Child Behavior Questionnaire–Short Form for Teachers: Informant Correspondences and Divergences

Hedwig Teglasi[1], Laura Schussler[2], Kathleen Gifford[2], Lee Ann Annotti[1], Catherine Sanders[1], and Huili Liu[1]

Abstract

Discrepancies among independent sources of information about presumably identical constructs argue against reliance on a single perspective. To fill the need for temperament questionnaires for teacher and parent informants, we adapted the popular parent/caregiver Child Behavior Questionnaire–Short Form for preschool and kindergarten teachers. Informant correspondences were low as expected, but patterns were consistent with hypotheses drawn from person perception models. Internal consistencies of the teacher scales were adequate, comparing favorably with those of parent-rated scales. Anticipated relations of temperament scales emerged with social competence and tasks of executive attention for both parent and teacher informants. Confirmatory factor analyses conducted separately for parent and teacher scales supported the familiar three-factor model when allowances were made for cross-loadings and correlated errors. A multigroup confirmatory factor analyses with parent and teacher data indicated that the factor structures of the two questionnaires are similar but not equivalent.

Keywords

Questionnaires are frequently used in studies investigating the role of temperament in children's adjustment (for a review, see Gartstein, Bridgett, & Low, 2012). The temperament construct is accepted as an important underpinning of developmental outcomes, including the acquisition of social competencies that mediate subsequent outcomes (e.g., Eisenberg, Morris, & Spinrad, 2005; Zhou, Main, & Wang, 2010). Self-regulatory aspects of temperament have conceptual and empirical links with executive functions (EFs), competencies that are essential for children's functioning and development (Bridgett, Oddi, Laake, Murdock, & Bachmann, 2013).

Although young children between the ages of 3 and 7 years often attend preschool or kindergarten, researchers have relied primarily on parents as informants. The Child Behavior Questionnaire (CBQ; Ahadi, Rothbart, & Ye, 1993; Rothbart, Ahadi, Hershey, & Fisher, 2001), designed for parents/caregivers, has been widely researched and, even in its short form (CBQ-SF; Putnam & Rothbart, 2006), is regarded as a highly differentiated and comprehensive measure with 15 scales that offer a detailed assessment of both *reactive* and *self-regulatory* aspects of temperament. To fill the need for a teacher version of this instrument, this study adapted the CBQ-SF for use by preschool and kindergarten teachers, calling it the CBQ-TSF. To our knowledge,

no study has focused on the development of a separate teacher version of the entire questionnaire, in either its original or short form.

Recognizing the value of multi-informant protocols, researchers have administered selected CBQ scales to teachers to supplement parent ratings (e.g., Blair & Razza, 2007; Eisenberg et al., 2003; Spinrad, et al., 2006),.. However, because their focus was not on the properties of the scale, the researchers typically selected a limited number of temperament scales, often compositing them into broader factor scores in order to minimize the number of analyses. In the current study, we first developed a teacher report version of the CBQ-SF, then examined convergence between parent- and teacher-reported temperament traits and examined relations of teacher- and parent-rated temperament traits with measures of social competence (SC) and performance on tests of EF. Given previous research

[1]University of Maryland, College Park, MD, USA
[2]Anne Arundel County Public Schools, Annapolis, MD, USA

Corresponding Author:
Hedwig Teglasi, Department of Counseling, Special Education and Higher Education, University of Maryland, 3214 Benjamin Building, College Park, MD 20742, USA.
Email: hteglasi@umd.edu

showing that individual scales within broader factors have differential relations with external correlates, such as preschool behavioral problems (see Gartstein, Putnam, & Rothbart, 2012; Moran, Lengua, & Zalewski, 2013), we examine external correlates separately for each scale. Finally, we conducted confirmatory factor analyses (CFAs) of the teacher and parent scales as well as multigroup CFAs.

Consistent findings across various psychology subfields (see Meyer et al., 2001) of low agreement between *independent* sources of information about purportedly similar psychological phenomena (e.g., between different informants' ratings and between questionnaires and performance tests or laboratory observations) argue against reliance on a single informant. Low correspondences across measures also hold in temperament research (e.g., Kagan, Snidman, McManis, Woodward, & Hardway, 2002). Historically, informant discrepancies have been regarded as unwelcome and treated as measurement error, but persistence of low agreement, even with instruments that are reliable, valid, and have similar factor structures, favors the view that each informant provides unique insights about individuals' functioning in the settings in which they are situated (see De Los Reyes, Thomas, Goodman, & Kundey, 2013; Eid & Diener, 2006).

Further highlighting the importance of assessing traits with multiple informants is the common finding that relations between two constructs, such as a predictor (e.g., temperament) and criterion (e.g., SC), tend to be context/informant specific. In their extensive review, Meyer et al. (2001) reported that the sizes of validity coefficients tend to be moderate to substantial when the *same* informant completes both measures (often reaching .50) and modest (rarely exceeding .30) when *different* informant complete each measure. Hence, to understand relations between theoretically linked psychological phenomena such as temperament and adjustment, it is necessary to give consideration to the sources of information about each construct and method of measurement.

Definition of Temperament

Temperament comprises individual differences in children's styles of engagement with their surroundings, influencing how they respond to various stimuli and how others respond to them. Despite various theoretical approaches, there is consensus on the following as defining features of temperament (see review by Zentner & Bates, 2008): Temperament encompasses variations within the normal range of behaviors pertaining to core domains of affect, activity, attention, and sensory sensitivity that appear early in life; are linked in complex ways to biological mechanisms; and are predictive of conceptually coherent developmental outcomes. Rothbart and her colleagues conceptualized temperament as subsuming two broad biologically based dispositions, *reactivity*

and *self-regulation*, whose expression is subject to influence by maturation and experience (Rothbart & Derryberry, 1981) and by stimuli in the surroundings (see Rothbart & Bates, 2006). Reactive aspects of temperament refer to individual differences in how stimuli elicit physiological, motoric, cognitive, and emotional responses; self-regulatory aspects refer to individuality in resources to modulate reactive response tendencies, primarily through effortful control (EC).

Although the CBQ was not developed with any particular higher order factor structure in mind, exploratory factor analyses of its 15 scales (originally designed to be used separately) have consistently yielded three theoretically relevant factors with both parent- and self-report questionnaires (see Putnam & Rothbart, 2006; Rothbart et al., 2001; Sleddens, Kremers, Candel, De Vries, & Thijs, 2011) even across cultures (Ahadi et al., 1993; Kusanagi, 1993). Two of the factors measure styles of *reactivity* to the surroundings, including Negative Affectivity (NA; Anger/Frustration, Discomfort, Fear, Sadness, and low Soothability) and Extraversion/Surgency (E/S; Activity Level, Impulsivity, High-Intensity Pleasure, low Shyness, Approach/Positive Anticipation, and Smiling/Laughter). A third factor measures capacities enabling *self-regulation* of reactive tendencies through the exercise of EC (Attentional Focusing, Inhibitory Control, Low-Intensity Pleasure, and Perceptual Sensitivity).

Despite the consistency and conceptual utility of the three-factor structure of the parent-rated CBQ family of scales, certain scales are not stable on their factors, cross-loading, or having their highest loading with different factors across samples (see review, Mervielde & De Pauw, 2012).

Temperament, Context, and Informant Perceptions

Correspondences among pairs of informants are particularly low when they differ in the contexts in which they are situated, such as parents and teachers (for a review, see De Los Reyes & Kazdin, 2005). Despite initial conceptions of temperament as consistent across contexts, considerable variation has been observed in children's temperament across situations and within dimensions (Fagot & O'Brien, 1994; Majdandzic & Van den Boom, 2007; Schaughency & Fagot, 1993). Contexts vary in the array of available stimuli that may elicit reactivity, in their requirements for self-regulation, and in the leeway accorded for the expression of emotionality.

To the extent that settings vary in the importance of certain traits to meet adaptive requirements, they are not "functionally equivalent" for those traits (see Mischel, 2004). Although differential pull of contexts would be reflected in mean differences across informants observing in different settings, it would not necessarily alter rank order correspondences in

temperament ratings. What likely drives low informant correspondence is that children differ in their responses to similar environmental stimuli, a central tenet of temperament theory (e.g., Rothbart & Bates, 2006). For example, behaviors associated with shyness are more often expressed in novel than familiar contexts, but only among children who are temperamentally disposed to shyness. If so, then "shyness" rated in a familiar context would not generalize to behavior in an unfamiliar context, and vice versa, thereby lowering informant correspondences when they observe in settings that are differentially salient for this trait. Although finding low informant correspondences on aggressive syndromes, one study showed higher agreement across parents and teachers of children enrolled in a summer program when informants reported similar social events encountered by children in their respective settings (Hartley, Zakriski, & Wright, 2011). Presumably, similarity of social situations permitted informants to observe similar evoking cues and functional requirements.

Parent and teacher informants differ not only in the contexts that they share with children but also in the nature of their relationships with those children. Multiple sources of influence join with context to shape the mind-sets that filter what informants notice, how they recollect observations, and what they are inclined to report on questionnaires about child behaviors (De Los Reyes & Kazdin, 2005). Models of person-perception offer avenues for assessors to clarify informants' perspectives. Funder's (1995) realistic accuracy model (RAM) proposes the following influences on raters' judgments: (a) the *relevance* of the measured trait to the informant, (b) the *availability* of the trait to observation by the informant, (c) *detection* of the trait by the informant, and (d) *utilization* (interpretation) of trait-relevant observations by the informant. Using this model, Tackett (2011) predicted divergent and convergent ratings of mothers and fathers about their child's personality and behavior problems, reasoning that agreement would be highest for traits that are relevant and available to observation and lowest for traits that are difficult to observe.

Likewise, parent and teacher temperament ratings are subject to influence by the *relevance* of the trait in the home and school setting and the *availability* of the trait to observation by the informant in that setting. Although overtly expressed traits (e.g., externalizing) are more easily observed than traits that are more internal (see Achenbach, McConaughy, & Howell, 1987; De Los Reyes & Kazdin, 2005), parents may be more likely to notice traits that contribute to their affective bonds with their children (see Mangelsdorf, Schoppe, & Buur, 2000), whereas teachers may be more concerned about traits that contribute to functioning in the classroom. Differences between home and school settings in behavioral norms and expectations about when and how to express emotions (e.g., Allan & Gilbert, 2002) may also influence temperament ratings. It seems reasonable that behavioral and emotional responses that depart from normative expectations are more noticeable to observers.

