



# HHS Public Access

Author manuscript

*Dev Psychol.* Author manuscript; available in PMC 2017 November 01.

Published in final edited form as:

*Dev Psychol.* 2016 November ; 52(11): 1805–1819. doi:10.1037/dev0000165.

## Does early childhood callous-unemotional behavior uniquely predict behavior problems or callous-unemotional behavior in late childhood?

**Rebecca Waller,**

Department of Psychology, University of Michigan, USA

**Thomas J. Dishion,**

Department of Psychology, Arizona State University, USA

**Daniel S. Shaw,**

Department of Psychology, University of Pittsburgh, USA

**Frances Gardner,**

Department of Social Policy and Intervention, University of Oxford, UK

**Melvin N. Wilson,** and

Department of Psychology, University of Virginia, USA

**Luke W. Hyde**

Department of Psychology, Center for Human Growth and Development, Institute for Social Research, University of Michigan, USA

### Abstract

Callous unemotional (CU) behavior has been linked to behavior problems in children and adolescents. However, few studies have examined whether CU behavior in *early childhood* predicts behavior problems or CU behavior in *late childhood*. This study examined whether indicators of CU behavior at ages 2–4 predicted aggression, rule-breaking, and CU behavior across informants at age 9.5. To test the unique predictive and convergent validity of CU behavior in early childhood, we accounted for stability in behavior problems and method effects to rule out the possibility that rater biases inflated the magnitude of any associations found. Cross-informant data were collected from a multi-ethnic, high-risk sample ( $N = 731$ ; female = 49%) at ages 2–4 and again at age 9.5. From age 3, CU behavior uniquely predicted aggression and rule-breaking across informants. There were also unique associations between CU behavior assessed at ages 3 and 4 and CU behavior assessed at age 9.5. Findings demonstrate that early-childhood indicators of CU behavior account for unique variance in later childhood behavior problems and CU behavior, taking into account stability in behavior problems over time and method effects. Convergence with a traditional measure of CU behavior in late childhood provides support for the construct validity of a brief early childhood measure of CU behavior.

---

Correspondence to: Luke W. Hyde, Department of Psychology, 530 Church Street, University of Michigan, Ann Arbor, MI 48109, USA; lukehyde@umich.edu.

**Conflict of interest statement:** No conflicts declared.

## Keywords

bifactor; callous-unemotional; conduct problems; measurement; prevention

---

Studies have demonstrated that many children and adults who show antisocial behavior, such as violence and substance use, exhibited behavior problems that emerged early in life from age 2 onwards (e.g., Campbell, 1995; Shaw, Gilliom, Ingoldsby, & Nagin, 2003). However, most young children with early behavior problems naturally desist from these behaviors (Côté, Vaillancourt, LeBlanc, Nagin, & Tremblay, 2006). Thus, the goal of preventative efforts is to identify families of children at high risk for early-starting and *stable* trajectories of behavior problems, particularly children likely to persist in such behaviors beyond the preschool years (Dishion et al., 2008; Webster-Stratton & Taylor, 2001). The toddler years represent a focal point for preventative efforts targeting behavior problems because they are years of rapid transition in children's physical and cognitive abilities that can pose challenge to parents in terms of responding to and managing corresponding child behaviors (Dishion et al., 2008; Shaw & Shelleby, 2014). Moreover, treatment research suggests that interventions implemented prior to school age, when behavior is potentially more malleable, are particularly efficacious (Dishion & Patterson, 1992; Reid, Webster-Stratton, & Baydar, 2004). However, across developmental stages, behavior problems are heterogeneous in etiology and persistence, posing challenges for how best to personalize treatment components (Frick & Morris, 2004).

## Callous unemotional (CU) behavior

One approach to identify children at risk for *persistent* behavior problems has focused on the presence of callous unemotional (CU) behavior (Frick, Ray, Thornton, & Kahn, 2014). CU behavior consists of low levels of empathy and guilt, and uncaring towards others (Frick & Morris, 2004), and predicts the development of severe behavior problems, particularly during late childhood and adolescence (Frick et al., 2014). Thus, assessing early childhood behaviors that tap the CU behavior construct may identify those children likely to *persist* in their behavior problems and could help to inform the creation of specialized treatment components (Waller, Gardner, & Hyde, 2013). Previous studies examining the predictive validity of CU behavior have typically focused on samples assessed in late childhood or adolescence. However, recent studies have suggested that CU behavior in the toddler and preschool years also predicts later behavior problems. In the first study to examine CU behavior in preschoolers, Kimonis and colleagues (2006) found that CU behavior predicted teacher-reported proactive aggression 9–12 months later in a sample of 2–5 year olds (Kimonis et al., 2006). However, initial aggression was not included in models, making it difficult to know whether CU behavior was simply indexing children with more severe *existing* behavior problems.

## Does early CU behavior at ages 2–4 uniquely predict behavior problems at age 9.5?

More recent studies have controlled for stability of behavior problems, demonstrating that early childhood CU behavior uniquely predicts increases in behavior problems over time. For example, CU behavior predicted increases in behavior problems one year later among a large, community sample of children aged 4–9 years old (Dadds, Fraser, Frost, & Hawes, 2005). In addition, two separate studies that used the same “brief-adapted” five-item parent-reported measure found that CU behavior at age 3 predicted later high, stable trajectories of teacher-reported aggression from ages 6–12 years old (Willoughby, Mills-Koonce, Gottfredson, & Wagner, 2014) and teacher-reported externalizing behavior at age 6 (Waller, Hyde, Grabell, Alves, & Olson, 2015). Moreover, this five-item measure of early CU behavior was related to important criterion variables, including low empathy and moral regulation (Waller, Hyde et al., 2015) and was preceded by lower temperamental fear and distress during infancy (Willoughby, Waschbusch, Moore, & Propper, 2011). Taken together, these studies demonstrate the separability of CU behavior from other dimensions of early behavior problem during early childhood and the value of CU behavior for predicting future externalizing behaviors. However, with the exception of the efforts of Willoughby and colleagues (2014), no previous studies initiated during early childhood have examined the predictive validity of CU behavior with follow-up periods exceeding three years. Moreover, previous studies suggest that CU behavior may be more strongly related to covert forms of antisocial behavior and proactive forms of aggression (Kimonis et al., 2006; Waller, Hyde et al., 2015; Waller, Wright et al., 2015), although differential prediction of a covert/overt distinction in behavior problems has yet to be tested across the early to late childhood period.

The first goal of the current study was thus to examine whether an early childhood measure of CU behavior predicted behavior problems in the late childhood period. In doing so, a brief-adapted measure of CU behavior was used that had previously been validated in the current sample at ages 2–4 (Hyde et al., 2013). The measure was termed deceitful-callous (DC) behavior because parent-reported items tapping both callousness (e.g., “doesn’t seem guilty after misbehaving”) and deceitfulness (e.g., “lies”, and “sneaky”) loaded together in exploratory and confirmatory factor analyses. At age 3, this measure of DC behavior robustly predicted stable, high trajectories of child behavior problems using latent growth curve modeling from ages 2–4 both within and across informants (Hyde et al., 2013). However, we have yet to test whether DC behavior predicts behavior problems in late childhood over and above the stability of behavior problems. To explore a potential overt/covert distinction in later behavior problems, we examined aggression and rule-breaking as somewhat separable outcomes of behavior problems, while controlling for earlier behavior problems.

## Does early childhood CU behavior uniquely predict late childhood CU behavior?

Beyond the question of whether DC behavior adds variance in the prediction of later behavior problems, a second question centers on *construct validity*. CU behavior has been shown to exhibit moderate stability within middle-late childhood (Obradovi , Pardini, Long, & Loeber, 2007) and during the preschool years (Hyde et al., 2013; Willoughby et al., 2011). However, no studies have tested whether brief-adapted CU behavior measures in early childhood are uniquely related to CU behavior assessed in late childhood via what are considered “gold standard” or purpose-developed measures. The 24-item Inventory of Callous-Unemotional Traits (ICU; Frick, 2004) is a widely-used measure that provides a full assessment of the affective deficits linked to CU behavior. The ICU exhibits a three-factor bifactor (3FBB) structure, with items simultaneously loading onto three “specific” factors (callous, uncaring, and unemotional) and a “general” CU behavior factor. Psychometric support for the 3FBB model has been demonstrated in the current sample at age 9.5 (Waller et al., 2015) and other samples assessed during adolescence (Essau, Sasagawa, & Frick, 2006).

However, no studies have examined whether early childhood CU behavior is related to the specific (uncaring, callous, and unemotional) or general CU behavior factors of the ICU within a bifactor framework. Moreover, studies have yet to examine the convergence of CU behavior measures in early versus late childhood, taking into account overlap with behavior problems. As no prior studies have examined links between early childhood CU behavior measures and the 3FBB model of the ICU, we considered analyses to be exploratory. Nevertheless, we hypothesized that higher DC behavior scores in early childhood would be uniquely related to higher scores on the general factor of the 3FBB ICU model given that both measures were derived to assess the same underlying and general construct capturing variance in callousness and uncaring for others. Finally, a valid concern about our DC behavior measure, or indeed of other brief measures of CU behavior, is that they are simply an index of severe behavior problems, which would explain any prediction of later behavior problems or convergence with later CU behavior measures. To alleviate this concern, and isolate specificity or *unique convergence* of measures of CU behavior over time, we also examined cross-lagged models that accounted for overlap with behavior problems in both *early* and *late* childhood.

