	–.29**	.04	–.09	–			
5. Parent distress	.39***	.29**	–.06	–.11	–		
6. Family SES	–.49***	–.03	–.27**	.15	–.53***	–	
7. Child gender	–.03	–.03	.20*	–.04	.02	–0.01	–
8. Child age	–.16	–.09	–.02	–.13	–.15	.07	–.08

OB observed, SR self-report  
 \* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$



**Fig. 1** Interaction effects from hierarchical regression models. (OB observed, SR self-report).  $^{+}p < .10$ ;  $^{***}p < .001$

### Moderation by distress and SES

To test whether the association between self-reported and observational measures of parenting was moderated by parent distress or household SES, we conducted four separate hierarchical linear regression models (see Fig. 1). All predictor variables were mean-centered to aid interpretation. Model 1a considered negative parenting moderated by parent distress. The first step included self-reported negative parenting, parent distress, family SES, and covariates of child age and gender predicting observed negative control. In the second step, we added the interaction term of self-reported negative parenting and parent distress. Model 1b considered negative parenting moderated by family SES in the second step. Models 2a and 2b followed the same procedure but with the measures of positive parenting in place of negative parenting. When interaction terms showed trends or significance, we probed interactions with post-hoc analyses to assess significance of simple slopes (Preacher et al. 2006).

### Negative control

Results from the models predicting observed negative control are presented in Table 2. In the first step with main effects only, self-reported negative parenting was not a significant predictor ( $\beta = -.16$ ,  $p = .12$ ). SES was a significant predictor ( $\beta = -.36$ ,  $p < .01$ ) and parent distress showed a trend towards significance ( $\beta = .24$ ,  $p = .05$ ).

In Model 1a, the interaction between self-reported negative parenting and distress emerged as significant in the second step ( $\beta = -.27$ ,  $p = .02$ ) and accounted for an additional 6.7% of the variance in observed parenting. The negative coefficient indicates that concordance between self-reported negative parenting and observed negative control was lower among parents who reported more distress. The simple slope of observed negative control on self-reported negative parenting was significant and negative one standard deviation above the mean of distress ( $b = -.03$ ,  $p < .001$ ), but non-significant one standard deviation below the mean ( $b = .01$ ,  $p = .12$ ).

**Table 2** Hierarchical regressions predicting observed negative control

	Step 1: Main effects		Model 1a: Parent distress interaction		Model 1b: Family SES interaction	
	B (SE)	$\beta$	B (SE)	$\beta$	B	$\beta$
Intercept	.024 (.004)		.027 (.004)		.025 (.004)	
Gender (female)	-.001 (.006)	-.024	-.002 (.005)	-.032	-.002 (.005)	-.027
Age in years	-.007 (.005)	-.125	-.005 (.005)	-.093	-.005 (.005)	-.096
Negative parenting (SR)	-.011 (.007)	-.155	-.007 (.007)	-.096	-.006 (.006)	-.093
Parent distress	.019 (.010)	.243 <sup>+</sup>	.025 (.009)	.320**	.016 (.009)	.208 <sup>+</sup>
Family SES	-.012 (.004)	-.358**	-.010 (.004)	-.308**	-.013 (.004)	-.376**
Distress $\times$ SR			-.038 (.015)	-.274*		
SES $\times$ SR					.017 (.008)	.268*
$\Delta R^2$			.067*	.072*		
Total $R^2$	.305**		.372**		.377**	