Informant Correspondence and External Correlates

If each informant's perception is grounded in observations in the context shared with the child, generalization of ratings to functioning may be best when the assessment and target contexts include common elements that are relevant to the rated attribute. One study, conducted under controlled laboratory conditions, demonstrated informant-specific external correlates of child disruptive behavior symptoms (De Los Reyes Henry, Tolan, & Wakschlag, 2009). When parents but not teachers reported disruptive behavior, the child tended to behave disruptively during parent–child interactions but not during interactions with the clinical examiner; and, when teachers but not parents reported disruptive behavior, the child was disruptive with the examiner but not with the parent. In a similar vein, stronger links between temperament and later externalizing behavior problems were found when mothers report on both than when mothers report on temperament and other informants report on behavior (Copeland, Landry, Stanger, & Hudziak, 2004). Likewise, parent-rated temperament (preadolescent EC) predicted problems at home but not at school, Rettew et al., 2011).

Current Study

Initial construct validity of the teacher form, the CBQ-TSF, derives from the CBQ-SF in terms of scale definitions and inclusion of items (detailed in the Method section). We relied on extensive pilot work to inform wording changes of items on the parent questionnaire (CBQ-SF) to assure appropriateness for the classroom without changing their intent.[1] We sought further support for construct validity via three traditional criteria: internal consistency reliability; two types of external correlates, informant based and performance based; and theoretically meaningful factor structure. To date, factor analyses of the CBQ scales have been conducted only with parent questionnaires.

We selected SC as the informant-based criterion due to its centrality to child adjustment (e.g., Blair, 2002; Masten & Coatsworth, 1998; Raver, 2002; Webster-Stratton & Reid, 2004) and its established associations with temperament (Moran et al., 2013), and we chose EF as the performance-based criterion for similar reasons, its importance to adjustment and its empirical links with the EC component of temperament. Although EF and EC are studied in different fields, both encompass self-regulatory processes that enable children to function in academic and social-emotional arenas (see Bridgett et al., 2013; Liew, 2012).

Hypotheses

Informant Correspondences

Although low parent–teacher correspondences (non-significant or <.30) are expected, following Tackett's (2011) reasoning, patterns of higher or lower agreement may be anticipated. Child attributes that contribute more to interactions with the informant are likely to be more *relevant*, more *available* to observation, and hence, more likely to be *detected* and *used* by the informant. Therefore, traits that are similarly relevant to functioning in home and school contexts and more overly expressed, and hence more observable, should have the highest parent–teacher correspondences.

Specifically, we expect relatively higher agreement on two self-regulatory aspects of temperament, Attention Focusing and Inhibitory Control, having been linked to outcomes from both parent and teacher perspectives (e.g., Blair & Razza, 2007), and relatively higher agreement on three E/S traits that are similar to the externalizing traits that are reportedly more observable: Activity, High-Intensity Pleasure, and Impulsivity.

Informant-Based External Correlate

We expect to replicate previously documented relations of temperament and SC but primarily within teacher informants in line with previous research showing low validity coefficients across different informants (Meyer et al., 2001). Specifically, within each informant, we expect SC to be (a) positively correlated with EC scales, particularly Attention Focusing and Inhibitory Control (e.g., Eisenberg et al., 2003; Eisenberg, Spinrad, & Morris, 2002); (b) inversely correlated with NA, particularly with Anger/Frustration (e.g., Coplan & Bullock, 2012; Corapci, 2008; Moran et al., 2013; Spinrad et al., 2006); and (c) positively associated with Smiling/Laughter (see review, Putnam, 2012).

Performance-Based External Correlates

Although EC and EF are studied in different fields (temperament and neurocognitive psychology, respectively) and are typically measured in different ways (questionnaires and performance tasks, respectively), they are conceptually similar in that both capture the exercise of control over one's behavior (Rueda, Posner, & Rothbart, 2011). Performance on EF tasks correlates with parent and self-reported EC (Gerardi-Caulton, 2000; Rothbart, Ellis, Rueda, & Posner, 2004; Simonds, Kieras, Rueda, & Rothbart, 2007) as well as with teacher-reported EC (e. g., Blair & Razza, 2007). However, effect sizes are usually low to moderate (as would be expected with different measurement methods). In the current study, we anticipate teacher-rated EC scales to correlate with EF tasks.

Relations of NA scales with EF tasks have been understudied. However, socialization of children into the culture of schools focuses on reigning in their intense emotions (e.g., Olson, Sameroff, Lunkenheimer, & Kerr, 2009; Vohs & Baumeister, 2011). To the extent that emotional outburst are not acceptable in school, children are required to exercise self-restraint (Inhibitory Control) to moderate their emotional expressions at school. Supporting this idea is the finding that the intensity of preschoolers' anger expression was associated with self-regulation when ratings were provided by school personnel but not by parents (Eisenberg, Fabes, Nyman, Bernzweig, & Pinuelas, 1994). Among children and adults, those with higher Anger/Frustration have lowest level of cognitive self-regulation, measured with tasks of EF (Wilkowski & Robinson, 2010). To the extent that venting intense emotions, such as Anger/Frustration, in the school context reflects a lapse of self-regulation, it would be reasonable to anticipate significant associations between teacher-rated NA and EF.

Factor Structure

Similarity of factor structures across groups or raters bolsters confidence in construct validity (see Georgas, van de Vijver, Weiss, & Saklofske, 2003; Reynolds & Carson, 2005). We anticipate CFAs to support the familiar three-factor structure for the CBQ-TSF but only when cross-loadings and post hoc refinements are included. Given that initial specifications of cross-loadings are based on parent data reviewed earlier, we expect that post hoc modifications will be necessary to obtain an adequate model–data fit for teachers. In a multiple-group CFA, we anticipate an adequate fit between the best parent model and the best teacher model but only when measurement invariance is not assumed.

Method

Development of the CBQ-SF for Teachers

The 94-item CBQ-SF[2] (Putnam & Rothbart, 2006), which includes 15 scales that assess temperament in children ages 3 to 7 years, was derived from the 195-item CBQ (Rothbart et al., 2001). After obtaining permission from the authors of the CBQ-SF (Putnam & Rothbart, 2006), modifications were introduced to develop the teacher version, retaining the conceptual framework of the CBQ-SF in terms of scale definitions and intent of the items.

The original instructions direct the respondent to "read each statement and decide whether it is a *true* or *untrue* description of "your child's reaction within the past six months." The 7-point Likert-type scale ranges from 1 (*extremely untrue of your child*) to 7 (*extremely true of your child*). The instructions also direct the respondent to mark

Not Applicable (N/A) if the behavior described is not observable in the setting. These instructions and response format were retained, except that the phrase *your child* was replaced with *the above-named child*. In addition, revisions were made to 20 items to increase their relevance to the preschool classroom, keeping their original meaning. For example, the item "gets angry when told she or he has to go to bed" was changed to "gets angry when told she or he is told to remain still during rest or quiet time." Subsequent modifications were made to six additional items after examining teacher comments during a pilot study in which 12 preschool teachers participated. All original items of the CBQ-SF were kept in the teacher version (CBQ-TSF) with a total of 26 revised items.

Participants

Preschool Sample. Participants included preschool students (134; 46.5% males, 51.4% females), their teachers (14), and their parents (106, primarily mothers). All of the children attended an on-campus preschool at a large university in the Mid-Atlantic region of the United States. The mean age of the preschoolers was 57.38 months (age range, 38-82 months). Although this sample was largely middle class, it was otherwise diverse, including 35.9% European Americans, 9.2% African Americans, 9.9% Asian Americans, and 9.9% "Other." Ethnicity data were missing for 13.4% of the children in the sample. All teacher raters were female, and no ethnicity data were available for these participants.

Kindergarten Sample. Participants included kindergarten students (105; 54 males, 51 females), as well as their parents/guardians (70) and teachers (28; 1 male and 27 females of whom 2 were Asian American and 25 European American), recruited from five private schools within the greater Washington, D.C., metropolitan area, and one private school from the greater Chicago area. The mean age of participants was 70.0 months (age range = 60-83 months). The majority of the kindergarten students, 63.8%, were European American, 11.4% were African American, 9.5% Hispanic American, and 4.8% did not disclose this information. The sample was largely middle class.

Procedures

Similar procedures were followed for both samples. Packets for each child with informed consent on file were prepared and distributed to teachers (in preschool, the CBQ-TSF and Social Competence and Behavior Evaluation [SCBE]; in kindergarten, the CBQ-TSF and Social Skills Improvement System [SSIS]) and to parents (in preschool, the CBQ-SF; in kindergarten, the CBQ-SF and SSIS). Researchers followed up with parents and teachers to collect the completed forms, which were then checked for missing items, and, if necessary, they called parents or teachers for a phone interview or redistributed the forms to secure complete data. The performance tasks were administered individually to each child by psychology doctoral students trained in the data collection protocol. In the preschool sample, nearly half of the parents contacted returned signed permission forms after a couple of reminders, and more than three quarters completed the questionnaires. Almost all of the preschool teachers completed the questionnaires for children whose parents gave permission. In the kindergarten sample, the proportion of contacted parents who participated was variable, averaging about a third. Again, teachers were more likely to complete the questionnaires than parents.

Measures

The CBQ-SF (Putnam & Rothbart, 2006) was introduced as an investigative tool that maintains the scope and depth of the original but is less time-consuming to administer. The CBQ-SF includes 94 items derived from the 195-item standard form (CBQ; Rothbart et al., 2001). Similar to the standard CBQ, each of the 15 scales of the CBQ-SF refers to individual differences in a primary temperamental characteristic that may be conceptualized as reactive or self-regulatory and is measured reliably. To illustrate, an item from the Activity Level scale is *Seems always in a big hurry to get from one place to another*. Response options range from 1 (*extremely untrue of this child*) to 7 (*extremely untrue of this child*) with the opportunity to indicate that the item does not apply (N/A). This questionnaire was administered to parents in both preschool and kindergarten samples.

The CBQ-TSF, an adaptation of the parent/caregiver version, includes the same scales and items except that 26 of the 94 items were modified to suit the classroom context. Internal consistencies (described later) were acceptable. This questionnaire was administered to teachers in both preschool and kindergarten samples.