## Does early CU behavior predict late-childhood outcomes accounting for method effects?

Despite the methodological rigor of controlling for earlier behavior problems, a lingering concern surrounding any predictive or construct validity of CU behavior measures centers on rater biases. In particular, no parent-reported behavior scales represent wholly *objective measures* of behavioral constructs, but reflect instead, to a greater or lesser extent, inherent biases derived from the parent’s perceptions of the child and their own personal memories, values, or mental states. One component of a parent’s behavior ratings is their tendency to report negative child attributes. It seems particularly pertinent to consider the meaning of

ratings when parents endorse their young child for behaviors central to the CU construct, including lying, sneakiness, lack of empathy, or reduced guilt. That is, are parents actually reporting on the emergence of “callousness” or do ratings reflect parents’ own negative biases or attributions about the child? Concerns about method effects highlight the need for studies that parse out parents’ ratings of children’s externalizing behavior and a tendency to rate the child negatively from variance that specifically taps the underlying CU behavior construct.

A separate and large body of literature has highlighted limitations of behavior rating scales and the need to account for method variance (e.g., Campbell & Fiske, 1959; Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Studies have examined multi-method, multi-trait (MMMT) models that use multiple “methods” (i.e., different informants/assessment methods) to obtain data on “traits” (i.e., underlying constructs the measure intends to assess) (Marsh & Grayson, 1995). MMT models represent a novel approach to test whether early childhood CU behavior predicts behavior problems or CU behavior in late childhood over and above earlier behavior problems and taking into account parent perceptions. That is, use of a MMT framework allows for the comparison of the predictive effects of variance in informant perceptions (“method”) versus effects of the unique variance in the underlying CU behavior and behavior problems constructs in early childhood (“trait”) (Figure 1). Specifically, we sought to examine the effects of a “method” factor capturing informant type (hereafter referred to as the “informant” factor). Latent informant factors capture variance in the ratings provided by two or more informants for all items across the measures of both CU behavior and general behavior problems. The predictive effect of variance in informant perceptions can then be compared to “trait” factors derived as latent factors capturing variance within all CU behavior items, but across both informant, and within all behavior problems items, but across informant (see Figure 1). This modeling approach facilitates an important and testable prediction: If the predictive capability reported for CU behavior measures is driven by negative or positive informant perceptions that have not typically been accounted for in studies, then an “informant” factor, but not a CU behavior “trait” factor, would uniquely predict later rule-breaking, aggressive, and CU behaviors.

### Does age of prediction matter?

Beyond questions centered on validity, a final question relates to the age at which measures of CU behavior are helpful as predictors of later outcomes. Across recent studies in early childhood including several separate samples, studies appear to have settled on testing the predictive validity of preschool CU behavior from *age 3* without formal examination of potential age effects (e.g., Hyde et al., 2013; Waller, Hyde et al., 2015; Willoughby et al., 2014). We have previously hypothesized that it may be developmentally inappropriate to assess CU behavior as young as age 2 because individual differences in the socioemotional capabilities that CU behavior indexes (i.e., empathy, prosociality) are only just emerging (Eisenberg & Fabes, 1990; Fehr, Bernhard, & Rockenbach, 2008). Moreover, a previous study in this sample found the internal validity of DC behavior to be lower at age 2 compared to ages 3 and 4 suggesting that at age 2, items were not tapping a coherent construct with developmental validity (i.e.,  $\alpha=.57$  at age 2 vs.  $\alpha=.64$  at age 3; Hyde et al., 2013). However, no studies have set out to systematically compare the differential predictive

validity of CU behavior assessed in the toddler period (i.e., age 2) versus CU behavior in the early preschool years (i.e., ages 3 or 4), which is important for improving our knowledge about when we can reliably and validly assess the underlying construct of CU behavior in very young children and for knowing *when* to target or tailor early-starting prevention efforts and. Therefore in a final study goal, we sought to evaluate whether there was differential prediction of behavior problems and CU behaviors in late childhood by DC behavior assessed at different ages during early childhood, contrasting prediction by DC behavior at ages 2 versus 3 and 4.

## Current study

Our overarching study goal was thus to examine both the *predictive* and *construct validity* of DC behavior at ages 2–4. We hypothesized that DC behavior would predict aggressive and rule-breaking behavior across informants (teacher and primary and alternative caregiver reports), over and above a problem behavior factor and controlling for “informant” factors indexing method effects within a MMT framework (Figure 1). Second, we hypothesized that there would be unique associations between DC behavior in early childhood and a “general” CU behavior factor at age 9.5, over and above earlier behavior problems (Figure 2), and controlling for concurrent externalizing behavior and informant method effects. Finally, we hypothesized that CU behavior would show predictive and construct validity at ages 3 and 4 but not at age 2, because CU behaviors are unlikely to be fully formed at age 2 and thus the measurement of these behaviors would be less predictive of later outcomes.

## Methods

### Participants

Participants included 731 mother–child dyads recruited between 2002 and 2003 from Women, Infants, and Children (WIC) Nutritional Supplement Program programs in the metropolitan areas of Pittsburgh, PA, and Eugene, OR, and in and outside of Charlottesville, VA (Dishion et al., 2008). Participants were recruited to be part of a randomized controlled trial of the Family Check-Up, a preventive intervention for use in high-risk environments to address normative challenges facing parents from toddlerhood onwards (Dishion et al., 2008). Families were invited to participate if they had a son or daughter between age 2 years 0 months and 2 years 11 months. Recruitment criteria were defined as 1 *SD* above normative means or established clinical cut points on screening measures in at least two of the following three domains: (a) child behavior problems (conduct or high-conflict relationships with adults), (b) primary caregiver problems (maternal depression, daily parenting stress, self-reported substance use), and (3) sociodemographic risk (low education or low family income) (Dishion et al., 2008). Thus, children in the study were selected as “high risk” based on established risk factors for behavior problems. At the same time, because the sample was a community (versus clinic) sample, and not all children met inclusion criteria based on clinically-meaningful frequencies of conduct problems, the sample contained wide variability in early disruptive behavior.

Of the 1,666 families with children of the appropriate age and who were contacted across study sites, 879 met the eligibility requirements (52% in Pittsburgh, 57% in Eugene, and



49% in Charlottesville), and 731 (83.2%) consented to participate. Children in the sample had a mean age of 29.9 months ( $SD=3.2$ ) at the age 2 assessment (2.5 years old). Across sites, primary caregivers self-identified as belonging to the following racial groups: 28% African American, 50% European American, 13% biracial, and 9% other groups. Thirteen percent also self-identified as belonging to a Hispanic ethnic group. During screening, more than 66% of enrolled families had an annual income < \$20,000. Forty-one percent of the sample had a high school/general education diploma, and 32% had 1–2 years of post-high school training. Following baseline assessments, half the sample was randomly assigned to receive the Family Check-Up intervention (Dishion et al., 2008). Although intervention effects have been examined by previous studies in this sample (Dishion et al., 2008; Hyde et al., 2013), our focus was on basic developmental processes. Thus intervention effectiveness was not a major focus of the current study and was only included as a covariate in all analyses.

### Procedures

Annual assessments were conducted from ages 2–10.5 (with the exception of age 6) at the family home. Assessments lasted approximately 2–3 hours and included questionnaires, interviews, assessor impressions, and videotaped observations. Beginning when children were age 7.5, we also collected data from teachers. The current study used questionnaire data collected separately from primary and alternative caregivers at ages 2–4 and 9.5 and from teachers at age 9.5. At age 2, primary caregivers were typically biological mothers (96%) and alternative caregivers were most commonly biological fathers (29%), grandmothers (13%), aunts (3%), or a female friend of the mother (2%). This informant pattern was similar at ages 3 and 4. At age 9.5, primary caregivers were still most likely biological mothers (73%). Alternative caregivers were biological fathers (24%), a stepfather or male partner of mothers (16%), grandmas (6%), or aunts (3%). Families (i.e., primary caregivers) were reimbursed \$100 for participation at age 2, \$120 at age 3, \$140 at age 4, and \$200 at age 9.5. Alternative caregivers were reimbursed \$20 for participation at ages 2–4 and \$40 at age 9.5.

Of 731 families who initially participated, 659 (90%) participated at age 3 and 620 (85%) at age 4. At ages 3 and 4, selective attrition analyses revealed no significant differences in project site, or children's behavior problems, race, ethnicity, and gender (Dishion et al., 2008). At age 9.5, primary caregiver-reported data were available for 586 (80%), alternative caregiver reports for 426 (58%), and teacher reports for 385 (53%). Selective attrition analyses using Chi Square- and t-tests suggested no significant differences in families for whom data were available by intervention status, family income, children's race or gender, or initial level of conduct problems ( $ps > .20$ ). However, parents of children for whom we did not have data available at age 9.5 had significantly lower education ( $p < .001$ ). Thus parental education is a covariate in all models.

## Measures

### Demographics questionnaire – covariates

Primary caregivers completed a demographics questionnaire at age 2 (Dishion et al., 2008). Consistent with past studies in this sample, primary caregiver education was coded as “less than high school” = 0 and “high school/beyond” = 1. Gross annual family income was coded as \$14,999 = 0 and \$15,000 = 1. Child gender was coded as female = 0 and male = 1. Child’s race was coded as “Caucasian/other” = 0 and “Black African-American/biracial” = 1. Ethnicity was coded as “non-Hispanic” = 0 and “Hispanic” = 1. Finally, as data were collected from multiple sites, which differed with respect to the urbanicity and ethnic/racial composition of participants, location was included as a covariate to account for potential differences. Cut-points were designed to represent meaningful differences between groups within our relatively high-risk sample. For example, the cut-point of \$14,999 a year includes families who were 25% or more below the poverty line for a family of four in the year the data was collected. Although reported results use these cut-points to be consistent with previous work in this sample, the pattern of findings was unchanged when we included the full scale quasi-continuous variables of parent education and family income.