SR self-report. All continuous variables mean centered

<sup>+</sup> $p < .10$ ; \* $p < .05$ ; \*\* $p < .01$

**Table 3** Hierarchical regressions predicting observed positive control

	Main effects		Model 2a: Parent distress interaction		Model 2b: Family SES interaction	
	B (SE)	$\beta$	B (SE)	$\beta$	B	$\beta$
Intercept	.499 (.022)		.498 (.021)		.506 (.022)	
Gender (female)	-.008 (.027)	-.029	-.008 (.029)	-.028	-.009 (.029)	-.031
Age in years	-.043 (.027)	-.171	-.043 (.027)	-.167	-.033 (.028)	-.130
Positive parenting (SR)	-.019 (.024)	-.088	-.011 (.025)	-.051	-.017 (.025)	-.079
Parent distress	-.027 (.046)	-.075	-.044 (.046)	-.123	-.019 (.045)	-.054
Family SES	.016 (.019)	.104	.011 (.019)	.070	.014 (.019)	.092
Distress $\times$ SR			-.085 (.042)	-.214*		
SES $\times$ SR					.048 (.027)	.208 <sup>+</sup>
$\Delta R^2$			.042 <sup>+</sup>	.043 <sup>+</sup>		
Total $R^2$	.063**		.105**		.106**	

SR self-report. All continuous variables mean centered

<sup>+</sup> $p < .10$ ; \* $p < .05$ ; \*\* $p < .01$

For Model 1b, the interaction term for self-reported negative parenting by SES was significant ( $\beta = .27, p = .03$ ) and accounted for an additional 7.2% of the variance in observed parenting. The positive coefficient indicates that concordance between self-reported negative parenting and observed negative control was higher among parents of higher SES. The simple slope of observed negative control on self-reported negative control was significant one standard deviation above the mean of SES ( $b = -.03, p < .05$ ) but not one standard deviation below the mean ( $b = -.01, p = .13$ ).

*Positive control*

There were no significant main effects of self-reported positive parenting, SES, or parent distress in the first step

predicting observed positive control (see Table 3). In Model 2a, the interaction term for parent distress by self-reported positive parenting emerged as significant in the second step ( $\beta = -.21, p = .04$ ), with a negative coefficient indicating that concordance between self-reported positive parenting and observed positive control was lesser at higher levels of parent distress. The simple slope of observed positive control on self-reported positive parenting showed a trend for significance one standard deviation above the mean of distress ( $b = -.05, p < .10$ ), but not one standard deviation below the mean ( $b = .03, p = .36$ ).

In Model 2b, the interaction of self-reported positive parenting by SES showed a trend toward significance ( $\beta = .21, p = .08$ ) and explained an additional 4.3% of the variance in observed positive control. Simple slopes showed a trend towards significance one standard deviation below

the mean of SES ( $b = -.06, p = .07$ ), suggesting that concordance was lower among families of lower SES.

## Discussion

Overall, the associations between self-reported and observational measures of parenting differed by level of parent distress and family socioeconomic status. Specifically, concordance between the methods was worse for parents who reported more depression symptoms and parents of lower SES. These patterns of interaction may indicate a number of underlying issues relevant to understanding measurement of parenting. Our study examined both positive and negative aspects of parenting in contrast with much of the extant work involving only one or the other. Furthermore, we utilized detailed, micro-social observational methods in contrast with more commonly used and less precise global coding approaches (Alexander et al. 1995; Bardack et al. 2017).

Regarding parent distress, the findings may reflect a phenomenon similar to depression-distortion in the literature on child behavior problems. Symptoms of depression may influence how parents perceive not only their children's behavior, but also their responses to and management of that behavior. As a moderator, distress does not simply predict worse self-report scores but changes the association between self-report and observed scores. Distressed parents who were rated by observers favorably might have been overly critical of themselves, as the depression-distortion hypothesis would predict. In contrast, distressed parents rated less favorably by observers tended to describe their parenting more positively than did non-distressed parents with similar observer ratings. We speculate that, among parents demonstrating less competent parenting behaviors, distress may inflate their self-perception of parenting because they are preoccupied with personal emotional distress while lacking insight for how their distress influences their parenting behaviors. This possibility warrants further empirical investigation.