The Developmental Neuropsychological Assessment (NEPSY, 1st ed., and NEPSY-II, 2nd ed.; Korkman, Kirk, & Kemp, 1998, 2007) is a commercially available test that measures functioning in several domains. Age-appropriate tasks were selected from the Attention and Executive Functioning domain to administer to preschool and kindergarten children. In the preschool sample, two subtests from the NEPSY-II were administered to children younger than 5 years, the Visual Attention and Statue subtests (*n* = 75). Four subtests from the NEPSY-II were administered to all children in the kindergarten sample, Auditory Attention, Design Fluency, Inhibition, and Statue. An additional subtest, Tower, was included from the earlier edition of the NEPSY. Each of the NEPSY tasks captures a narrow facet of attention/EF and are not designed to be aggregated into an overall score; therefore, subtest scaled scores were used in the analyses.

The SSIS Rating Scales (Gresham & Elliott, 2008) is a commercially available questionnaire that includes parent and teacher versions of three scales. The appropriate version of the 46-item SC scale was administered to parents and teachers of children in the kindergarten sample, and standard scores were used. The SC scale includes items about communication, cooperation, assertion, responsibility, empathy, engagement, and self-control (e.g., "Follows your directions"). Items are rated according to how often the behavior occurs on a 4-point Likert-type scale (1 = *never*, 4 = *always*). In the current study, internal consistencies were high for teacher (.86) and parent (.87) Total Social Skills scores, similar to authors' report (Gresham & Elliott, 2008; Gresham, Elliott, Cook, Vance & Kettler, 2010).

The SCBE, Preschool Edition (ages 2.5-6 years), Short Form (LaFreniere & Dumas, 2003), is a commercially available 30-item scale, designed for teachers to complete by responding to 6-point Likert-type scale to indicate the child's "typical behavior or emotional state" (e.g., "Works easily in a group") based on how often it occurs (1 = *almost never occurs* to 6 = *almost always occurs*). Internal consistencies of the scales comprising SC are high (ranging from .91 to .79), and normative data are extensive (LaFreniere & Dumas, 2003). This scale was administered to teachers in the preschool sample, and standard scores were used.

Results

Preliminary Analyses

Since teacher ratings are clustered within classrooms, intraclass correlation coefficients (ICCs) were examined to estimate the effect of this grouping. In the preschool sample, ICC values for 13 of the scales were small and nonsignificant, ranging from .00 to .05. Two reached significance (Approach Positive and Soothability, .06 and .11, respectively). Since the ICC values were small (with one exception), subsequent analyses did not correct for nesting effects. However, since children were assigned to classrooms based on age, and there are modest correlations of age in months with some temperament scales in preschool, correlational analyses controlled for age. In the kindergarten sample, the ICC analyses did not converge because of insufficient numbers of students rated by each teacher.

Follow-up procedures kept missing data to a minimum particularly in the preschool sample, with less than 1% of items left incomplete. In the kindergarten sample, about 6% of teacher-rated items and less than 1% of parent-rated items were missing. Missing data were primarily due to informants choosing the N/A option on the CBQ. All missing items were assigned that item's mean score as rated by the relevant informant. The CBQ questionnaires from two teachers were not usable. The most frequent items that were left unanswered by kindergarten teachers were those on the

High-Intensity Pleasure and Low-Intensity Pleasure scales, but no pattern was discerned for parents.

Analyses were conducted separately for preschool and kindergarten samples, with a focus on investigating the properties, external correlates, and factor structure of the teacher version of the CBQ-SF.

Internal Consistency Reliability

As an index of internal consistency, alpha values of .70 or above are widely regarded as the standard. DeVellis (1991) considered alphas of .60 as undesirable but not unacceptable. Table 1 presents the internal consistencies of the CBQ-SF and CBQ-TSF scales in both samples as well as for those of the CBQ-SF reported by Putnam and Rothbart (2006). In the preschool sample, all but two CBQ-TSF scales had internal consistencies at or above .70. The Low-Intensity Pleasure and Sadness scales had internal consistencies of .67 and .68, respectively, but item analyses showed that deletion of one or more items would not result in improvement. In the kindergarten sample, only one CBQ-TSF scale fell below .70, Sadness (.69).

Internal consistencies for the CBQ-SF were lower than for the CBQ-TSF. In the preschool sample, six scales fell below .70 but none below .60. In the kindergarten sample, five scales fell below .70, three of which were below .60 (.57, .55, & .56, respectively for Low-Intensity Pleasure, Sadness, and Approach/Positive Anticipation).

Informant Correspondences

To examine correspondences between parent and teacher informants on temperament scales, we use correlational analyses (Pearson's). In describing the effect sizes, we refer to correlation coefficients of about .10, .30, and .50 as "small", "moderate", and "large," respectively (Cohen, 1988).

As anticipated, informant correspondences in both preschool and kindergarten samples were low. Only 7 of 15 correlations in the preschool sample reached significance, with only 2 exceeding .30 (Impulsivity and Shyness). Among these 7 correlations were 4 of the 5 that were expected to show informant convergence (Activity Level, High-Intensity Pleasure, Impulsivity, and Inhibitory Control, but not Attention Focusing). In the kindergarten sample, only 6 of 15 correlations reached significance, but 5 exceeded .30 (Attentional Focusing, Soothability, High-Intensity Pleasure, Impulsivity, and Inhibitory control). All five of the scales that were hypothesized to show convergent relations were included in these six. The exception was Falling Reactivity/Soothability.

It is interesting to note that in this study, low informant correspondences (Pearson's *r* = .17) were also found with the parent and teacher forms of the commercially available SSIS in the kindergarten sample.

Table 1. Rater Correspondences and Internal Consistencies of CBQ-SF and CBQ-TSF.

Scale (N items)	Correspondence (Pearson's coefficient)		Alpha coefficient				Putnam and Rothbart (2006)		
	Preschool	Kindergarten (K)	CBQ-SF, preschool (N = 106)	CBQ-TSF, preschool (N = 134)	CBQ-SF, K (N = 71)	CBQ-TSF, K (N = 94)	Sample 1 (N = 138)	Sample 2 (N = 289)	Sample 3 (N = 169)
Effortful Control									
Attentional Focusing (6)	.14	.45**	.78	.79	.78	.81	.73	.70	.70
Inhibitory Control (6)	.25*	.63**	.65	.82	.73	.83	.62	.68	.72
Low-Intensity Pleasure (8)	.19	−.03	.66	.67	.57	.76	.82	.60	.68
Perceptual Sensitivity (6)	.10	.01	.76	.71	.68	.86	.69	.60	.73
Negative Affectivity									
Anger/Frustration (6)	.24*	21	.80	.86	.83	.92	.78	.72	.69
Discomfort (6)	.10	.05	.86	.83	.75	.82	.70	.69	.82
Fear (6)	−.04	.11	.74	.70	.78	.87	.54	.64	.60
Sadness (7)	.21*	.01	.65	.68	.55	.69	.61	.43	.46
Soothability/Falling Reactivity (6)	.20	.31**	.79	.79	.83	.76	.71	.67	.80
Extraversion/Surgency									
Activity Level (7)	.22*	.28*	.69	.88	.75	.87	.74	.65	.72
Approach/Positive Anticipation (6)	.23*	.12	.68	.81	.56	.81	.58	.57	.70
High-Intensity Pleasure (6)	.27**	.46**	.74	.89	.76	.93	.70	.66	.76
Impulsivity (6)	.40**	.41**	.73	.83	.62	.80	.54	.62	.74
Shyness (6)	.60**	.22	.86	.88	.89	.83	.79	.82	.87
Smiling and Laughter (6)	.14	.05	.61	.87	.74	.85	.64	.55	.77

Note. CBQ-SF = Child Behavior Questionnaire–Short Form; CBQ-TSF = Child Behavior Questionnaire–Teacher Short Form.
*p < .05. **p < .01.

External Correlates

To detect differential relations of scales within factors with external correlates, correlational analyses were conducted separately for each of the 15 scales rather than collapsing them into broader factors scores, a procedure commonly used in temperament research.

Informant-Based External Correlates. To alleviate concern about the possibility that correlations may be inflated due to item overlaps such as those found between temperament scales and symptom checklists (e.g., Lemery, Essex, & Smider, 2002; Lengua, West, & Sandler, 1998), the wording of items was examined, showing no overlaps of temperament with SC questionnaires.

Within-Informant Correlations. In the preschool sample, within-informant correlations were available only for teachers (see Table 2), whereas in the kindergarten sample, within-informant associations were available for teachers and parents (see Table 3).

When preschool teachers rated both temperament and SC (with the SCBE), 11 of 15 correlations reached significance (see Table 2), and 9 of these associations (all but 2) exceeded .30. In line with expectations, all four of the EC scales and all five of the NA scales correlated significantly with SC. Among the E/S scales, as predicted, Smiling/Laughter correlated with SC. Although not predicted, Shyness was also correlated significantly.

When kindergarten teachers rated both temperament and SC, 9 of the 15 correlations reached significance, with 8

equaling or exceeding .30. Similar to the pattern found in preschool, all four teacher-rated EC scales correlated significantly with SC (three equaled or exceeded .30). Of the five NA scales (all of which reached significance in the preschool sample), three scales reached significance (all equaled or exceeded .30; Anger/Frustration, Sadness, and Soothability). Finally, as expected, within kindergarten teachers, Smiling/ Laughter (E/S) correlated (exceeding .30).

When kindergarten parents rated both temperament and SC, 8 of 15 temperament scales correlated significantly with SC, with 5 exceeding .30. Two of the four EC scales reached significance (Attention Focusing and Inhibitory Control), and these two scales are the most frequently chosen in temperament studies to represent EC and have the most empirical support for links with SC. Four of the five NA scales (all but Sadness) reached significance (with two exceeding .50). Again, of the E/S scales, Smiling/Laughter correlated as expected with SC (exceeding .30), but so did Shyness.

Between-Informant Correlations. In the preschool sample, between-informant data were available only for parent-rated temperament and teacher-rated SC. With one exception (Shyness $r = −.24$; see Table 2), none of the temperament scales showed significant cross-informant relations with SC. In the kindergarten sample, cross-informant correlations were available in both directions: when parents rated temperament and teachers rated SC and when teachers rated temperament and parents rated SC (Table 3). Only two temperament dimensions, both in the EC factor, Attention

Table 2. External Correlates of Parent- and Teacher-Rated Temperament Scales in the Preschool Sample.