### Early childhood measures

**DC behavior (primary & alternative caregiver-reported; ages, 2, 3, & 4)**—We assessed CU behavior in early childhood using a brief adapted measure of DC behavior derived and validated in a previous study (Hyde et al., 2013). The measure includes items from the Child Behavior Checklist (CBCL) (Achenbach & Rescorla, 2000), Eyberg Child Behavior Inventory (ECBI) (Robinson, Eyberg, & Ross, 1980) and Adult-Child Relationship Scale (Pianta, 2001). Items were chosen if they reflected lack of guilt, lack of affective behavior, deceitfulness, or were similar to ICU items (Frick, 2004). In our earlier study, we constructed this factor using exploratory factor analysis on half the sample and confirmatory factor analysis on the other half. Five items loaded onto a factor, which was termed DC behavior to reflect the item content containing items indexing callousness and deceitful behavior: “child doesn’t seem guilty after misbehaving” (CBCL, 0–2 scale; 0 = not true, 1 = somewhat true; 2 = very true), “punishment doesn’t change behavior” (CBCL), “child is selfish/won’t share” (CBCL), “child lies” (Eyberg, 1–7 scale; 1 = never; 4 = seldom; 7 = always), and “child is sneaky/tries to get around me” (ACRS, 1–5 scale; 1 = definitely note; 3 = not sure; 5 = definitely) (Hyde et al., 2013). The measure demonstrated poor internal consistency at age 2 ( $\alpha = .57$ ), which improved at ages 3 ( $\alpha = .64$ ) and 4 ( $\alpha = .73$ ). Internal consistencies are comparable with CU behavior measures in older samples of children (Frick et al., 2014) and in other preschool samples (Willoughby et al., 2011).

**Behavior problems (primary & alternative caregiver-reported; ages, 2, 3, & 4)**—Behavior problems were assessed at ages 2–4 via the ECBI (Robinson et al., 1980), a 36-item behavior checklist. The ECBI assesses behavior problems in children between 2–16 years of age via two factors, one that focuses on the perceived *intensity* of behaviors, and another that identifies the degree to which the behavior is a *problem* for caregivers. The current study used the intensity factor<sup>1</sup>, which is an index of the *frequencies* of early conduct problems (versus parents’ perceptions of the behavior being problematic). Primary and



alternative caregivers rated items on a seven-point Likert scale (e.g., 1 = never; 4 = sometimes; 7 = always), providing an index of the intensity of problem behaviors (e.g., “temper tantrums”, “physically fights”, and “destroys toys and other objects”). One item that was used in the DC behavior measure (“lies”) was removed to avoid content overlap. There was acceptable internal consistency of both scales from ages 2 to 4 ( $\alpha = .86 - .94$ ) (Dishion et al., 2008). We specifically chose to use the ECBI since it overlapped less with the DC behavior factor in item content than the CBCL and because it was the measure that was originally used to screen children into this study.

### Late childhood behavior problems (age 9.5)

**Aggressive and rule-breaking behavior (primary caregiver, alternative caregiver, & teacher reports; age 9.5)**—Primary and alternative caregivers completed the CBCL (Achenbach & Rescorla, 2000) and teachers completed the Teacher Report Form of the Child Behavior Checklist (TRF) (Achenbach, 1991). Both questionnaires include an Externalizing (33 items for the CBCL and 34 for the TRF) behavior scale, comprising two subscales of aggressive behavior (e.g., defiant, talks back) and rule-breaking (e.g., steals). Separate models were examined for aggressive versus rule-breaking subscales. Thus, we could test whether CU behavior versus behavior problems trait factors showed differential associations with these overlapping but separable components of the broader externalizing construct. However, for our third study goal of examining specificity in the prediction of ICU scores, we included the broadband externalizing scale (rather than separate aggressive and rule-breaking subscales) in cross-lagged models to test potential unique links between early versus late childhood CU behavior. High internal consistencies were found for primary caregiver, alternative caregiver, and teacher reports of externalizing behavior at age 9.5 (range,  $\alpha = .93 - .95$ ).

**Callous unemotional (CU) behavior (primary caregiver & alternative caregiver; age 9.5)**—We assessed CU behavior in late childhood via primary and alternative caregiver reports on the ICU (Frick, 2004) (e.g., “does not show emotions”; “feels guilty after wrongdoing”). Items are rated on a 4-point scale (0 = not true; 1 = somewhat true; 2 = very true; 3 = definitely true). We used a “general” CU behavior and “specific” uncaring, callous, and unemotional factors at age 9.5 for both primary and alternative caregiver reports, based on a 3FBF structure that was validated in a previous study in this sample (Waller et al., 2015). Unfortunately, we did not collect teacher- or youth-reported versions of the ICU. High internal consistencies were found for total ICU “general” factor ( $\alpha = .87$ ) and “specific” callous ( $\alpha = .78$ ) and uncaring ( $\alpha = .81$ ) scores, and acceptable internal consistency for the “specific” unemotional score ( $\alpha = .65$ ).

### Analytic strategy

All models were computed using Weighted Least Squares Means Variance estimation (WLSMV) in Mplus 7.2 (Muthén & Muthén, 2014) to take into account the ordinal nature of

---

<sup>1</sup>Note that when we re-computed models using the Eyberg *Problem* factor, which identifies a count of how many problem behaviors are actually considered problematic by the parent, instead of the Intensity factor, the pattern of findings was similar but the effect of DC behavior on late childhood outcomes was *stronger*. Thus, we present findings using the Intensity factor as a more conservative test of the predictive validity of DC behavior.

items (Flora & Curran, 2004). WLSMV estimation accounts for missing data in four steps that include two steps using maximum likelihood estimation, which has been shown to be more efficient than listwise deletion and produces unbiased results with up to 50% missing at random (Enders & Bandalos, 2001). We re-computed analyses including intervention group, child gender, ethnicity, and race, primary caregiver education and age, and family income to check whether accounting for the effects of these demographic covariates influenced findings. Models included all participants except if they were missing on individual covariates ( $n = 723 - 731$ ) or if they were missing all items for the entire behavior problems or DC behavior measures ( $n = 679 - 713$ ).

### **Aim 1: Does DC behavior at ages 2–4 uniquely predict behavior problems at age 9.5?**

We computed zero-order correlations between DC behavior and behavior problems scores at ages 2–4 and aggressive and rule-breaking behavior at age 9.5. We next examined MMMT models, specifying items from across primary and alternative caregiver reports of DC behavior (5 items each; 10 in total) and behavior problem (35 items each; 70 in total) to load onto two “trait” factors, while simultaneously specifying all primary caregiver-reported (i.e., across DC behavior and behavior problem items) and alternative caregiver-reported items (i.e., across DC behavior and behavior problem items) to load onto two separate “informant” factors. We tested whether trait versus informant factors uniquely predicted aggressive or rule-breaking behavior across primary caregiver, alternative caregiver, and teacher reports at 9.5 (Figure 1).

### **Aim 2: Does early childhood DC behavior uniquely predict late childhood CU behavior?**

We computed zero-order correlations between DC behavior at ages 2–4 and extracted ICU “general” CU behavior and three “specific” factor scores at age 9.5 modeled using the bifactor framework validated in an earlier study<sup>2</sup>. We next used multiple regression models to examine whether there were unique links between DC behavior and ICU extracted bifactor scores, controlling for earlier behavior problems. We examined separate within-informant (primary vs. alternative caregiver reports) and across-informant models (Figure 2). In a final test of the convergence between DC behavior and ICU factor scores, we re-ran MMMT models, and included extracted “general” CU behavior factor scores and concurrent externalizing behavior at age 9.5 within a cross-lagged framework. Thus, we accounted for behavior problems severity in early and late childhood and tested whether DC behavior still uniquely predicted later CU behavior.

### **Aim 3: Does age of prediction matter?**

Across all analyses, we examined separate models for predictor variables assessed at ages 2, 3, or 4 to test developmental specificity. We present results for ages 2 and 3 in tables. For brevity, we only discuss in the text any differences in the results for ages 2 versus 3 to address the goal of examining developmental specificity. We only present and discuss results for models testing ages 2 and 3 predictors, as results were highly similar for DC behavior at ages 3 versus 4.

<sup>2</sup>For computational and interpretation ease we used *extracted* factor scores based on modeling in Waller et al. (2015) because models including both a bifactor and MMMT approach contained too many parameters to be calculated in a sample of this size.

## Results

### Aim 1: Does DC behavior at ages 2–4 uniquely predict behavior problems at age 9.5?

Descriptive statistics for study variables are presented in Table 1. There were modest-moderate zero-order correlations between DC behavior at age 3 and primary and alternative caregiver reports of rule-breaking and aggressive behavior at age 9.5 within and across informant (range,  $r = .16 - .31$ ,  $p < .01$ ) (Table 2). These correlations suggested that DC behavior was related to behavior problems in late childhood. However, zero-order correlations between DC behavior at age 3 and teacher-reported outcomes at age 9.5 were not significant. Next, we examined MMT models that included both trait (DC behavior vs. behavior problem) and method (primary vs. alternative caregiver reports) factors as predictors of outcomes at age 9.5 across informants and settings (Tables 3 & 4; Figure 1). At age 3, we found that the DC behavior *trait* factor uniquely predicted primary and alternative caregiver-reported rule-breaking and aggression, controlling for problem behavior and method factors (Table 4)<sup>3</sup>. However, we also found robust effects of age 3 *method factors* on outcomes both within and across informant. Specifically, within informant, primary and alternative caregiver method factors uniquely predicted both aggression and rule-breaking at 9.5. There were no significant effects of age 3 trait or method factors on teacher-reported outcomes at 9.5, with the exception of the alternative caregiver method factor predicting *lower* teacher-reported rule-breaking. However, given that the zero order correlations between both caregivers reports and later teacher reports approached zero, the multivariate negative prediction is likely spurious.