Regarding family SES, differential concordance may reflect cultural differences in beliefs about parenting self-efficacy, how individuals interpret questionnaire items, and what individuals consider favorable and socially desirable parenting behavior. We found significant disagreement in parent-report and observational measures of parenting among lower SES families such that lower SES families who were rated more favorably by observers reported less favorable parenting. Parents from lower SES backgrounds who were observed as demonstrating more positive control may lack parenting self-efficacy and thus undervalue and underreport competence-promoting parenting (Hoff et al. 2002). On the other hand, lower SES families rated less

favorably by observers reported more favorable parenting than higher SES parents rated less favorably by observers. This could reflect the tendency of some lower SES parents to engage in punitive and controlling behaviors that they may consider appropriate. Considering that most researchers come from relatively privileged backgrounds, it is likely that higher SES parents interpret measure items in a manner more similar to the interpretations of the researchers who design self-report measures and coding schemes to evaluate observed parenting. For example, warmth in response to children's negative affect may be viewed as nurturing and supportive by researchers and parents of higher SES, but may be viewed as permissive or "spoiling" by parents of lower SES, who tend to value obedience and social conformity to a greater degree than do higher SES parents (Hoff et al. 2002). Thus, lower SES parents who demonstrate sensitivity in interactions with their children might give themselves less favorable self-reports that would conflict with observer ratings.

Relatedly, it is possible that the discordance between self-report and observational measures among parents with greater distress scores and lower SES backgrounds indicates that such parents may be less skilled at "faking good" both on self-reports and during observational tasks. However, the literature indicates that reactivity to observation is not a substantial source of measurement error in observational assessments of parenting (Aspland and Gardner 2003; Gardner 2000). Self-presentation bias is considerably more likely in self-report than observational methods (Gardner 2000; Repetti et al. 2015). As another alternative, we acknowledge the possibility that discordance could be driven by bias on the part of researchers, who may inadvertently inflate parenting scores based on apparent SES differences. However, it is unlikely that such bias accounts for the findings due to our use of micro-social momentary coding procedures, which have been shown to reduce halo effects compared to more global observational coding schemes (Alexander et al. 1995; Bardack et al. 2017). Observer bias also fails to explain why low SES parents with more favorable observational ratings would self-report their parenting in a less favorable manner compared to low SES parents with less favorable observational ratings.

It is important to note that parent distress and family SES were significantly correlated, sharing 28% of variance. As such, the moderation effects by SES could partially reflect impacts of parent distress through their common variance. Alternatively, distress could reflect effects of SES. Because distress and SES do not always co-occur, however, it is useful to understand how each factor independently interacts with self-report in predicting observational measures. Both distress and family SES can serve as markers of likely differences in measurement concordance, and one or both



factors may be salient to future studies depending upon their populations of interest. While these findings highlight challenges with different methods for assessing parenting, they also offer progress for understanding and accounting for well-documented discrepancies in results from differing methods of parenting assessment.

### Implications for Future Research

Results of the current study have several implications for measurement of parenting quality. First, limitations of self-reported parenting are apparent not only due to lack of significant bivariate correlations with observational measures, but also because patterns of agreement differ by highly relevant parenting covariates. We found that observational measures that were carefully matched to items and subscales in parent self-report measures did not correlate significantly. Presuming that the observational measures function as a criterion, we infer that information provided by parents through self-report should be considered in the context of both family SES and parent distress. At minimum, when researchers are predicting outcomes such as child behavior from self-reported parenting, parent distress and SES should be measured and included as covariates. Interaction terms of self-reported parenting by distress and SES might provide additional clarity when statistical power permits their inclusion.

These patterns of disagreement in parenting are most salient when researchers seek to examine parenting quality within populations that range substantially in levels of parent distress and family SES. As such, the issue is not limited to clinical or other high-risk samples. In fact, concordance between measures may actually be improved when samples are more homogeneous. For this reason, future studies that must rely on parent self-report would benefit from the development of measures that show validity not only within high risk or clinical samples but across samples that range in risk and cultural background. Ideally, validity data should demonstrate associations between parent-report and criterion measures such as observational coding that are similar rather than varying across cultural groups and across ranges of parent distress and SES. One consideration for the development of such measures would be to reduce differences in socially desirable responding, either by determining that perceived social desirability of individual items is consistent across levels of parent distress and SES or by identifying items that differentiate parenting quality with less face validity. Cultural differences in what constitutes positive or negative parenting also require further explication and development.