	NEPSY		
CBQ scales	Visual Attention	Statue	Social Competence (teacher)
Effortful Control			
Attention Focusing			
Parent	−.03	.17	.11
Teacher	.14	.36**	.48**
Inhibitory Control			
Parent	.02	.11	.21
Teacher	.28*	.45**	.45**
Low-Intensity Pleasure			
Parent	.09	.37**	.02
Teacher	−.12	.07	.31**
Perceptual Sensitivity			
Parent	−.03	.14	.06
Teacher	.14	.03	.31**
Negative Affectivity			
Anger/Frustration			
Parent	−.14	−.14	−.04
Teacher	−.19	−.18	−.39**
Discomfort			
Parent	−.16	−.17	.07
Teacher	−.32*	−.06	−.24*
Fear			
Parent	.18	−.23	−.08
Teacher	−.04	−.25*	−.37*
Sadness			
Parent	−.02	−.12	.10
Teacher	−.24	−.19	−.25*
Soothability/Falling Reactivity			
Parent	.14	.16	.06
Teacher	.30*	.22	.43**
Extraversion/Surgency			
Activity Level			
Parent	.10	−.01	−.01
Teacher	−.11	−.25*	.01
Approach/Positive			
Parent	−.03	−.06	.09
Teacher	.02	−.18	.11
High-Intensity Pleasure			
Parent	.12	.24	.13
Teacher	.02	.04	.11
Impulsivity			
Parent	.08	.10	.16
Teacher	−.08	−.20	.08
Shyness			
Parent	−.23*	−.19	−.24**
Teacher	−.07	−.06	−.36***
Smiling/Laughter			
Parent	.15	.28*	.12
Teacher	.08	.17	.46**

Note. CBQ = Child Behavior Questionnaire; NEPSY = Developmental Neuropsychological Assessment.
*p < .05. **p < .01. ***p < .001.

Table 3. External Correlates of Parent- and Teacher-Rated Temperament Scales in the Kindergarten Sample.

CBQ scales	NEPSY					Social Competence (parent)	Social Competence (teacher)
	Auditory Attention	Design Fluency	Inhibition	Statue	Tower		
Effortful Control							
Attention Focusing							
Parent	.33**	.17	.22	.17	.23	.29*	.30*
Teacher	.39**	.32**	.38**	.20	.27*	.11	.53**
Inhibitory Control							
Parent	.35**	.23	.32**	.27*	.14	.46**	.33**
Teacher	.42**	.31**	.28**	.21*	.31**	.20	.54**
Low-Intensity Pleasure							
Parent	−.05	−.09	−.06	.07	.09	.23	.12
Teacher	.28**	.13	.30**	.10	−.06	−.07	.38**
Perceptual Sensitivity							
Parent	.04	.02	.11	.28*	−.21	.11	.18
Teacher	.09	.07	.18	−.08	−.01	−.14	−.26*
Negative Affectivity							
Anger/Frustration							
Parent	−.23	−.12	−.23	−.08	.09	−.51**	−.19
Teacher	−.29*	−.38**	−.22	−.33**	−.21	−.17	−.44**
Discomfort							
Parent	.02	.11	.18	.08	−.10	−.25*	.08
Teacher	.03	−.16	−.10	−.14	−.19	−.01	−.14
Fear							
Parent	.00	.23	−.02	.23	−.11	−.27*	.03
Teacher	.05	−.01	.09	−.10	−.26*	.08	.08
Sadness							
Parent	.01	−.12	−.18	.08	−.07	−.18	.02
Teacher	−.07	−.31**	−.14	−.35**	−.23*	−.15	−.30**
Soothability/Falling Reactivity							
Parent	.11	.21	.10	.20	−.03	.55**	.21
Teacher	.30**	.36**	.20	.09	.10	.16	.33**
Extraversion/Surgency							
Activity Level							
Parent	−.18	−.12	−.05	−.16	−.02	−.08	−.12
Teacher	−.12	−.20	−.22*	−.31*	−.12	−.08	−.35**
Approach/Positive							
Parent	−.10	−.03	.01	.06	.12	.07	.07
Teacher	−.04	−.12	−.05	−.21	.03	−.06	−.11
High-Intensity							
Parent	−.19	−.16	−.28*	−.20	.10	−.16	−.16
Teacher	−.01	−.14	−.12	−.22*	−.10	.15	−.22
Impulsivity							
Parent	−.21	−.03	−.25*	−.27*	.04	.04	−.02
Teacher	−.05	−.02	.03	−.25*	−.06	−.23	.21
Shyness							
Parent	.12	.08	.02	.24*	.17	−.50**	.07
Teacher	−.08	−.21	.08	.10	−.13	.11	−.20
Smiling/Laughter							
Parent	−.11	.06	−.17	−.08	.00	.36**	.12
Teacher	−.04	.07	.12	.05	.01	.07	.31**

Note. CBQ = Child Behavior Questionnaire; NEPSY = Developmental Neuropsychological Assessment.
*p < .05. **p < .01.

Focus and Inhibitory Control, reached significance (.30 and .33, respectively) but only when parents rated temperament and teachers rated SC.

Performance-Based External Correlates

Correlational analyses (Pearson's) were conducted to examine relations of teacher-rated temperament with performance on NEPSY tasks. Although hypotheses were limited to EC and NA scales, to be inclusive, Tables 2 and 3 show links with all temperament scales (for the preschool and kindergarten samples and parent as well as teacher informants, respectively).

Three of the four teacher-rated EC scales were associated with NEPSY tasks. Inhibitory control was associated with all of the NEPSY tasks administered in both samples. Attention Focus also showed considerable associations with NEPSY tasks (four of five for kindergarten teachers and one of two for preschool teachers). Low-Intensity Pleasure was associated with fewer NEPSY tasks, two of five associations in kindergarten, and none in preschool. No significant association emerged between Perceptual Sensitivity and NEPSY tasks for teacher informants. We also note that parent-rated EC scales also correlated with NEPSY tasks, most strongly with Inhibitory Control in the kindergarten sample (with three of five tasks).

With respect to NA, reasoning that the expression of intense negative affect in the classroom setting signals a lapse in self-regulation, we anticipated teacher-rated NA scales to correlate inversely with NEPSY tasks and found support for the hypothesis. Among kindergarten teachers, the following NA scales showed negative correlations with NEPSY task performance: Anger/Frustration (three of five), Sadness (three of five), and Soothability (two of five). In preschool, three of five teacher-rated NA scales were inversely associated with NEPSY tasks: Discomfort (one of two), Fear (one of two), and Soothability (one of two). It is noteworthy that none of the parent-rated NA scales was correlated with NEPSY tasks in either sample.

Factor Analyses

Recommendations regarding acceptable sample sizes for factor analyses have varied widely, with some proposing absolute numbers (ranging from $n = 100$ to $n = 200$; see Anderson & Gerbing, 1984; Boomsma, 1982) and others preferring a certain number of cases per parameter (e.g., five per model parameter; see Jackson, 2001, 2003). More recent approaches to determining adequacy of sample size focus on characteristics of the design and construct reliability (number of indicators per factor and magnitude of factor loadings). According to simulations conducted by Gagne and Hancock (2006), under certain conditions, samples of 25 were sufficient for converging replications (without

improper parameter estimates). Given about five indicators per factor and loadings averaging about .60, a sample size of 100 would be considered adequate.

We conducted a series CFAs in the preschool sample using Mplus Version 6.12 (Muthen & Muthen, 1998-2011) with maximum likelihood estimation. To examine the factor structures separately for teacher and parent data sets, we tested three types of models. In Type 1 models (1 and 4, for parents and teachers, respectively), each of the 15 scales was assigned to its originally designated factor. In Type 2 models (2 and 5, for parents and teachers, respectively), cross-loadings were specified in accord with previously reported findings with parent ratings. Type 3 models (3 and 6, for parents and teachers, respectively) included modifications to the Type 2 model suggested by the software.

Based on previous factor analyses of parent-rated CBQ scales (Rothbart et al., 2001; Sleddens et al., 2011), the following cross-loadings were allowed in Type 2 models: Approach/Positive (E/S and N/A), Inhibitory Control (EC and E/S), Attention Focus (EC and E/S), Smiling/Laughter (E/S and EC), and Shyness (E/S and NA).

The criteria to evaluate model–data fit (displayed in Table 4) are described next: For the root mean square error of approximation, lower values indicate a better fit with values of .01, .05, and .08 viewed, respectively, as excellent, good, and mediocre, and values greater than .10 as poor (Browne & Cudeck, 1992; MacCallum, Browne, & Sugawara, 1996). The root mean square error of approximation, an absolute measure of fit that adjusts for degrees of freedom and sample size, is biased toward being smaller with increasing degrees of freedom and larger sample sizes.

The comparative fit index is an incremental measure of model fit indicating the proportion in the improvement of the model relative to a null model, and values of greater than .90 indicate adequate fit. The standardized root mean squared residual, an absolute fit index, is a standardized summary of the average covariance residuals, and values lower than .10 indicate adequate fit.

Because of the limitations of the chi-square statistic in CFAs (Barret, 2007), we rely on the other indices to evaluate model–data fit.

As summarized in Table 4, the model fit indices for both informants improve when cross-loadings are allowed (Type 2 model–data fit is better than Type 1 model–data fit). Despite the improvement, indicators fell short of criteria for a good fit for Type 2 models. However, with the addition of modifications in Type 3 models (3 and 6), the model–data fit indicators improve further and are consistent with a good fit for parent and teacher versions of the CBQ, with the exception of the chi-square tests, a pattern that is not unusual. The Akaike information criterion and Bayesian information criterion values are both comparative fit indices with smaller values showing better fit. Whereas the

Table 4. Model–Data Fit Indices.

	Parent			Teacher			Multigroup, Model 7
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	
AIC	3954.83	3882.21	3851.49	4361.56	4321.00	4225.90	8079.50
BIC	4041.45	3989.83	3961.74	4452.99	4434.59	4350.58	8394.08
χ^2	243.04	154.41	121.70	284.08	227.52	124.42	606.75
	$df = 102$, $p < .0001$	$df = 94$, $p = .0001$	$df = 93$, $p = .0245$	$df = 102$, $\|p < .0001$	$df = 94$, $p < .0001$	$df = 90$, $p = .0095$	$df = 385$, $p < .0001$
RMSEA (<.06)	.116	.079	.055	.123	.11	.057	.067
90% CI	[.098, .135]	[.056, .101]	[.021, .08]	[.106, .140]	[.092, .128]	[.029, .08]	[.057, .077]
p(RMSEA ≤.5)	<.0001	<.022	<.367	<.0001	<.0001	<.308	<.005
CFI (≥.90)	.621	.838	.923	.786	.843	.96	.842
SRMR (<.09)	.128	.083	.075	.128	.108	.076	.093

Note. AIC = Akaike information criterion; BIC = Bayesian information criterion; *df* = degrees of freedom; RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index; SRMR=standardized root mean squared residual.