### Aim 2: Does early childhood DC behavior *uniquely* predict late childhood CU behavior?

There were modest-moderate zero-order correlations between DC behavior at age 3 and “general” CU behavior factor scores at age 9.5 within and across informant (range,  $r = .20 - .24$ ,  $p < .01$ ), suggesting convergence of these measures over time. In multiple regression models, we also examined *unique* associations between early childhood DC behavior and “general” versus “specific” ICU factors, controlling for earlier behavior problems (Figure 2). At age 3, both primary and alternative caregiver reports of DC behavior uniquely predicted “general” CU behavior scores over and above behavior problems (Table 5). Consistent with zero-order associations, primary caregiver reports of behavior problems at ages 2–4 uniquely predicted “specific” callous scores. As a final test of construct validity, we examined convergence between DC behavior measure with CU behavior in late childhood, controlling for earlier and concurrent behavior problems and method factors. This cross-lagged approach and inclusion of method factors meant that we isolated *unique convergence* in links between the DC behavior and CU behavior measures over time and ruled out the possibility that effects in previous models were driven by severity. In support of this hypothesis, the DC behavior factor at age 3 was uniquely related to CU behavior at age 9.5

<sup>3</sup>We ran models controlling for intervention group, project location, child gender, ethnicity and race, family income, and parent education and age. Boys were reported as having higher aggressive and rule-breaking behavior across informants at age 9.5. Children whose parents had below a high school education were reported as showing higher rule-breaking across informants. Teachers reported African-American children as having higher aggressive and rule-breaking behavior. Primary and alternative caregivers reported non-Hispanic children as showing lower aggression.

across primary and alternative caregiver reports, and over and above method factors and both earlier and concurrent behavior problems (Figure 3).

### Aim 3: Does age of prediction matter?

Compared to the zero-order correlations for DC behavior at age 2, zero-order correlations between DC behavior at age 3 and primary and alternative caregiver reports of rule-breaking and aggressive behavior at age 9.5 were larger in magnitude, confirmed quantitatively using Fisher *r*-to-*z* transformations (Table 2; age 4 results similar and not presented for brevity). In MMT models, the DC behavior factor at age 2 predicted primary caregiver-reported aggression and rule-breaking at age 9.5, as well as alternative caregiver-reported aggression (trend-level), over and above the behavior problem trait factor and method factors. However, the most consistent predictive effects at age 2 were for the *method* factors, with particularly robust within-informant effects. Moreover, at age 2 (but not age 3), both primary and alternative caregiver *method* factors also predicted aggression and rule-breaking across informant. Further, the age 2 primary caregiver method factor predicted teacher-reported aggression. Finally, the magnitude of the association between DC behavior and behavior problems at age 2 with “general” CU behavior factor scores at age 9.5 was similar. In contrast, by age 3, DC behavior was *uniquely* related to later CU behavior, with non-significant associations between age 3 behavior problems and later CU behavior (Table 3). Thus, at ages 3 and 4, the DC behavior factor showed consistent unique prediction of later aggression and rule-breaking, and convergence with a fuller measure of CU behavior. In contrast, at age 2, significant effects were distributed across the DC behavior and behavior problem trait factors, and were most consistent for *method* factors, suggesting lower validity in the underlying DC behavior *construct* at this earlier age (Table 5).

## Discussion

The current study examined whether an early childhood measure of CU behavior (DC behavior) predicted rule-breaking, aggressive, and CU behaviors in late childhood. First, we demonstrated unique effects of DC behavior at ages 3 and 4 on aggression and rule-breaking at age 9.5 within and across primary and alternative caregiver reports, parsing overlap with early behavior problems, and within a MMT framework that accounted for informant method effects. Second, we found that DC behavior uniquely predicted “general” CU behavior at age 9.5 assessed using a gold-standard CU behavior measure. Moreover, cross-lagged models demonstrated unique convergence within and across informants in associations between early childhood DC behavior and late childhood CU behavior, controlling for earlier and concurrent behavior problems, as well as informant factors. Early behavior problems, but not DC behavior, uniquely predicted “specific” callous scores, after accounting for variance in “general” CU behavior. Finally, we found evidence to suggest that the predictive and construct validity of DC behavior is stronger from age 3 onwards. We focus our discussion on each of these findings.

### DC behavior at ages 2–4 uniquely predicts behavior problems at age 9.5

There was robust prediction of aggression and rule-breaking by DC behavior at ages 3 and 4, consistent with recent studies that have demonstrated that early childhood CU behavior

uniquely predicts future behavior problems (Waller, Hyde et al., 2015; Willoughby et al., 2014). However, we did not find evidence to support differential prediction of rule-breaking versus aggression. Rather, our DC behavior measure predicted higher levels of both, suggesting that this measure may be useful in identifying children at greatest risk of severe behavior problems across dimensions. We interpret this finding within a long-established classification that focuses on both overt and covert dimensions within youth antisocial behavior (e.g., Loeber & Schmaling, 1985). Of relevance, a “versatile/mixed” antisocial subgroup with aggressive (e.g., fighting) and rule-breaking (e.g., theft) problems are thought to pose the *greatest* risk for more serious forms of violence and aggression into adolescence (see Dishion, 2014 for a review). Findings from the current study suggest that DC behavior may tap a construct over and above the early emergence of covert, rule breaking behaviors alone that fits with conceptualizations of a “versatile/mixed” antisocial group<sup>4</sup>.

A novel aspect of this study was an examination of models controlling for *informant method* factors, guided by a large literature highlighting the importance of method variance (Campbell & Fiske, 1959; Podsakoff et al., 2003). MMT models revealed significant effects of early childhood DC behavior on later outcomes over and above the effects of “informant” factors, suggesting that the unique predictive validity of our DC behavior measure cannot solely be accounted for by rater effects. However, we also found that the informant factors across ages 2–4 predicted aggressive and rule-breaking behavior over and above the effects of DC behavior and behavior problem “trait” factors. Strongest associations were within informant (e.g., primary caregiver informant factor predicting primary caregiver-reported outcomes), highlighting that potential rater biases could account for many of the within-informant associations reported in previous studies that have used single informant ratings to assess whether CU behavior is related to future behavior problems (Frick et al., 2014).

Interestingly however, we also found some *cross*-informant prediction by method factors. Thus, there was evidence that informant factors actually predicted *someone else’s* ratings of the child’s behavior up to seven years later (i.e., associations were not inflated by within-rater bias). One possibility is that the cross-informant prediction for primary and alternative caregivers could have been driven, at least in part, by shared perceptions of the child, including perceptions resulting from similarity in ratings between two caregivers who have previously discussed the child’s behavior (i.e., spouses/co-parents/parent or grandparent). This explanation does not, however, address method factors predicting *teacher*-reported outcomes (though parent and teacher perceptions could affect each other during parent-teacher communications). Indeed, it is noteworthy that, despite robust prediction of aggression and rule-breaking within and across primary and alternative caregiver report, the DC behavior factor at ages 2–4 did not consistently predict teacher-reported outcomes. One explanation is that the lack of agreement across parent and teacher reports reflects “true”

---

<sup>4</sup>Note that we re-computed models examining links between trait (DC behavior and behavior problems) and informant (primary and alternative caregivers) factors and later aggression, rule-breaking and CU behavior in just the control group ( $N = 364$ ). The pattern of findings was broadly similar, suggesting unique effects of DC behavior on outcomes from age 3 onwards, and significant effects of informant effects particularly for within informant. We found some evidence to suggest differential prediction of later rule-breaking by earlier DC behavior versus prediction of later aggression by earlier behavior problems; otherwise the pattern of findings was almost identical in this subsample (results available from study authors on request).

differences in the behavior of children in home versus school based on different expectations and contingences across these settings. Thus, the predictive validity of CU behavior as rated by a parent may be limited to the home setting. Future studies are needed to examine corroboration across informants from home versus school settings, and particularly whether discrepancies in agreement differentially predict long-term outcomes (see De Los Reyes, Thomas, Goodman, & Kundery, 2013). Further, while teachers were required to have known students for a minimum of two months to complete questionnaires, the inconsistent associations found may reflect lower knowledge that teachers had of children based on fewer observations of the child. Future studies could therefore address this issue by scheduling school data collection for the end of the school year to maximize the time that teachers have had to get to know children.

Beyond the lack of significant associations between the DC behavior *trait/construct* factor and teacher-reported outcomes, we did find, somewhat surprisingly, that the primary caregiver *method/informant* factor at age 2 predicted both teacher-reported aggression and rule-breaking. Thus, it may be that at this very early age, parents see something “negative” in children’s early behavior that we are not measuring well, but that has important predictive validity. For example, only a handful of items in the scale used to assess behavior problems addressed inattention (e.g., “he/she has a short attention span”) or hyperactivity (“he/she is overactive or restless”). Therefore, having parsed variance specific to the DC behavior and behavior problems factors, the remaining variance with predictive validity could have been parents picking up on ADHD behaviors. Alternatively, the informant factor may reflect other processes relating to the parent, including harsh parenting or depressive symptoms, which exacerbate risk for behavior problems or CU behavior via coercive parent-child interactions or negative developmental processes resulting in later (teacher-reported) externalizing behaviors (e.g., Dishion, 2014). The informant factor could also reflect other types of learned behaviors not captured by the items in these analyses (i.e., mistrust or resentment of a specific adult), a heritable trait, or a parental projection onto or belief about the child that is not reflected in the child’s behavior. These “unobserved” variables could further increase risk for children to show behavior problems. Future studies that explicitly examine whether informant perceptions influence children’s outcomes could address these issues either by including observations of parenting behaviors, specific parental characteristics, the inclusion of items in models that assess broader dimensions of child behavior (e.g., internalizing; prosociality), multi-wave models examining reciprocal effects between informant and trait factors, or genetically-informed designs to test various possibilities explaining method factor predictive effects from very early childhood.