Another potentially fruitful avenue for future research might identify alternative methods for assessing parenting

quality that do not depend upon parent self-report or the intensive and costly aspects of many observational measures. For example, recent work shows great promise for the Five Minute Speech Sample, a measure that utilizes brief audio recordings of open-ended parent responses to prompts that can be coded by trained researchers with demonstrable reliability and validity as an assessment of parent-child relationship quality (Sher-Censor 2015; Weston et al. 2016). Adapting and validating such methods for use in more naturalistic settings could further enhance their utility (Gardner 2000).

The observed patterns of moderation are noteworthy not only because they represent the complexity of measurement issues, but also because parent distress and family SES predict differences in parenting quality in the broader developmental literature. It is likely that measurement issues highlighted by our study both exaggerate and obscure the links between parenting, parent emotional distress, and SES within the literature that attempts to summarize and aggregate findings across studies that have utilized different parenting measurement approaches. Our findings indicate that measurement of parenting via observation and via self-report should not be considered equivalent. As such they should not be aggregated conceptually without consideration of how family characteristics like parent distress and SES influence measurement agreement. Conceptual reviews that carefully separate findings by method of parenting assessment would better characterize what is known.

Similarly, studies that include both self-report and observational measures with discrepant patterns of results may be better understood with attention to moderated effects. For example, studies assessing impacts of parenting interventions often identify change in self-reported, but not observational parenting (Bor et al. 2002; Eddy et al. 1998). In some cases, the intervention might be impacting relevant family characteristics other than parenting, such as parent distress, that drive these findings. Furthermore, the changes apparent in self-reported parenting may reflect processes of change in parent attitudes as distress remits or perceptions of parenting styles and self-efficacy change. These changes in parent attitudes and perceptions may then lead to improvements in observed parenting in the future.

### Strengths and Limitations

Noteworthy strengths of the current study include high quality observational parenting measures as well as a relatively diverse community sample. In order to rely upon the observational coding as a criterion measure, we employed a micro-social coding scheme with small units of behavior and limited inference to reduce likelihood of rater bias (Alexander et al. 1995; Bardack et al. 2017).

The information generated by coders with good reliability was then aggregated carefully to best correspond with constructs assessed by self-report measures of both positive and negative control. With our community sample, we were able to observe differences in agreement by a range of parent distress and family SES.

Despite these strengths, we were limited by a fairly small sample size of 102 dyads. While SES did range substantially, the sample was generally of lower risk with most parents reporting educational backgrounds of college degrees or higher. It is possible that patterns would differ with more representation at the lower end of SES. With 93% of our small sample consisting of mothers as primary caregivers, we lacked statistical power to assess possible differences between agreement of methods with maternal vs. paternal reports. We also utilized the CES-D measure of depression with continuous scores to represent current levels of parent emotional distress rather than attempting to assess clinical depression or differentiate parents who might qualify for diagnoses. Our findings suggest that differences across the range of distress are meaningful, though it may be the case that qualitative differences would emerge between groups by depression diagnosis.

It is also important to note that, despite our best efforts to achieve conceptual mapping between the self-reported and observational parenting constructs, it is possible that some disagreement attributed to measurement may instead have arisen from conceptual mismatch. As a specific example, parents reported on inconsistency as part of the negative parenting measure while observational coding of negative control did not include such behaviors. Furthermore, findings based on our specific focus of positive and negative control may not generalize to other aspects of parenting, such as affective warmth and responsiveness. We also recognize the need for caution when presuming that our observational measures of positive and negative control represent the best criterion. While its short-comings are well documented, self-report of parenting maintains certain strengths. Parents clearly have the most experience with their own parent-child relationships and the largest samples of behavior from which to draw. They can consider aspects of the parent-child relationships across all contexts whereas observational measure rely upon what can be seen or inferred during brief, highly structured laboratory tasks. For example, seemingly self-critical parents, whose observational scores were more favorable than self-reports, may have been reporting problematic behaviors or interactions that occurred in their daily lives but were not apparent in the structured laboratory tasks.