Bayesian information criterion adjusts for sample size, the Akaike information criterion does not. However, the two indices were comparable in all of the analyses. The best fitting models (3 and 6, for parents and teachers, respectively) are shown in Figures 1 and 2, and the parameter estimates are shown in Table 5.

Examination of the best parent model (Model 3) shows that 14 of the 15 scales loaded significantly on their assigned factors; the exception was Attention Focusing, which could not be assigned to any factor as it was distributed across all three. Modifications of Model 3, eliminated the loading from construct E/S to Attention Focus (EC) and allowed correlated errors between Shyness (E/S) and Impulsivity and between Shyness (E/S) and High-Intensity Pleasure (E/S). These two pairs of scales appear to share variance not accounted for by the latent factor (E/S) on which they load.

Examination of the best teacher model (Model 6) shows that every scale correlated significantly with its assigned factor. For this model, post hoc modifications included eliminating and adding cross-loadings and allowing correlated errors among several observed variable pairs. Figure 2 shows the loadings and cross-loadings of scales on each of the three factors, and Table 5 includes the parameter estimates.

Finally, to test for measurement invariance, we conducted multigroup CFAs comparing the model with and without the assumption of configural invariance. The multigroup CFA that did not assume measurement invariance (Model 7; see Figure 3) included the best fitting models for parents and teachers at the measurement level (3 and 6), allowing different factor loading structures for parent and teacher data sets. Because parents and teachers rated the same group of students, the error terms were allowed to correlate across the two sets of scales. At the structural level, we allowed the latent factors to be fully correlated. The fit

indicators, shown in Table 4, support the conclusion of a marginally good model–data fit.

The model assuming configural invariance (Model 8) constrained both sets of data to the structure that best fit the teacher data (when parent ratings were used to constrain both sets, the model did not converge). Since Model 8 was a poorer fit with the data than Model 7, findings did not support configural measurement invariance, and because configural invariance is the least stringent test of measurement invariance, no further tests were conducted.

Discussion

In the context of widely documented informant discrepancies, we adapted the CBQ-SF (Putnam & Rothbart, 2006) designed for parents/caregivers, to assess temperament in children ages 3 to 7 years, for use by teachers (CBQ-TSF) of preschool and kindergarten children. Although informant correspondences were low as expected, they followed a pattern predicted from person perception models (Funder, 1995; Tackett, 2011). In addition, scales were internally consistent and meaningfully related to SC and performance tasks of EF. Finally, with allowance for cross-loadings, the scales of the CBQ-TSF fit the familiar three-factor structure. Yet, despite sharing similar factors structures, the fit between parent and teacher scales is not good if factor loadings are constrained to be the same. Hence, assumptions of configural measurement invariance were not met.

Internal Consistency Reliability

Overall, in the current study, alpha coefficients for all 15 CBQ-TSF scales met criteria for adequate internal consistency, and coefficients tended to be higher than those found with the parent/caregiver questionnaire, the CBQ-SF. Alpha coefficients of the CBQ-SF found in this investigation were
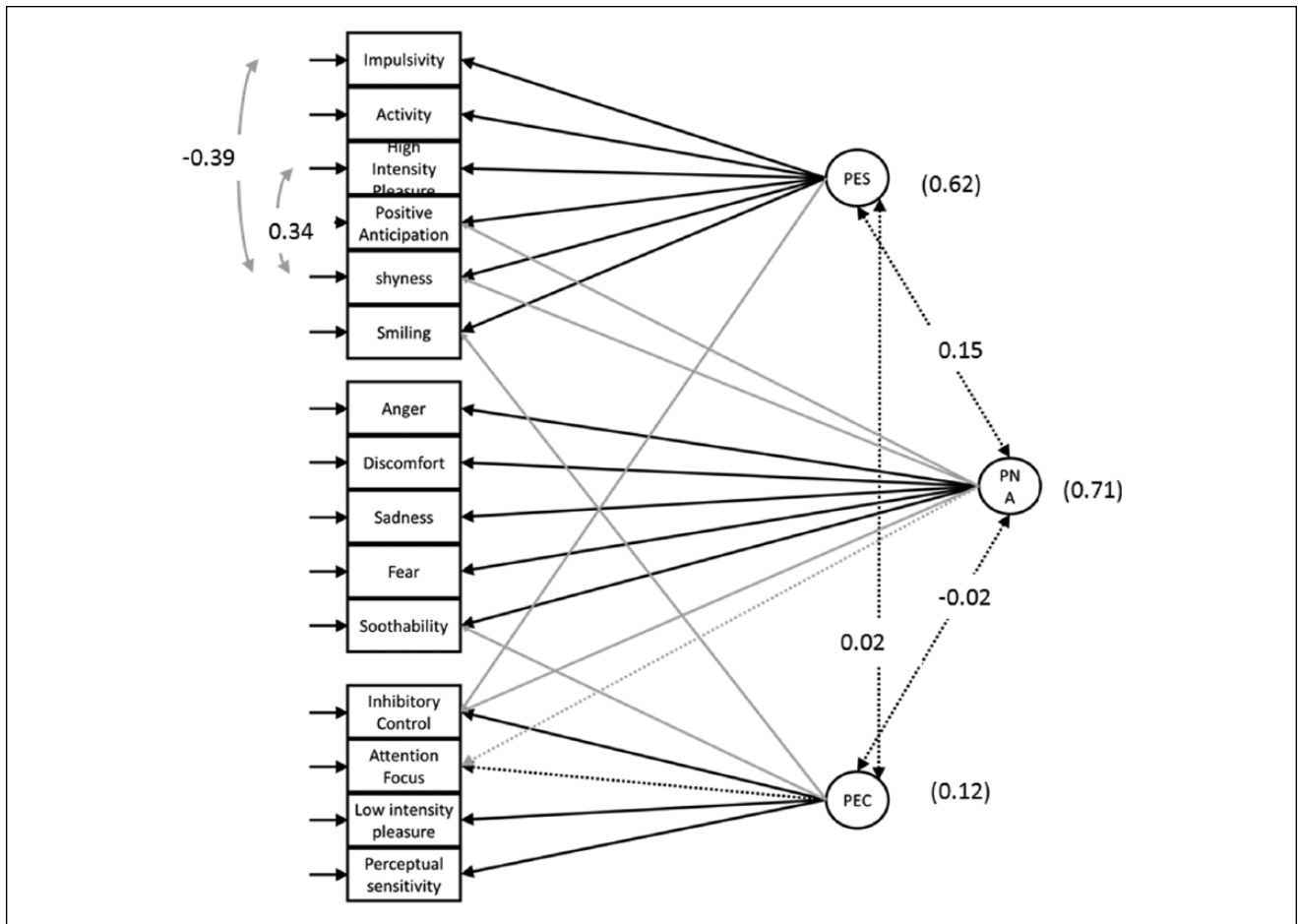
Figure 1. Model 3: Best fitting model for parent data.
Note. PES = Parent Extraversion/Surgency; PNA = Parent Negative Affectivity; PEC = Parent Effortful Control. Solid lines represent significant loadings. Dotted lines represent insignificant loadings at .05 level. Black lines represent originally assigned scales. Gray lines represent added cross-loadings. Values in parentheses are factor variances.

comparable with or higher than those reported by Putnam and Rothbart (2006), who explained that they obtained higher internal consistencies with higher SES samples. Subsequent research showed that valid results have been found when using the CBQ with children living in poverty (Richard, Davis, & Burns, 2008).

Informant Correspondences

Low parent–teacher correspondences across each of the 15 scales were anticipated based on prior research and were lower than those reported between mother–father dyads (see Putnam & Rothbart, 2006; Rothbart et al., 2001). A similar pattern of lower informant correspondence across different than similar contexts also applies to the SC questionnaire used in this study. Parent–teacher agreements on various scales of the SSIS Rating Scales (Gresham &

Elliott, 2008) were modest (on average, $r = .30$)[3] and lower than agreements between observers in the same context (teacher–teacher and parent–parent dyads, averaging $r = .58, .55$, respectively; see Gresham, Elliott, Cook, Vance, & Kettler, 2010).