### **Early childhood DC behavior *uniquely* predicts late childhood CU behavior**

Our measure of DC behavior uniquely predicted CU behavior “general” factor scores at age 9.5 modeled within a bifactor framework while also controlling for earlier behavior problems. The convergence of early childhood DC behavior with “general” CU behavior scores in late childhood is a useful test of construct validity, as both our DC behavior measure and the ICU tap variance relating to a lack of empathy and deficits in guilt, albeit via different items and at different ages. At the same time, we caution the use of the somewhat artificial statistical modeling approach of our analyses. While we demonstrated



links between two CU behavior measures in early and late childhood, we do not intend to imply stability in the CU behavior construct at either the mean or individual levels. Rather, these analyses show that rank order in early childhood is predictive of rank order of CU behavior in middle childhood and cannot address developmental changes in mean levels over time. Indeed, a growing body of literature highlights that CU behavior is far from “trait” like, immutable, or any more stable than behavior problems, showing heterogeneity in terms of trajectories across childhood (Fontaine, McCrory, Boivin, Moffitt, & Viding, 2011) and appearing malleable in response to parenting practices and other sources of environmental influence (Waller et al., 2013). Nevertheless, we see results as representing a “proof of concept” of our early childhood measure of DC behavior, which used items from behavior checklists that were not originally designed to capture the CU behavior construct. To confirm this conclusion, we examined cross-lagged models that controlled for behavior problems both at ages 3 and 9.5, and that included informant method factors. We found that DC behavior at age 3 continued to uniquely predict CU behavior at age 9.5 within and across informant, meaning that we could discount the possibility that associations between CU behavior measures in early versus late childhood reflected severity or informant perceptions, and supporting the CU behavior construct as having developmental meaning, even at this young age.

These findings are in line with a separate body of evidence in the developmental literature indicating that “CU-like” behavior could meaningfully exist and be measured in preschool children because individual differences in core characteristics related to CU behavior emerge at ages 2–3 years old, including the capacity for empathic concern (Eisenberg & Fabes, 1990), sharing rewards with others (Fehr et al., 2008), and the distinction between “nice” versus “nasty” Theory of Mind (Ronald, Happe, Hughes, & Plomin, 2005). At the same time, it is vital to remain mindful of the potential hazards associated with labeling very young children as “callous and unemotional”, including need for continued evaluation of the developmental appropriateness of items used to assess CU behavior, the assumption that individual differences reflect psychopathology versus developmental delay (or other processes, e.g., autism), and the importance of recognizing changes in personality and temperament features across childhood (Seagrave & Grisso, 2002). Nevertheless, we continue to believe ultimately in the translational potential of this basic research, focusing particularly on the *positive* utility of identifying children who are at high risk of escalating behavior problems based on the presence of CU behavior and/or individual differences in empathic concern, prosociality, or moral regulation. Thus, we emphasize that our conceptualization and measurement of CU behavior may help to identify young children with specific socioemotional and behavioral needs, who are less likely to desist from early-starting conduct problems and may require targeted and personalized treatments.

### **DC behavior is more reliably predictive of outcomes from age 3 onwards**

Finally, in addition to examining links with behavior outcomes within MMMT models, we tested whether the validity of our early childhood measure of CU behavior showed *developmental specificity* by comparing age 2 findings to those for ages 3 and 4. It is noteworthy that the majority of previous studies that have assessed the predictive validity of CU behavior in early childhood have focused on examining prediction from age 3 onwards

(Waller, Hyde et al., 2015; Willoughby et al., 2014). In support of the developmental focus of these studies, we found that the magnitude of zero-order correlations between DC behavior and later behavior problems was greater at ages 3 and 4 versus age 2. Further, while DC behavior at age 2 predicted primary caregiver-reported outcomes, there was no significant cross-informant prediction of outcomes (i.e., reported by alternative caregivers). In contrast, by age 3 DC behavior showed consistent prediction of outcomes within and across informants. Finally, DC behavior at age 3 onwards showed consistent within and across informant links with the ICU, a purpose-developed CU behavior measure. In contrast, there were comparable predictive effect sizes for associations between behavior problems versus DC behavior at age 2 and CU behavior at age 9.5, suggesting that items indexing CU-like behavior are less meaningful and potentially less easily differentiated from early behavior problems at age 2. These findings are in line with our conclusions in a previous study in this sample and are further supported by the lower internal consistency of the DC behavior measure at age 2 (Hyde et al., 2013). Taken in conjunction with findings from our first study goal, measures of early CU-like behavior appear to exhibit more reliable predictive and construct validity when assessed from age 3 onwards.

### Strengths and Limitations

There were a number of strengths to the current study. In particular, we examined associations between early CU behavior and later aggressive, rule-breaking, and CU behavior over 5.5-to-7.5 year follow-up periods, incorporated reports of behavior from three different informants, and examined associations within bifactor, MMMT, and cross-lagged frameworks, all within a relatively large, diverse, and high-risk sample. However, findings should be considered alongside several limitations. First, we focused on low-income children with risk factors across multiple domains, including sociodemographic risk, family risk, and early child problem behavior. Thus, it is unclear whether results would generalize to children from higher-income families, families with fewer risk factors, or clinic-referred populations. Second, a more thorough examination of any associations between trait and method factors would involve testing MMMT models in both early (ages 2–4) and late (age 9.5) childhood. However, the number of required parameters made this model too computationally difficult to estimate within this sample. Finally, although there was generally corroboration across models using primary versus alternative caregiver reports, we found inconsistent prediction of teacher-reported outcomes, including one finding opposite to the expected direction. We accounted for informant perceptions/method factors within models testing associations between caregiver-reported outcomes. Thus, these associations are likely robust to the potential confounding effects of shared method variance. Nevertheless, the lack of prediction of teacher-reported outcomes suggests limitations to the concept that children with either early behavior problems or CU behavior can be characterized as having a stable and cross-context “trait”. Future studies are needed to explore whether discrepancies in what parents and teachers report reflect true differences in children’s behavior across settings.

### Conclusions and future directions

Our findings have implications for preventative interventions. First, we demonstrated that from age 3 onwards, DC behavior robustly predicted worse behavior problems in late

childhood. Extrapolating from these findings, early childhood measures of CU behavior may help identify children most in need of intervention, which could enable treatment components to be tailored to fit socioemotional needs related to CU behavior (i.e., lower empathic concern). This conclusion is particularly salient given the recent inclusion of a CU behavior specifier for the diagnosis of child conduct disorder into the DSM-5 (“with limited prosocial emotions”; American Psychiatric Association, 2013), meaning that clinicians are already making diagnostic and treatment decisions based on the presence of child CU behavior. However, the finding that aggression and rule-breaking were also consistently predicted by informant factors highlights that variance relating to ideas or beliefs of the informant (i.e., parent) can, in many cases, add as much variance to the prediction of outcomes as the supposed, underlying “trait”. Thus, in relation to intervention implications, it is vital to consider the fact that parental characteristics, attitudes, caregiving practices, and the broader family ecology continue to represent key mechanisms for identifying at-risk, vulnerable children and those families who may be most in need of early intervention (Dishion et al., 2008; Shaw & Shelleby, 2014). Second, we demonstrated that our “brief-adapted”, five-item DC behavior measure in early childhood converged with CU behavior assessed in late childhood via a “gold-standard” measure. In line with other recent studies, this finding supports the notion that, at least from age 3 onwards, CU behavior-like items on common behavior questionnaires appear quantitatively distinct from other dimensions within childhood behavior problems. Finally, beyond the fact that we examined convergence of two CU behavior measures over time and that MMT models included CU behavior “trait” factors, we emphasize that our conceptualization of CU behavior represents simply one way to identify children who may be at most risk of poor outcomes, and who would benefit from empirically-supported and tailored preventative interventions. Thus, we do not imply that early CU behavior can or should be conflated with CU “traits” (e.g., unchangeable, highly stable), nor that CU behavior should be equated with psychopathic traits or psychopathy in adulthood, rather that CU behavior can be considered a risk factor for these later indicators of severe trajectories of antisocial behavior or aggression. In sum, we believe that our findings highlight potentially new ways to identify children by age 3 who are less likely to naturally desist from early behavior problems and may have different treatment needs across childhood.

## Acknowledgments

This research was supported by Grants 5R01 DA16110 and 5R01 DA16110-02 from the National Institutes of Health, awarded to Thomas J. Dishion, Daniel S. Shaw, Melvin N. Wilson, and Frances Gardner. We thank families and staff of the Early Steps Multisite Study. Some findings reported in full here were presented at the Society for Research in Child Development (SRCD) in Philadelphia, PA, USA in March 2015.

## Abbreviations

<b>CU</b>	callous-unemotional
<b>DC</b>	deceitful-callous
<b>MMMT</b>	multi-method multi-trait