In summary, we found that parent distress and family SES moderated the concordance between self-report and observational measured of parenting in terms of both positive control and negative parenting. These patterns of

moderation likely reflect differences in parents' interpretations of behavior, self-report measure items, cultural perceptions, and aspects of social desirability. Efforts to further understand the nature of discrepancies among parenting measurement methods have potential to improve our understanding of the extant research on parenting and to inform future study designs.

**Author Contributions** J.H.: collaborated with study design and coding of observational data, assisted with data analysis, and wrote the paper. E.G.: assisted with coding of observational data, analyzed the data, and collaborated in editing the manuscript. J.O.: designed and executed the larger study, assisted with data analyses, and collaborated in writing and editing the manuscript.

**Funding** This study was funded in part by grants from the William T. Grant Foundation and the Canadian Institute for Advanced Research.

#### Compliance with Ethical Standards

**Conflict of Interest** The authors declare that they have no competing interests.

**Ethical Approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

**Informed Consent** Informed consent was obtained from all individual participants included in the study.

## References

- Alexander, J. F., Newell, R. M., Robbins, M. S., & Turner, C. W. (1995). Observational coding in family therapy process research. *Journal of Family Psychology, 9*, 355–365. doi:[10.1037/0893-3200.9.4.355](https://doi.org/10.1037/0893-3200.9.4.355).
- Arnold, D. S., O'Leary, S. G., Wolff, L. S., & Acker, M. M. (1993). The parenting scale: A measure of dysfunctional parenting in discipline situations. *Psychological Assessment, 5*, 137–144. doi:[10.1037/1040-3590.5.2.137](https://doi.org/10.1037/1040-3590.5.2.137).
- Aspland, H., & Gardner, F. (2003). Observational measures of parent-child interaction: An introductory review. *Child and Adolescent Mental Health, 8*, 136–143.
- Bardack, S., Herbers, J.E., & Obradović, J. (2017). Unique contributions of dynamic versus global measures of parenting predicting children's adjustment to school. *Journal of Family Psychology*. Advance online publication. doi:[10.1037/fam0000296](https://doi.org/10.1037/fam0000296).
- Bor, W., Sanders, M. R., & Markie-Dadds, C. (2002). The effects of the triple P-positive parenting program on preschool children with co-occurring disruptive behavior and attentional/hyperactive difficulties. *Journal of Abnormal Child Psychology, 30*, 571–587. doi:[10.1023/A:1020807613155](https://doi.org/10.1023/A:1020807613155).
- Boyle, M. H., & Pickles, A. (1997). Maternal depressive symptoms and ratings of emotional disorder symptoms in children and adolescents. *Journal of Child Psychology and Psychiatry, 38*, 981–992. doi:[10.1111/j.1469-7610.1997.tb01615.x](https://doi.org/10.1111/j.1469-7610.1997.tb01615.x).

- Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, *53*, 371–399.
- Briggs-Gowan, M. J., Carter, A. S., & Schwab-Stone, M. (1996). Discrepancies among mother, child, and teacher reports: Examining the contributions of maternal depression and anxiety. *Journal of Abnormal Child Psychology*, *24*, 749–765. doi:10.1007/BF01664738.
- Bruckner, C. T., & Yoder, P. (2006). Interpreting kappa in observational research: Baserate matters. *American Journal on Mental Retardation*, *111*, 433–441. doi:10.1352/0895-8017.
- Chi, T. C., & Hinshaw, S. P. (2002). Mother–child relationships of children with ADHD: The role of maternal depressive symptoms and depression-related distortions. *Journal of Abnormal Child Psychology*, *30*, 387–400. doi:10.1023/A:1015770025043.
- Dekovic, M., Janssens, J. M., & Gerris, J. R. M. (1991). Factor structure and construct validity of the block child rearing practices report (CRPR). *Psychological Assessment*, *3*, 182–187. doi:10.1037/1040-3590.3.2.182.
- De Los Reyes, A., & Kazdin, A. E. (2006). Informant discrepancies in assessing child dysfunction relate to dysfunction within mother–child interactions. *Journal of Child and Family Studies*, *15*, 643–661. doi:10.1007/s10826-006-9031-3.
- Durbin, C. E., & Wilson, S. (2012). Convergent validity of and bias in maternal reports of child emotion. *Psychological Assessment*, *24*, 647. doi:10.1037/a0026607.
- Eddy, J. M., Dishion, T. J., & Stoolmiller, M. (1998). The analysis of intervention change in children and families: Methodological and conceptual issues embedded in intervention studies. *Journal of Abnormal Child Psychology*, *26*, 53–69. doi:10.1023/A:1022634807098.
- Fabes, R. A., Eisenberg, N., & Bernzweig, J. (1990). *The coping with children's negative emotions scale: Procedures and scoring*. Available from authors. Arizona State University.
- Fergusson, D. M., Lynskey, M. T., & Horwood, L. J. (1993). The effect of maternal depression on maternal ratings of child behavior. *Journal of Abnormal Child Psychology*, *21*, 245–275. doi:10.1007/BF00917534.
- Gardner, F. (2000). Methodological issues in the direct observation of parent–child interaction: Do observational findings reflect the natural behavior of participants? *Clinical Child and Family Psychology Review*, *3*, 185–198.
- Gartstein, M. A., Bridgett, D. J., Dishion, T. J., & Kaufman, N. K. (2009). Depressed mood and maternal report of child behavior problems: Another look at the depression–distortion hypothesis. *Journal of Applied Developmental Psychology*, *30*, 149–160. doi:10.1016/j.appdev.2008.12.001.
- Gonzales, N. A., Cauce, A. M., & Mason, C. A. (1996). Interobserver agreement in the assessment of parental behavior and parent–adolescent conflict: African American mothers, daughters, and independent observers. *Child Development*, *67*, 1483–1498. doi:10.1111/j.1467-8624.1996.tb01809.x.
- Harvey, E. A., Fischer, C., Weieneth, J. L., Hurwitz, S. D., & Sayer, A. G. (2013). Predictors of discrepancies between informants' ratings of preschool-aged children's behavior: An examination of ethnicity, child characteristics and family functioning. *Early Childhood Research Quarterly*, *28*, 668–682. doi:10.1016/j.ecresq.2013.05.002.
- Hawes, D. J., & Dadds, M. R. (2006). Assessing parenting practices through parent-report and direct observation during parent-training. *Journal of Child and Family Studies*, *15*, 555–568. doi:10.1007/s10826-006-9029-x.
- Heberle, A. E., Briggs-Gowan, M. J., & Carter, A. S. (2015). A person-oriented approach to identifying parenting styles in mothers of early school-age children. *Infant and Child Development*, *24*, 130–156. doi:10.1002/icd.1888.
- Hoff, E., Laursen, B., & Tardif, T. (2002). Socioeconomic status and parenting. In M. H. Bornstein (Ed.), *Handbook of parenting Volume 2: Biology and ecology of parenting* (pp. 231–252). Mahwah, NJ: Erlbaum.
- Hurley, K. D., Huscroft-D'Angelo, J., Trout, A., Griffith, A., & Epstein, M. (2014). Assessing parenting skills and attitudes: A review of the psychometrics of parenting measures. *Journal of Child and Family Studies*, *23*, 812–823. doi:10.1007/s10826-013-9733-2.
- Kochanzka, G., Kuczynski, L., & Radke-Yarrow, M. (1989). Correspondence between mothers' self-reported and observed child-rearing practices. *Child Development*, *60*, 56–63. doi:10.1111/j.1467-8624.1989.tb02694.x.
- Leerkes, E. M., & Crockenberg, S. C. (2003). The impact of maternal characteristics and sensitivity on the concordance between maternal reports and laboratory observations of infant negative emotionality. *Infancy*, *4*, 517–539. doi:10.1207/S15327078IN0404\_07.
- Lovejoy, M. C., Verda, M. R., & Hays, C. E. (1997). Convergent and discriminant validity of measures of parenting efficacy and control. *Journal of Clinical Child Psychology*, *26*, 366–376. doi:10.1207/s15374424jccp2604\_5.
- Lui, J. H., Johnston, C., Lee, C. M., & Lee-Flynn, S. C. (2013). Parental ADHD symptoms and self-reports of positive parenting. *Journal of Consulting and Clinical Psychology*, *81*, 988. doi:10.1037/a0033490.
- Margolin, G., Oliver, P. H., Gordis, E. B., O'Hearn, H. G., Medina, A. M., Ghosh, C. M., & Morland, L. (1998). The nuts and bolts of behavioral observation of marital and family interaction. *Clinical Child and Family Psychology Review*, *1*, 195–213. doi:10.1023/A:1022608117322.
- Morsbach, S. K., & Prinz, R. J. (2006). Understanding and improving the validity of self-report of parenting. *Clinical Child and Family Psychology Review*, *9*, 1–21. doi:10.1007/s10567-006-0001-5.
- Najman, J. M., Williams, G. M., Nikles, J., Spence, S., Bor, N., O'Callaghan, M. O., & Shuttleswood, J. (2001). Bias influencing maternal reports of child behavior and emotional state. *Social Psychiatry and Psychiatric Epidemiology*, *36*, 186–194. doi:10.1007/s001270170062.
- Obradović, J., Portilla, X. A., & Ballard, P. J. (2016). Biological sensitivity to family income: Differential effects on early executive functioning. *Child Development*, *87*, 374–384. doi:10.1111/cdev.12475.
- Parent Practices Interview. (2011) The Incredible Years Project. University of Washington Parenting Clinic. <http://www.incredibleyears.com/Measures/ppi-ilt3.pdf>.
- Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, *31*, 437–448. doi:10.3102/10769986031004437.
- 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.
- Repetti, R. L., Reynolds, B. M., & Sears, M. S. (2015). Families under the microscope: Repeated sampling of perceptions, experiences, biology, and behavior. *Journal of Marriage and Family*, *77*, 126–146. doi: 10.1111/jomf.12143.
- Richters, J. E. (1992). Depressed mothers as informants about their children: A critical review of the evidence for distortion. *Psychological Bulletin*, *112*, 485–499. doi:10.1037/0033-2909.112.3.485.
- Scott, S., Briskman, J., & Dadds, M. R. (2011). Measuring parenting in community and public health research using brief child and parent reports. *Journal of Child and Family Studies*, *20*, 343–352. doi:10.1007/s10826-010-9398-z.