Variations in parent–teacher correspondences that emerged in the current study were largely consistent with expectations based on Funder's (1995) RAM. Similar to Tackett's (2011) approach, we identified certain traits as more easily observable to parent and teacher informants because of their *relevance* to functioning in both contexts (Attention Focus and Inhibitory Control) and because of their more *overt expression* (Activity, High-Intensity Pleasure, and Impulsivity). With few exceptions, these traits were among those reaching significance across informants. For example, in kindergarten, of the six significant parent–teacher association, only one, Soothability/Falling Reactivity, was not
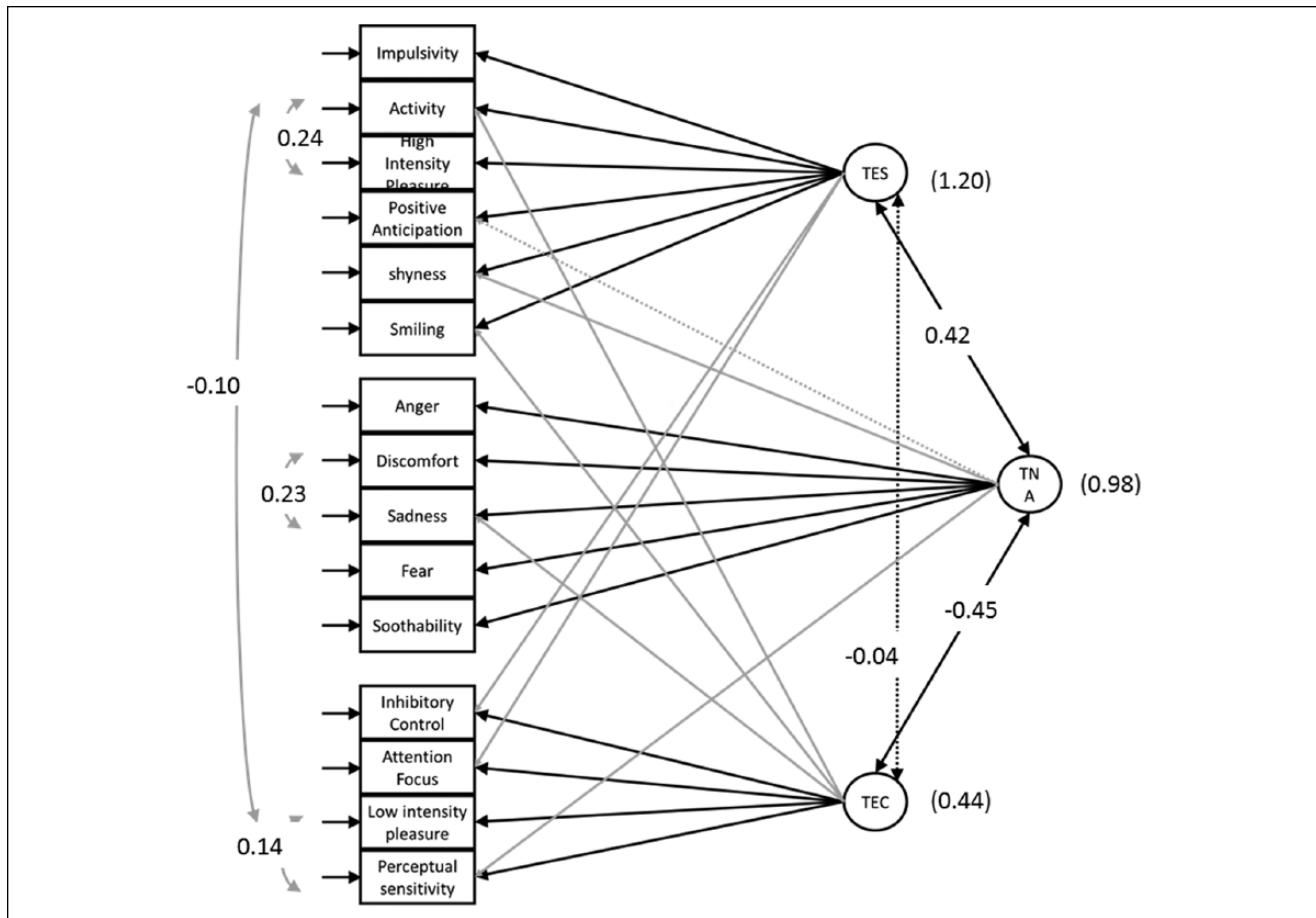
Figure 2. Model 6: Best fitting model for teacher data.
Note. TES = Teacher Extraversion/Surgency; TNA = Teacher Negative Affectivity; TEC = Teacher Effortful Control. Solid lines represent significant loadings. Dotted lines represent insignificant loadings at .05 level. Black lines represent originally assigned scales. Gray lines represent added cross-loadings. Values in parentheses are factor variances.

predicted. Although not anticipated, it stands to reason that children's ability to calm down on their own or in response to external regulation would be important to adjustment, particularly as children get older.

External Correlates

In light of consistent demonstration of low validity coefficients across different informants and different methods of measurement (Meyer et al., 2001), we examined relations with external correlates separately for each informant. Likewise, cross-cultural adaptations of measures rely on patterns of external correlates of constructs separately within each cultural milieu (see, Hambleton & Kang Lee, 2013; van de Vijver & Poortinga, 2005). Our focus was on informant-based and performance-based external correlates of the teacher questionnaire, the CBQ-TSF, but for the sake of inclusiveness, we also conducted similar analyses with the parent/caregiver questionnaire, the CBQ-SF.

Informant-Based External Correlates

Within-Informant Correlations. When teachers completed both questionnaires, strong support was found for the hypothesized link between temperament and SC, and for the most part, pattern emerging in preschool and kindergarten were similar. As expected, scales subsumed within the EC and NA scales were consistently related to SC. Moderate to high correlations emerged between SC and all four teacher-rated EC scales in preschool and kindergarten samples. With respect to NA, all five scales rated by preschool teachers were significantly associated with SC, with moderate to high effect sizes for three: Anger/Frustration, Fear, and Soothability. In kindergarten, three NA scales showed moderate to high correlations with SC: Anger/Frustration, Sadness, and Soothability.

One E/S scale, Smiling/Laughter was expected to correlate positively with SC, and this hypothesis was supported. As an indicator of positive emotionality, this scale is

Table 5. Factor Loading Estimates for Models 3 and 6.

Factor	Parent			Teacher		
	Parameter estimation	SE	p	Parameter estimation	SE	p
E/S by						
Impulsivity	I	0	N/A	I	0	N/A
Activity	0.822	0.149	.000	1.057	0.068	.000
High-Intensity	0.806	0.149	.000	0.735	0.086	.000
Approach	0.345	0.119	.004	0.638	0.069	.000
Shyness	−0.645	0.183	.000	−0.946	0.096	.000
Smiling	0.176	0.084	.036	0.516	0.087	.000
Inhibitory	−0.404	0.113	.000	−0.503	0.089	.000
Attention				−0.396	0.087	.000
NA by						
Anger	I	0	N/A	I	0	N/A
Discomfort	0.662	0.201	.001	0.599	0.118	.000
Sadness	0.539	0.129	.000	0.746	0.159	.000
Fear	0.638	0.158	.000	0.14	0.056	.012
Soothability	−0.644	0.139	.000	−0.805	0.116	.000
Approach	0.27	0.115	.019	0.043	0.078	.583
Shyness	0.547	0.155	.000	0.478	0.115	.000
Inhibitory	−0.251	0.107	.019			
Attention	−0.259	0.136	.057			
Perceptual Sensitivity				0.333	0.126	.008
EC by						
Inhibitory	I	0	N/A	I	0	N/A
Attention	0.232	0.336	.489	0.823	0.131	.000
Low-Intensity	0.935	0.285	.001	0.439	0.124	.000
Perceptual Sensitivity	1.682	0.473	.000	0.791	0.203	.000
Smiling	0.902	0.278	.001	0.858	0.156	.000
Soothability	1.742	0.467	.000			
Sadness				0.616	0.203	.002
Activity				−0.348	0.096	.000

Note. E/S = Extraversion/Surgency; NA = Negative Affectivity; EC = Effortful Control; N/A = not applicable.

relevant to children's social effectiveness, with the caveat that extreme exuberance carries negative implications as well (see Putnam, 2012). Although not predicted, Shyness had a negative correlation with SC.

Overall, findings of associations between temperament and SC within teacher informants support the construct validity of the CBQ-TSF. The scales that are most often noted in the literature as correlating with SC also emerged with significant associations within both teacher and parent informants: Attention Focus, Inhibitory Control, Anger/Frustration, Soothability, and Smiling Laughter.

Between-Informant Correlations. As anticipated, few correlations between temperament and SC emerged when parents completed one questionnaire and teachers completed the other. In the preschool sample, between-informant data were available only for parent-rated temperament and teacher-rated SC. Only Shyness showed a modest negative correlation with SC. In the kindergarten sample, teachers and parents rated both temperament and SC. However, only two temperament scales, both assigned to the EC factor, Attention Focus and Inhibitory Control, reached significance across informants but only when parents rated temperament and teachers rated SC.

The unexpected finding that between-informant relations reached significance only when parents rated temperament and teachers rated SC warrants future study. Prior research investigating relations of temperament with SC in this study's age range often rely on parent-rated temperament and teacher-rated SC (e.g., Blair, Denham, Kochanoff, & Whipple, 2004; Izard, Fine, Schultz, Ackerman, & Youngstrom, 2001; Mathieson & Banerjee, 2010).

Performance-Based External Correlates. We anticipated and found meaningful associations between teacher-rated EC and NEPSY performance, thereby garnering support for the
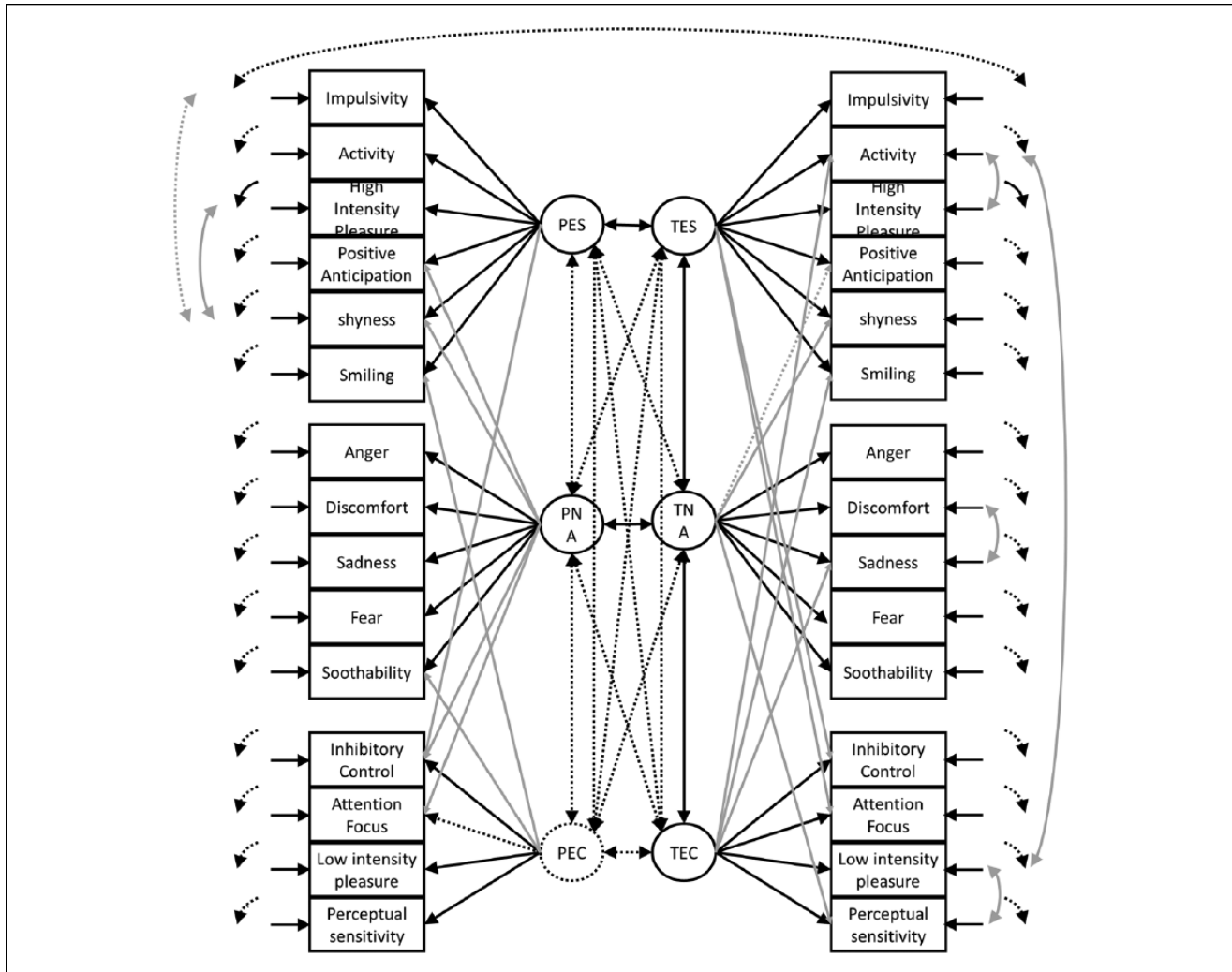
Figure 3. Model 7: Multi-group CFA model without measurement invariance assumption.
Note. CFA = confirmatory factor analysis; PES = Parent Extraversion/Surgency; PNA = Parent Negative Affectivity; PEC = Parent Effortful Control; TES = Teacher Extraversion/Surgency; TNA = Teacher Negative Affectivity; TEC = Teacher Effortful Control. Solid lines represent significant loadings. Dotted lines represent insignificant loadings at .05 level. Black lines represent originally assigned scales. Gray lines represent added cross-loadings.