## References

- Achenbach, TM. Manual for the Teacher's Report Form and 1991 profile. Burlington, VT: University of Vermont Department of Psychiatry; 1991.
- Achenbach, TM.; Rescorla, L. ASEBA preschool forms & profiles: An integrated system of multi-informant assessment. Burlington, VT: University of Vermont Department of Psychiatry; 2000.
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5. Washington, DC: American Psychiatric Association; 2013.
- Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*. 1959; 56:81–105. DOI: 10.1037/h0046016 [PubMed: 13634291]
- Campbell SB. Behavior problems in preschool children: a review of recent research. *Journal of Child Psychology and Psychiatry*. 1995; 36:113–149. DOI: 10.1111/j.1469-7610.1995.tb01657.x [PubMed: 7714027]
- Côté S, Vaillancourt T, LeBlanc JC, Nagin DS, Tremblay RE. The development of physical aggression from toddlerhood to pre-adolescence: A nation wide longitudinal study of Canadian children. *Journal of Abnormal Child Psychology*. 2006; 34:68–82. DOI: 10.1007/s10802-005-9001-z
- Dadds MR, Fraser J, Frost A, Hawes DJ. Disentangling the underlying dimensions of psychopathy and conduct problems in childhood: a community study. *Journal of Consulting and Clinical Psychology*. 2005; 73:400–410. DOI: 10.1037/0022-006X.73.3.400 [PubMed: 15982138]
- De Los Reyes A, Thomas SA, Goodman KL, Kundery SM. Principles underlying the use of multiple informants' reports. *Annual Review of Clinical Psychology*. 2013; 9:123–149. DOI: 10.1146/annurev-clinpsy-050212-185617
- Dishion, T. A Developmental Model of Aggression and Violence: Microsocial and Macrosocial Dynamics Within an Ecological Framework. In: Lewis, M.; Rudolph, KD., editors. *Handbook of Developmental Psychopathology*. Springer; US: 2014. p. 449-465.
- Dishion TJ, Patterson GR. Age effects in parent training outcome. *Behavior Therapy*. 1992; 23:719–729. DOI: 10.1016/S0005-7894(05)80231-X
- Dishion TJ, Shaw D, Connell A, Gardner F, Weaver C, Wilson M. The Family Check-Up with high-risk indigent families: Preventing problem behavior by increasing parents' positive behavior support in early childhood. *Child Development*. 2008; 79:1395–1414. DOI: 10.1111/j.1467-8624.2008.01195.x [PubMed: 18826532]
- Eisenberg N, Fabes RA. Empathy: Conceptualization, measurement, and relation to prosocial behavior. *Motivation and Emotion*. 1990; 14:131–149. DOI: 10.1007/BF00991640
- Enders CK, Bandalos DL. The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*. 2001; 8:430–57. DOI: 10.1207/S15328007SEM0803\_5
- Essau CA, Sasagawa S, Frick PJ. Callous-unemotional traits in a community sample of adolescents. *Assessment*. 2006; 13:454–469. DOI: 10.1177/1073191106287354 [PubMed: 17050915]
- Fehr E, Bernhard H, Rockenbach B. Egalitarianism in young children. *Nature*. 2008; 454:1079–1083. DOI: 10.1038/nature07155 [PubMed: 18756249]
- Flora DB, Curran PJ. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*. 2004; 9:466–491. DOI: 10.1037/1082-989X.9.4.466 [PubMed: 15598100]
- Fontaine NM, McCrory EJ, Boivin M, Moffitt TE, Viding E. Predictors and outcomes of joint trajectories of callous–unemotional traits and conduct problems in childhood. *Journal of Abnormal Psychology*. 2011; 120:730. doi: 10.1037/a0022620 [PubMed: 21341879]
- Frick PJ. The Inventory of Callous-Unemotional Traits. Unpublished rating scale. 2004
- Frick PJ, Morris AS. Temperament and developmental pathways to conduct problems. *Journal of Clinical Child and Adolescent Psychology*. 2004; 33:54–68. DOI: 10.1207/S15374424JCCP3301\_6 [PubMed: 15028541]
- Frick PJ, Ray JV, Thornton LC, Kahn RE. Annual Research Review: A developmental psychopathology approach to understanding callous unemotional traits in children and adolescents with serious conduct problems. *Journal of Child Psychology and Psychiatry*. 2014; 55:532–548. DOI: 10.1111/jcpp.12152 [PubMed: 24117854]

- Hyde LW, Shaw DS, Gardner F, Cheong J, Dishion TJ, Wilson M. Dimensions of callousness in early childhood: Links to problem behavior and family intervention effectiveness. *Development and Psychopathology*. 2013; 25:347–363. DOI: 10.1017/S0954579412001101 [PubMed: 23627949]
- Kimonis ER, Frick PJ, Boris NW, Smyke AT, Cornell AH, Farrell JM, Zeanah CH. Callous-unemotional features, behavioral inhibition, and parenting: independent predictors of aggression in a high-risk preschool sample. *Journal of Child and Family Studies*. 2006; 15:741–752. DOI: 10.1007/s10826-006-9047-8
- Loeber R, Schmalzing KB. Empirical evidence for overt and covert patterns of antisocial conduct problems: A meta-analysis. *Journal of Abnormal Child Psychology*. 1985; 13:337–353. DOI: 10.1007/BF00910652 [PubMed: 4008760]
- Marsh, HW.; Grayson, D. Latent variable models of multitrait-multimethod data. Thousand Oaks, CA, US: Sage Publications, Inc; 1995.
- Muthén, LK.; Muthén, BO. Mplus user's guide. 7. 2014. (1998–2014)
- Obrodovi J, Pardini DA, Long JD, Loeber R. Measuring interpersonal callousness in boys from childhood to adolescence: An examination of longitudinal invariance and temporal stability. *Journal of Clinical Child and Adolescent Psychology*. 2007; 36:276–292. DOI: 10.1080/15374410701441633 [PubMed: 17658974]
- Pianta, RC. STRS: Student-teacher Relationship Scale: professional manual. Psychological Assessment Resources; 2001.
- Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*. 2003; 88:879–903. DOI: 10.1037/0021-9010.88.5.879 [PubMed: 14516251]
- Reid MJ, Webster-Stratton C, Baydar N. Halting the development of conduct problems in Head Start children: The effects of parent training. *Journal of Clinical Child and Adolescent Psychology*. 2004; 33:279–291. DOI: 10.1207/s15374424jccp3302\_10 [PubMed: 15136193]
- Robinson EA, Eyberg SM, Ross AW. The standardization of an inventory of child conduct problem behaviors. *Journal of Clinical Child & Adolescent Psychology*. 1980; 9:22–28. DOI: 10.1080/15374418009532938
- Ronald A, Happe F, Hughes C, Plomin R. Nice and nasty theory of mind in preschool children: Nature and nurture. *Social Development*. 2005; 14:664–684. DOI: 10.1111/j.1467-9507.2005.00323.x
- Seagrave D, Grisso T. Adolescent development and the measurement of juvenile psychopathy. *Law and human behavior*. 2002; 26:219–239. DOI: 10.1023/A:1014696110850 [PubMed: 11985299]
- Shaw DS, Gilliom M, Ingoldsby EM, Nagin DS. Trajectories leading to school-age conduct problems. *Developmental Psychology*. 2003; 39:189–200. DOI: 10.1037/0012-1649.39.2.189 [PubMed: 12661881]
- Shaw DS, Shelleby EC. Early-starting conduct problems: Intersection of conduct problems and poverty. *Annual Review of Clinical Psychology*. 2014; 10:503–528. DOI: 10.1146/annurev-clinpsy-032813-153650
- Waller R, Gardner F, Hyde LW. What are the associations between parenting, callous-unemotional traits, and antisocial behavior in youth? A systematic review of evidence. *Clinical Psychology Review*. 2013; 33:593–608. DOI: 10.1016/j.cpr.2013.03.001 [PubMed: 23583974]
- Waller R, Wright AG, Shaw DS, Gardner F, Dishion TJ, Wilson MN, Hyde LW. Factor Structure and Construct Validity of the Parent-Reported Inventory of Callous-Unemotional Traits Among High-Risk 9-Year-Olds. *Assessment*. 2015; 22:561–580. DOI: 10.1177/1073191114556101 [PubMed: 25371449]
- Waller R, Hyde LW, Grabell AS, Alves ML, Olson SL. Differential associations of early callous unemotional, oppositional, and ADHD behaviors: multiple domains within early-starting conduct problems? *Journal of Child Psychology and Psychiatry*. 2015; 56:657–666. DOI: 10.1111/jcpp.12326 [PubMed: 25251938]
- Waller R, Shaw DS, Neiderhiser JM, Ganiban JM, Natsuaki MN, Reiss D, Trentacosta CJ, Leve LD, Hyde LW. Towards an understanding of the role of the environment in the development of early callous behavior. *Journal of Personality*. 2015; in press. doi: 10.1111/jopy.12221
- Webster-Stratton C, Taylor T. Nipping early risk factors in the bud: Preventing substance abuse, delinquency, and violence in adolescence through interventions targeted at young children (0–8

years). *Prevention Science*. 2001; 2:165–192. DOI: 10.1023/A:1011510923900 [PubMed: 11678292]

Willoughby MT, Mills-Koonce RW, Gottfredson NC, Wagner NJ. Measuring callous unemotional behaviors in early childhood: factor structure and the prediction of stable aggression in middle childhood. *Journal of Psychopathology and Behavioral Assessment*. 2014; 36:30–42. DOI: 10.1007/s10862-013-9379-9 [PubMed: 24729655]

Willoughby MT, Waschbusch DA, Moore GA, Propper CB. Using the ASEBA to screen for callous unemotional traits in early childhood: factor structure, temporal stability, and utility. *Journal of Psychopathology and Behavioral Assessment*. 2011; 33:19–30. DOI: 10.1007/s10862-010-9195-4 [PubMed: 21483647]