- Sessa, F. M., Avenoli, S., Steinberg, L., & Morris, A. (2001). Correspondence among informants on parenting: Preschool children, mothers, and observers. *Journal of Family Psychology, 15*, 53–68. doi:[10.1037/0893-3200.15.1.53](https://doi.org/10.1037/0893-3200.15.1.53).
- Sher-Censor, E. (2015). Five minute speech sample in developmental research: A review. *Developmental Review, 36*, 127–155. doi:[10.1016/j.dr.2015.01.005](https://doi.org/10.1016/j.dr.2015.01.005).
- Weston, S., Hawes, D. J., & Pasalich, D. S. (2016). The five minute speech sample as a measure of parent-child dynamics: Evidence from observational research. *Journal of Child and Family Studies*. Advance online publication. doi:[10.1007/s10826-016-0549-8](https://doi.org/10.1007/s10826-016-0549-8).
- Winsler, A., Madigan, A. L., & Aquilino, S. A. (2005). Correspondence between maternal and paternal parenting styles in early childhood. *Early Childhood Research Quarterly, 20*, 1–12. doi:[10.1016/j.ecresq.2005.01.007](https://doi.org/10.1016/j.ecresq.2005.01.007).
- Zaslow, M. J., Weinfield, N. S., Gallagher, M., Hair, E. C., Ogawa, J. R., Egeland, B., & De Temple, J. M. (2006). Longitudinal prediction of child outcomes from differing measures of parenting in a low-income sample. *Developmental Psychology, 42*, 27. doi:[10.1037/0012-1649.42.1.27](https://doi.org/10.1037/0012-1649.42.1.27).