construct validity of the CBQ-TSF. Across both samples, Inhibitory Control correlated with the most NEPSY tasks (all) followed by Attention Focus and then by Low-Intensity Pleasure. For parents, the pattern was similar to that of teachers in kindergarten but not preschool. Healey, Brodzinsky, Bernstein, Rabinowitz, and Halperin (2010) also reported informant-specific relations between preschool temperament and neuropsychological functioning, but they used different temperament scales, the CBQ with parents and the Temperament Assessment Battery for Children–Revised (Martin & Bridger, 1999).

Although our emphasis was on examining relations between EC and NEPSY with the CBQ-TSF, we reasoned that the expression of negative emotions, typically being less acceptable in the classroom than at home, would signal a lapse in self-regulation in the school setting that would adversely influence performance. Emerging patterns were consistent with this line of reasoning, showing that teacher-rated NA scales correlated inversely with NEPSY task performance (i.e., Anger/Frustration, Sadness, and Soothability). However, these relations between teacher-rated NA and NEPSY emerged primarily in the kindergarten sample. One explanation is that requirements for self-regulation increase from preschool to kindergarten, wherein unfettered expression of negative affect is more problematic and, in accord with RAM (Funder, 1995), more likely to be observed, detected, and used by teachers. It is interesting to note that parent-rated NA did not correlate with NEPSY but for one modest association (with fear in preschool).

Overall, the patterns of correlations of temperament with performance tasks support construct validity of the CBQ-TSF. The availability of parallel scales enables future studies to clarify the meaning of traits from each informant's perspective in reference to the adaptive requirements of contexts in which they are rated and to demands of tasks presented. For instance, differential relations of teacher- and parent-rated NA to NEPSY task performance warrant future research about the salience of emotions in the home and school contexts.

Factor Structures

A series of CFAs were conducted separately with the CBQ-TSF and the CBQ-SF scales to test various three-factor models. For each informant, models with scales assigned to factors as originally theorized did not meet requirements for a good fit, and although allowance for cross-loadings improved the model, fit indicators did not meet criteria. The model–data fit was good only when refinements suggested by the software were incorporated and these suggestions were different for parent and teacher models. In a multigroup CFA, fitting together the "good" parent and "good" teacher models without assumptions of configural invariance, the fit was moderately good. However, the fit was worse when the factor loadings of the two data sets were constrained to be the same (as the teacher data). Given that the model assuming configural invariance was worse than the unconstrained model, support was not found for configural measurement invariance.

The "good-fitting" model for the CBQ-TSF included several cross-loadings found previously in research with the parent scales (see Rothbart et al., 2001; Sleddens et al., 2011): Approach/Positive Anticipation (E/S) with NA, Shyness (E/S) with NA, Smiling/Laugher (E/S) with EC, and Inhibitory Control (EC) and Attention Focus (EC) with E/S. Three additional cross-loading scales emerged with teacher but not parent ratings, all involving connections with EC: Activity (E/S) with EC, Sadness (NA) with EC, and Perceptual Sensitivity (EC) with NA.

Departures from the theorized factor structure do not negate the utility of the questionnaire. For instance, cross-loading of Smiling/Laughter, an indicator of positive emotionality, with EC is consistent with documented associations between positive emotional states and broadened attention focus (e.g., Gable & Harmon-Jones, 2010). Likewise, cross-loadings of EC with NA may capture the self-regulatory influence of traits such as Inhibitory Control and Attention Focus in order to enable individuals to moderate their NA and override behavioral impulses to perform behaviors that are more planned and purposeful (Rothbart, Ellis, & Posner, 2004). Use of the individual scales of the CBQ-SF and CBQ-TSF permits a differentiated and comprehensive approach to measuring temperament, and the three-factor structure provides a useful organizational framework.

Although this study is limited in the small sample of children and of teachers, power was sufficient to detect meaningful patterns, consistent with the aim of this investigation. The CBQ-TSF adds to the repertoire of research tools to investigate the temperament construct. Future research with multiple informants in home and school settings may help untangle influences on temperament ratings due to context of observation and relationship with the child.

Notes

1. The senior author of the CBQ-SF reviewed the changes.
2. Additional details regarding the scales are available from the first author or may be accessed at http://www.bowdoin.edu/~sputnam/rothbart-temperament-questionnaires/
3. Parent–teacher correspondence on the SSIS in this study was also low.

References

Achenbach, T. M., McConaughy, S. H., & Howell, C. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, *101*, 213-232.

Ahadi, S. A., Rothbart, M. K., & Ye, R. (1993). Children's temperament in the U.S. and China: Similarities and differences. *European Journal of Personality*, *7*, 359-378.

Allan, S., & Gilbert, P. (2002). Anger and anger expression in relation to perceptions of social rank, entrapment, and depressive symptoms. *Personality and Individual Differences*, *32*, 551-565.

Anderson, J., & Gerbing, D. W. (1984). The effects of sampling error on convergence, improper solutions and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, *49*, 155-173.

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, *42*, 815-824.

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of child functioning at school entry. *American Psychologist*, *57*, 111-127.

Blair, K., Denham, S., Kochanoff, A., & Whipple, B. (2004). Playing it cool: Temperament, emotion regulation and social behavior in preschoolers. *Journal of School Psychology*, *42*, 419-443.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math

and literacy ability in kindergarten. *Child Development*, *78*, 647-663.

Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analysis models. In K. G. Joreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction* (Pt. 1, pp. 149-173). Amsterdam, Netherlands: North-Holland.

Bridgett, D. J., Oddi, K. B., Laake, L. M., Murdock, K. W., & Bachmann, M. N. (2013). Integrating and differentiating aspects of self-regulation: Effortful control, executive functioning, and links to negative affectivity. *Emotion*, *13*, 47-63.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Beverly Hills, CA: Sage.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum.

Copeland, W., Landry, K., Stanger, C., & Hudziak, J. J. (2004): Multi-informant assessment of temperament in children with externalizing behavior problems. *Journal of Clinical Child & Adolescent Psychology*, *33*, 547-556.

Coplan, R. J., & Bullock, A. (2012). Temperament and peer relations. In M. Zentner & R. Shiner (Eds.), *Handbook of temperament* (pp. 442-461). New York, NY: Guilford.

Corapci, F. (2008). The role of child temperament on Head Start preschoolers' social competence in the context of cumulative risk. *Journal of Applied Developmental Psychology*, *29*, 1-16.

De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, *131*, 483-509.

De Los Reyes, A., Henry, D. B., Tolan, P. H., & Wakschlag, L. S. (2009). Linking informant discrepancies to observed variations in young children's disruptive behavior. *Journal of Abnormal Child Psychology*, *37*, 637-652.

De Los Reyes, A., Thomas, S. A., Goodman, K. L., & Kundey, S. M. A. (2013). Principles underlying the use of multiple informants. *Annual Review of Clinical Psychology*, *9*, 123-149.

DeVellis, R. F. (1991). *Scale development: Theory and applications* (Applied Social Research Methods Series, Vol. 26). Newbury Park, CA: Sage.

Eid, M., & Diener, E. (2006). *Handbook of multimethod measurement in psychology*. Washington, DC: American Psychological Association.

Eisenberg, N., Fabes, R. A., Nyman, M., Bernzweig, J., & Pinuelas, A. (1994). The relations of emotionality and regulation to children's anger-related reactions. *Child Development*, *65*, 109-128.

Eisenberg, N., Morris, A. S., & Spinrad, T. L. (2005). Emotion-related regulation: The construct and its measurement. In D. Teti (Ed.), *Handbook of research methods in developmental science* (pp. 423-442). Oxford, England: Blackwell.

Eisenberg, N., Spinrad, T. L., & Morris, A. S. (2002). Regulation, resiliency, and quality of social functioning. *Self and Identity*, *1*, 121-128.

Eisenberg, N., Valiente, C., Morris, A. S., Fabes, R. A., Cumberland, A., Reiser, M., . . . Losoya, S. (2003). Longitudinal relations among parental emotional expressivity, children's regulation, and quality of socioemotional functioning. *Developmental Psychology*, *39*, 2-19.

Fagot, B. I., & O'Brien, M. (1994). Activity level in young children: Cross-age stability, situational influences, correlates with temperament, and the perception of problem behaviors. *Merrill Palmer Quarterly*, *40*, 378-398.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, *102*, 652-670.

Gable, P., & Harmon-Jones, E. (2010). The motivational dimensional model of affect; implications for breadth of memory, attention, and cognitive categorization. *Cognition and Emotion*, *24*, 322-337.

Gagne, P., & Hancock, G. R. (2006). Measurement model quality, sample size, and solution propriety in confirmatory factor models. *Multivariate Behavioral Research*, *41*, 65-83.

Gartstein, M. A., Bridgett, D. J., & Low, C. M. (2012). Asking questions about temperament: Self and other report measures across the life span. In M. Zentner & R. Shiner (Eds.), *Handbook of temperament* (pp. 442-461). New York, NY: Guilford.

Gartstein, M. A., Putnam, S. P., & Rothbart, M. K. (2012). Etiology of preschool behavior problems: Contributions of temperament attributes in early childhood. *Infant Mental Health Journal*, *33*, 197-211.

Georgas, J., van de Vijver, F. J. R., Weiss, L. G., & Saklofske, D. H. (Eds.). (2003). *Culture and children's intelligence: Cross-cultural analysis of the WISC-III*. San Diego, CA: Academic Press.