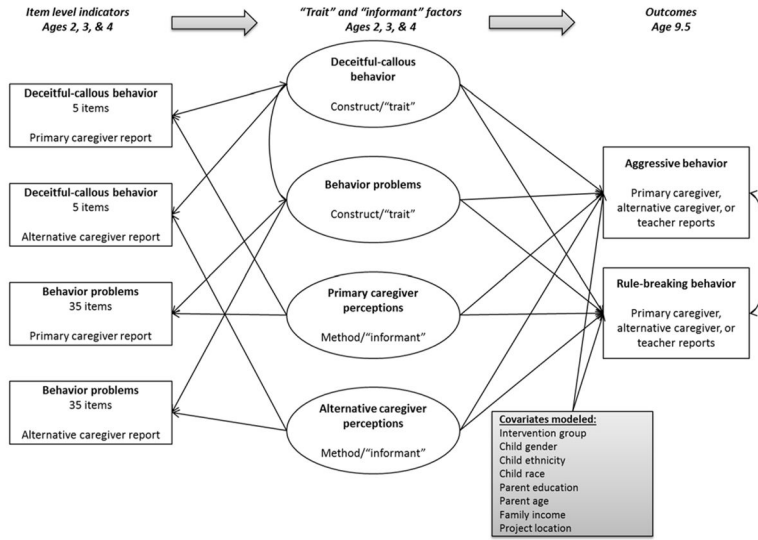
Author Manuscript

Author Manuscript

Author Manuscript

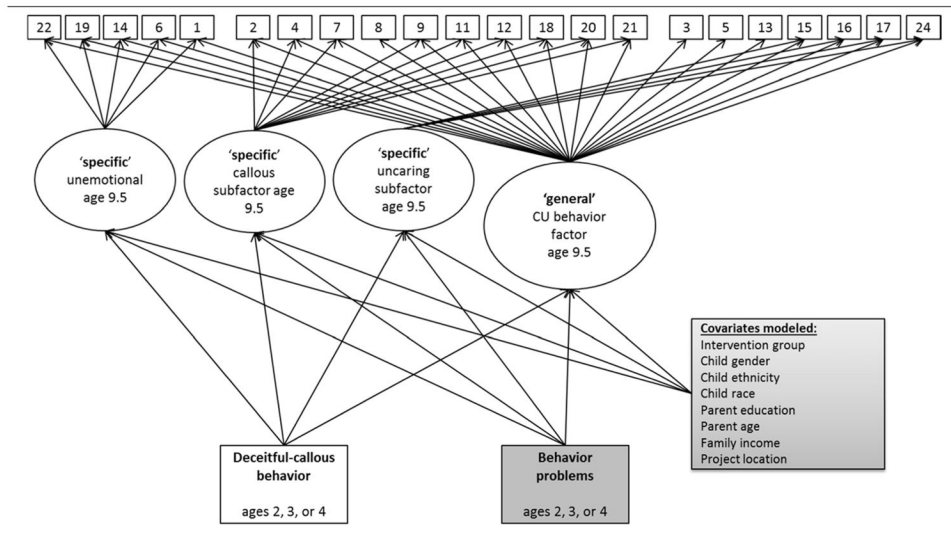
Author Manuscript





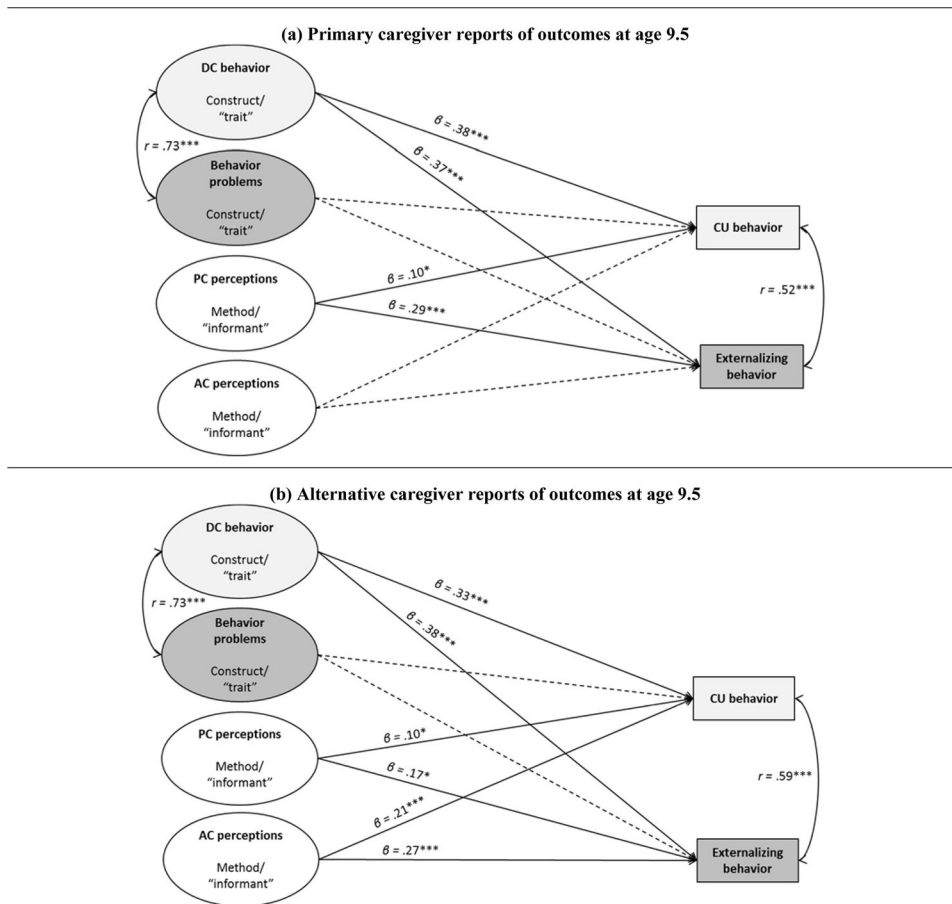
**Figure 1. Multi-trait multi-method model examining whether early CU behavior (deceitful-callousness) and behavior problem “trait” factors at ages 2, 3, and 4 versus primary and alternative caregiver reported “informant” factors at ages 2, 3, and 4 uniquely predict outcomes at age 9.5**

*Note.* Item level indicators not shown (for ease of readability): 5 DC behavior items based on Hyde et al. (2013); 35 behavior problems items (Eyberg Scale; Robinson et al., 1980). Prediction by measures assessed at ages 2, 3 and 4 tested in separate models. Within-time correlation between CU behavior and behavior problems factors specified to account for their overlap. We re-ran models controlling treatment status, project location, child gender, race, and ethnicity, family income, and parent education and age – findings unchanged



**Figure 2. Model showing hypothesized links between deceitful-calling behavior at ages 2, 3, or 4 predicting the three-factor bifactor model of the ICU at age 9.5, controlling for earlier behavior problems**

*Note.* Bifactor model based on Waller et al. (2015). Double headed arrows show within-time correlation between deceitful-calling behavior and behavior problems. Single-headed arrows show regression paths that test unique association between deceitful-calling behavior and behavior problems at ages 2–4 and ICU factors at age 9.5. For computational ease, we estimated effects using extracted bifactor scores but pattern of findings was similar when run within a bifactor framework. We ran separate models for prediction at ages 2, 3, and 4 to examine developmental differences. We re-ran models controlling treatment status, project location, child gender, race, and ethnicity, family income, and parent education and age–findings unchanged.



**Figure 3. Model showing age 3 multi-trait multi-method model within cross-lagged framework predicting primary versus alternative caregiver reports of CU behavior and externalizing behavior at age 9.5**

*Note.* PC = primary caregiver; AC = alternative caregiver; CU = callous unemotional; DC = deceitful-callous. Models tested whether a DC behavior trait, behavior problem trait, or primary and alternative caregiver method factors at age 3 predicted later CU behavior or externalizing behavior, controlling for within-time correlations (cross-lagged framework) across two reports of outcomes at age 9.5 - primary caregiver in (a) and alternative caregiver in (b). We re-ran models controlling for treatment status, project location, child gender, race, and ethnicity, family income, and parent education and age – findings unchanged.

**Table 1**

## Descriptives of study variables

	<i>N</i>	<i>M (SD)</i>	<b>Range</b>
<b>Age 2</b>			
DC behavior (PC)	725	7.72 (2.86)	2–17
DC behavior (AC)	414	6.16 (2.43)	2–15
Behavior problems (PC)	687	127.00 (27.83)	52–233
Behavior problems (AC)	362	113.28 (28.14)	43–213
<b>Age 3</b>			
DC behavior (PC)	649	7.62 (3.13)	2–18
DC behavior (AC)	411	6.73 (3.05)	2–18
Behavior problems (PC)	615	125.05 (31.87)	50–219
Behavior problems (AC)	384	112.60 (29.53)	41–203
<b>Age 9.5</b>			
Aggressive behavior (PC)	586	7.90 (6.79)	0–32
Aggressive behavior (AC)	426	6.99 (6.16)	0–31
Aggressive behavior (T)	385	5.30 (7.49)	0–37
Rule-breaking behavior (PC)	586	2.81 (2.77)	0–20
Rule-breaking behavior (AC)	427	2.51 (2.58)	0–19
Rule-breaking behavior (T)	385	2.26 (2.87)	0–16
CU behavior total score (PC)	533	18.33 (8.83)	0–52
CU behavior total score (AC)	392	19.41 (8.61)	0–49

*Note.* PC = primary caregiver; AC = alternative caregiver; T = teacher; CU = callous-unemotional; DC = deceitful-callous. We report means and standard deviations for observed summed scores for DC behavior and behavior problems at ages 2 and 3 and CU behavior at age 9.5 for ease of interpretation. The DC behavior summed scale comprises scores on five items from three different behavior questionnaires: the CBCL, 0–2 scale (0 = not true, 1 = somewhat true; 2 = very true); the Eyberg, 1–7 scale (1 = never; 4 = seldom; 7 = always); and the ACRS, 1–5 scale (1 = definitely not; 3 = not sure; 5 = definitely). The behavior problems scale comprises 35 items from the Eyberg, 7-point scale (1 = never; 4 = sometimes; 7 = always (see **Methods**). While summary statistics for summed scores are presented in this table, the majority of subsequent analyses modeled latent factors for measures (see Figures 1–3). Rule-breaking and aggressive behavior scores were log-transformed for subsequent analyses to account for negative skew.

Zero-order correlations between primary and alternative caregiver reports of DC behavior and behavior problems scores at ages 2 and 3 and primary caregiver, alternative caregiver, and teacher reports of aggressive and rule-breaking behavior at age 9.5

**Table 2**

	Age 9.5 outcomes across informants					
	Aggressive behavior			Rule-breaking		
	PC	AC	T	PC	AC	T
<b>Age 2</b>						
DC behavior (PC)	.14**	.08*	.06	.16***	.06	.05
DC behavior (AC)	.14**	.14	.03	.13*	.12*	.04
Behavior problems (PC)	.24***	.16***	.05	.21***	.09*	.08 <sup>†</sup>
Behavior problems (AC)	.13**	.13 <sup>†</sup>	.02	.09*	.09	.004
<b>Age 3</b>						
DC behavior (PC)	.31***	.18***	.06	.27***	.16***	.04
DC behavior (AC)	.21***	.17***	.08	.18**	.17***	.05
Behavior problems (PC)	.35***	.13*	.05	.31***	.10*	.05
Behavior problems (AC)	.19***	.21**	.06	.15**	.18**	.02
<b>Age 9.5</b>						
Aggressive behavior (PC)						
Aggressive behavior (AC)	.34***					
Aggressive behavior (T)	.14**	.08				
Rule-breaking behavior (PC)	.74***	.29***	.16***			
Rule-breaking behavior (AC)	.26***	.71***	.09*	.33***		
Rule-breaking behavior (T)	.13*	.08*	.71***	.20***	.11*	

Note.

<sup>†</sup>  $p < .10$ ,

\*  $p < .05$ ,

\*\*  $p < .01$ ,

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

\*\*\*  
 $d < .001$ .

PC = primary caregiver; AC = alternative caregiver; CU = callous unemotional; DC = deceitful-callous. Pattern of findings for age 4 DC behavior and behavior problems with outcomes at age 9.5 were similar to those for age 3. Thus, these findings are not presented for brevity, but available on request. Note that inter-correlations between DC behavior and behavior problems from ages 2–4 have been reported in a previous study (see Hyde et al., 2013). Fisher  $r$ -to- $z$  transformations showed that correlations with aggressive and rule-breaking behavior at age 9.5 were significantly greater in magnitude for age 3 DC behavior than age 2 DC behavior within primary caregiver informant reports (range,  $z = -2.2 - 3.43$ ,  $p < .001$ ) and for primary caregiver predicting alternative caregiver reports (range,  $z = 1.93 - 1.94$ ,  $p < .05$ ).



**Table 3** Multi-trait multi-method models examining age 2 primary and alternative caregiver reports of DC behavior and behavior problems predicting aggressive, rule-breaking, and CU behavior at age 9.5

	Aggression						
	Primary caregiver reports		Alternative caregiver reports		Teacher reports		
	B(SE)	$\beta$	B(SE)	$\beta$	B(SE)	$\beta$	
Informant factors	Primary caregiver factor	.04(.01)	.25***	.02(.007)	.16**	.02(.01)	.10*
	Alternative caregiver factor	.02(.01)	.16**	.02(.008)	.13*	-.002(.01)	-.01
Construct/ 'trait' factors	Behavior problems factor	.02(.01)	.13**	.02(.009)	.11 <sup>†</sup>	.003(.01)	.02
	DC behavior factor	.03(.01)	.19***	.02(.01)	.17*	.01(.01)	.08

	Rule-breaking						
	Primary caregiver reports		Alternative caregiver reports		Teacher reports		
	B (SE)	$\beta$	B(SE)	$\beta$	B(SE)	$\beta$	
Informant factors	Primary caregiver factor	.02(.003)	.24***	.01(.004)	.13**	.01(.005)	.10*
	Alternative caregiver factor	.01(.004)	.12*	.01(.005)	.13*	.001(.006)	.001
Construct/ 'trait' factors	Behavior problems factor	.01(.004)	.10*	.005(.005)	.06	.003(.006)	.04
	DC behavior factor	.02(.01)	.23***	.01(.006)	.10	.003(.006)	.04

Note.

<sup>†</sup>  $p < .10$ ,

\*  $p < .05$ ,

\*\*  $p < .01$ ,

\*\*\*  $p < .001$ .

PC = primary caregiver; AC = alternative caregiver; CU = callous unemotional; DC = deceitful-callous. See Figure 2 for hypothesized model. We examined whether DC behavior and behavior problem 'trait' factors and primary caregiver and alternative caregiver 'method' factors at age 3 predicted aggression and rule-breaking at ages 9,5 assessed via primary and alternative caregiver and teacher reports (i.e., across informant and settings). We controlled for within-time correlations of DC behavior and behavior problems factors. We ran models controlling for intervention group, project location, child gender, ethnicity and race, family income, and parent education – findings unchanged.

**Table 4** Multi-trait multi-method models examining age 3 primary and alternative caregiver reports of DC behavior and behavior problems predicting aggressive, rule-breaking, and CU behavior at age 9.5

	Aggression						
	Primary caregiver reports		Alternative caregiver reports		Teacher reports		
	B(SE)	$\beta$	B(SE)	$\beta$	B(SE)	$\beta$	
<b>Informant factors</b>	Primary caregiver factor	.04(.01)	.28***	.01(.007)	.08	-.004(.01)	-.02
	Alternative caregiver factor	.01(.01)	.05	.03(.01)	.19**	-.01(.01)	-.03
<b>Construct/ trait factors</b>	Behavior problems factor	.02(.01)	.10	.02(.01)	.13	-.004(.02)	-.02
	DC behavior factor	.05(.01)	.32***	.04(.01)	.28**	.03(.02)	.18, $p = .10$

	Rule-breaking						
	Primary caregiver reports		Alternative caregiver reports		Teacher reports		
	B (SE)	$\beta$	B (SE)	$\beta$	B (SE)	$\beta$	
<b>Informant factors</b>	Primary caregiver factor	.02(.004)	.26***	.01(.004)	.07	.002(.003)	.02
	Alternative caregiver factor	.002(.004)	.03	.01(.004)	.13*	-.01(.006)	-.15*
<b>Construct/ trait factors</b>	Behavior problems factor	.003(.01)	.03	-.001(.007)	-.02	-.001(.01)	-.01
	DC behavior factor	.03(.01)	.33***	.03(.001)	.37***	.01(.01)	.13

Note.

<sup>†</sup>  $p < .10$ ,

\*  $p < .05$ ,

\*\*  $p < .01$ ,

\*\*\*  $p < .001$ .

PC = primary caregiver; AC = alternative caregiver; CU = callous unemotional; DC = deceitful-callous. We examined whether DC behavior and behavior problem factors or primary caregiver and alternative caregiver method factors at age 3 predicted aggression and rule-breaking at ages 9.5 (across informants and settings). We controlled for within-time correlations of DC behavior and behavior problems factors. We ran models controlling for the following relevant demographic covariates: intervention group, project location, child gender, ethnicity and race, family income, and parent education. The pattern of findings was unchanged when these covariates were included in models, and thus results are not presented for brevity.

Zero-order correlations and multiple regression coefficients for associations between primary and alternative caregiver reports of CU behavior versus behavior problems scores at ages 2 and 3 and primary caregiver reports of CU behavior at age 9.5

**Table 5**

Primary caregiver-reported CU behavior at age 9.5 – extracted factors scores based on Waller et al., 2015)											
“general” CU behavior factor			“specific” callous			“specific” uncaring			“specific” unemotional		
<i>r</i>	<i>B</i> (SE)	<i>β</i>	<i>r</i>	<i>B</i> (SE)	<i>β</i>	<i>r</i>	<i>B</i> (SE)	<i>β</i>	<i>r</i>	<i>B</i> (SE)	<i>β</i>
<b>Age 2 within informant</b>											
Behavior problems (PC)	.17***	.004(.001)	.13**	.003(.001)	.13**	-.001	.001(.001)	-.02	.01	-.001 (.001)	-.03
DC behavior (PC)	.18***	.89(.33)	.12**	.01(.27)	.001	.09	.20(.23)	.05	.09*	.60(.28)	.10 <sup>†</sup>
<b>Age 2 across informant</b>											
Behavior problems (AC)	.07	.001(.003)	.01	.002(.002)	.10	-.02	.001(.002)	-.02	-.13 <sup>†</sup>	-.003 (.002)	-.10
DC behavior (AC)	.11*	1.16(.90)	.10	.11(.71)	.01	.01	.33(.52)	.05	-.12*	-.59 (.70)	-.07
<b>Age 3 within informant</b>											
Behavior problems (PC)	.18**	.002(.002)	.06	.004(.001)	.17**	.11*	.002(.001)	.10	-.06	-.003 (.001)	-.11 <sup>†</sup>
DC behavior (PC)	.24***	1.13(.30)	.21***	.08(.26)	.02	.07	.04(.19)	.01	.01	.33(.24)	.08
<b>Age 3 across informant</b>											
Behavior problems (AC)	.17**	.003(.003)	.08	.003(.002)	.11	.02	-.001 (.002)	-.04	.06	-.001 (.002)	-.02
DC behavior (AC)	.20***	.82(.36)	.18*	-.07(.26)	-.02	.09	.45(.26)	.14	.12 <sup>†</sup>	.45(.26)	.13 <sup>†</sup>
<b>Age 4 within informant</b>											
Behavior problems (PC)	.19***	.001(.002)	.05	.004(.002)	.18**	.08	.001(.001)	.01	-.05	.001 (.001)	-.01
DC behavior (PC)	.24***	.76(.24)	.21***	-.09(.20)	-.03	.11*	.23(.14)	.11 <sup>†</sup>	-.07	-.18(.19)	-.06
<b>Age 4 across informant</b>											
Behavior problems (AC)	.09	-.001(.003)	-.04	.004(.002)	.15*	.13*	.002(.001)	.11	-.09	-.01 (.002)	-.21*
DC behavior (AC)	.21***	1.41(.43)	.27***	-.06(.31)	-.02	.11 <sup>†</sup>	.22(.23)	.07	.02	.57(.34)	.14 <sup>†</sup>

Note.

<sup>†</sup> *p* < .10.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

\*  $p < .05$ ,  
\*\*  $p < .01$ ,  
\*\*\*  $p < .001$ . See Figure 2.

PC = primary caregiver; AC = alternative caregiver; CU = callous unemotional; DC = deceitful-callous. We specified within-time correlation between DC behavior and behavior problems at ages 2–4 to account for their overlap as reported in previous studies in this sample (see Hyde et al., 2013). Fisher  $r$ -to- $z$  transformations showed that correlations with CU behavior at age 9.5 were significantly greater in magnitude for age 3 DC behavior than for age 2 DC behavior within alternative caregiver reports ( $z = 1.94, p < .05$ ) and across primary and alternative caregivers reports (range,  $z = 2.11 - 2.34, p < .05$ ). We re-ran models controlling treatment status, project location, child gender, race, and ethnicity, family income, and parent education – findings unchanged