Gerardi-Caulton, G. (2000). Sensitivity to spatial conflict and the development of self-regulation in children 24-36 months of age. *Developmental Science*, *3*, 397-404.

Gresham, F. M., Cook, C. R., Vance, M. J., Elliott, S. N., & Kettler, R. (2010). Cross-informant agreement for ratings of social skill and problem behavior ratings: An investigation of the Social Skills Improvement System-Rating Scales, *Psychological Assessment*, *22*, 157-166.

Gresham, F. M., & Elliott, S. N. (2008). *Social Skills Improvement System Rating Scales*. Minneapolis, MN: Pearson Assessments.

Hambleton, R. K., & Kang Lee, M. (2013). Methods for Translating and Adapting Tests to Increase Cross-Language Validity. In D. H. Saklofske, C. R. Reynolds & V. L. Schwean (Eds.), *The Oxford Handbook of Child Psychological Assessment* (pp. 172-181).

Hartley, A. G., Zakriski, A. L., & Wright, J. C. (2011). Probing the depths of informant discrepancies: Examining contextual influences on convergence and divergence. *Journal of Clinical Child & Adolescent Psychology*, *40*, 54-66.

Healey, D. M., Brodzinsky, L. K., Bernstein, M., Rabinovitz, B., & Halperin, J. M. (2010). Moderating effects of neurocognitive abilities on the relationship between temperament and global functioning. *Child Neuropsychology*, *16*, 20-31.

Izard, C., Fine, S., Schultz, D., Mostow, A., Ackerman, B., & Youngstrom, E. (2001). Emotion knowledge as a predictor of social behavior and academic competence in children at risk. *Psychological Science*, *12*, 18-23.

Jackson, D. L. (2001). Sample size and number of parameter estimates in maximum likelihood confirmatory factor analysis:

A Monte Carlo investigation. *Structural Equation Modeling*, *8*, 205-223.

Jackson, D. L. (2003). Revisiting sample size and number of parameter estimates: Some support for the *N:q* hypothesis. *Structural Equation Modeling*, *10*, 128-141.

Kagan, J., Snidman, N., McManis, M., Woodward, S., & Hardway, C. (2002). One measure, one meaning: Multiple measures, clearer meaning. *Development and Psychopathology*, *14*, 463-475.

Korkman, M., Kirk, U., & Kemp, S. (1998). *NEPSY: A developmental neuropsychological assessment*. Chicago, IL: Psychological Corporation.

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-II*. San Antonio, TX: Pearson.

Kusanagi, E. (1993). A psychometric examination of the Children's Behavior Questionnaire. *Annual Report of the Research and Clinical Center for Child Development*, *15*, 25-33.

LaFreniere, P. J., & Dumas, J. E. (2003). *Social competence and behavior evaluation: Preschool edition*. Los Angeles, CA: Western Psychological Services.

Lemery, K. S., Essex, M. J., & Smider, N. A. (2002). Revealing the relation between temperament and behavior problem symptoms by eliminating measurement confounding: Expert ratings and factor analyses. *Child Development*, *73*, 867-882.

Lengua, L. J., West, S. G., & Sandler, I. N. (1998). Temperament as a predictor of symptomatology in children: Addressing contamination of measures. *Child Development*, *69*, 164-181.

Liew, J. (2012). Effortful control, executive functions, and education: Bringing self-regulatory and social-emotional competencies to the table. *Child Development Perspectives*, *6*, 105-111.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*, 130-149.

Majdandzic, M., & Van den Boom, D. C. (2007). Multimethod longitudinal assessment of temperament in early childhood. *Journal of Personality*, *75*, 121-168.

Mangelsdorf, S., Schoppe, S., & Buur, H. (2000). The meaning of parental reports: A contextual approach to the study of temperament and behavior problems in childhood. In D. Molfese & V. J. Molfese (Eds.), *Temperament and personality development across the lifespan* (pp. 121-140). Mahwah, NJ: Erlbaum.

Martin, R. P., & Bridger, R. (1999). *The Temperament Assessment Battery for Children – Revised*. Athens, GA: The University of Georgia.

Masten, A. S., & Coatsworth, J. D. (1998). The development of competence in favorable and unfavorable environments: Lessons from research on successful children. *American Psychologist*, *53*, 205-220.

Mathieson, K., & Banerjee, R. (2010). Pre-school peer play: The beginnings of social competence. *Educational and Child Psychology*, *27*, 9-20.

Mervielde, I., & De Pauw, S. S. W. (2012). Models of child temperament. In M. Zentner & R. L. Shiner (Eds.), *Handbook of temperament* (pp. 21-40). New York, NY: Guilford.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., . . .Reed, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, *56*, 128-165.

Mischel, W. (2004). Toward an integrative science of the person. *Annual Review of Psychology*, *55*, 1-22.

Moran, L. R., Lengua, L. J., & Zalewski, M. (2013). The interaction between negative emotionality and effortful control in early social-emotional development. *Social Development*, *22*, 340-362.

Muthén, L. K., & Muthén, B. O. (1998-2011). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.

Olson, S. L., Sameroff, A. J., Lunkenheimer, E. S., & Kerr, D. C. R. (2009). Self-regulator processes in the development of early disruptive behavior: The preschool to school transition. In S. L. Olson & A. L. Sameroff (Eds.), *Biopsychosocial regulatory processes in the development of childhood behavior problems*. New York, NY: Cambridge University Press.

Putnam, S. P. (2012). Positive emotionality. In M. Zentner & R. L. Shiner (Eds.), *Handbook of temperament* (pp. 105-123). New York, NY: Guilford.

Putnam, S. P., & Rothbart, M. K. (2006). Development of short and very short forms of the Children's Behavior Questionnaire. *Journal of Personality Assessment*, *87*, 103-113.

Raver, C. C. (2002). Emotions matter: Making the case for the role of young children's emotional development for early school readiness. *Society for Research in Child Development*, *16*, 1-20.

Rettew, D. C., Van Oort, F., Verhulst, F. C., Buitelaar, J. K., Ormel, J., Hartman, C. A., . . .Hudziak, J. J. (2011). When parent and teacher ratings don't agree: The Tracking Adolescents' Individual Lives Survey (TRAILS). *Journal of Child and Adolescent Psychopharmacology*, *21*, 389-397.

Reynolds, C. R., & Carson, A. D. (2005). Methods for assessing cultural bias in tests. In C. L. Frisby & C. R. Reynolds (Eds.), *Comprehensive handbook of multicultural school psychology* (pp. 795-823). New York, NY: Wiley.

Richard, H., Davis, D. W., & Burns, B. M. (2008). An evaluation of the "Children's Behavior Questionnaire" (CBQ) for use with children from low-income families. *Journal of Early Childhood and Infant Psychology*, *4*, 111-124.

Rothbart, M. K., Ahadi, S. A., Hershey, K., & Fisher, P. (2001). Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. *Child Development*, *72*, 1394-1408.

Rothbart, M. K., & Bates, J. E. (2006). Temperament. In W. Damon & R. M. Lerner (Series Eds.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 99-166). New York, NY: Wiley.

Rothbart, M., & Derryberry, D. (1981). Development of individual differences in temperament. In M. Lamb & A. Brown (Eds.), *Advances in developmental psychology* (Vol. 1, pp. 37-86). Hillsdale, NJ: Erlbaum.

Rothbart, M. K., Ellis, L. K., Rueda, M. R., & Posner, M. I. (2003). Developing mechanisms of temperamental effortful control. *Journal of Personality*, *71*, 1113-1143.

Rothbart, M. K., Ellis, L. K., & Posner, M. I. (2004). Temperament and self-regulation. In R. F. Baumeister & K. D. Vohs (Eds.), *Handbook of self-regulation: Research, theory, and applications* (pp. 357-370). New York: Guilford Press.

Rueda, M. R., Posner, M. I., & Rothbart, M. K. (2011). Attentional control and self-regulation. In R. F. Baumeister & K. D. Vohs

(Eds.), *Handbook of self regulation: Research, theory, and applications* (2nd ed., pp. 284-299). New York, NY: Guilford.

Schaughency, E. A., & Fagot, B. I. (1993). The prediction of adjustment at age 7 from activity level at age 5. *Journal of Abnormal Child Psychology*, *21*, 29-50.

Simonds, J., Kieras, J. E., Rueda, M. R., & Rothbart, M. K. (2007). Effortful control, executive attention, and emotional regulation in 7–10-year old children. *Cognitive Development*, *22*, 474-488.

Sleddens, E. F., Kremers, S. P., Candel, M. J., De Vries, N. N., & Thijs, C. (2011). Validating the Children's Behavior Questionnaire in Dutch children: Psychometric properties and a cross-cultural comparison of factor structures. *Psychological Assessment*, *23*, 417-426.

Spinrad, T. L., Eisenberg, N., Cumberland, A., Fabes, R. A., Valiente, C., Shepard, S. A., . . . Guthrie, I. K. (2006). Relation of emotion-related regulation to children's social competence: A longitudinal study. *Emotion*, *6*, 498-510.

Tackett, J. L. (2011). Parent informants for child personality: Agreement, discrepancies, and clinical utility. *Journal of Personality Assessment*, *93*, 539-544.

van de Vijver, F., & Poortinga, Y. (2005). Conceptual and methodological issues in adapting tests. In R. K. Hambleton, P. F. Merenda & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 39–64). Mahwah, NJ : Lawrence Erlbaum .

Vohs, K. D., & Baumeister, R. F. (Eds.). (2011). *Handbook of self-regulation: Research, theory, and applications* (2nd ed.). New York, NY: Guilford.

Webster-Stratton, C., & Reid, M. J. (2004). Strengthening social and emotional competence in children— The foundation for early school readiness and success: Incredible years classroom social skills and problem solving curriculum. *Infants & Young Children*, *17*, 96-113.

Wilkowski, B. M., & Robinson, M. D. (2010). The anatomy of anger: An integrative cognitive model of trait anger and reactive aggression. *Journal of Personality*, *78*, 9-38.

Zentner, M., & Bates, J. E. (2008). Child temperament: An integrative review of concepts, research programs, and measures. *European Journal of Developmental Science*, *2*, 7-37.

Zhou, Q., Main, A., & Wang, Y. (2010). The relations of temperamental effortful control and anger frustration to Chinese children's academic achievement and social adjustment: A longitudinal study. *Journal of Educational Psychology*, *102*, 180-196.