

Thin Slices of Child Personality: Perceptual, Situational, and Behavioral Contributions

Jennifer L. Tackett and Kathrin Herzhoff
University of Houston

Shauna C. Kushner and Nicholas Rule
University of Toronto

The present study examined whether thin-slice ratings of child personality serve as a resource-efficient and theoretically valid measurement of child personality traits. We extended theoretical work on the observability, perceptual accuracy, and situational consistency of childhood personality traits by examining intersource and interjudge agreement, cross-situational consistency, and convergent, divergent, and predictive validity of thin-slice ratings. Forty-five unacquainted independent coders rated 326 children's (ages 8–12) personality in 1 of 15 thin-slice behavioral scenarios (i.e., 3 raters per slice, for over 14,000 independent thin-slice ratings). Mothers, fathers, and children rated children's personality, psychopathology, and competence. We found robust evidence for correlations between thin-slice and mother/father ratings of child personality, within- and across-task consistency of thin-slice ratings, and convergent and divergent validity with psychopathology and competence. Surprisingly, thin-slice ratings were more consistent across situations in this child sample than previously found for adults. Taken together, these results suggest that thin slices are a valid and reliable measure to assess child personality, offering a useful method of measurement beyond questionnaires, helping to address novel questions of personality perception and consistency in childhood.

Keywords: child personality, thin slice, personality assessment, cross-situation consistency, person perception

Personality psychologists frequently emphasize the inherent power in personality traits for predicting a wide variety of consequential outcomes, including physical and mental health, occupational status and success, relationship outcomes, and well-being (Ozer & Benet-Martínez, 2006; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007). Further, we also know that such prediction emerges even from early individual differences in childhood dispositions (Caspi, Roberts, & Shiner, 2005; Hampson, 2008; Shiner, Masten, & Roberts, 2003). Childhood personality traits predict outcomes across a variety of domains, including parenting (Van den Akker, Deković, Asscher, & Prinzie, 2014), health behaviors (Hampson, 2008; Slobodskaya, 2011), and academic success (Shiner, 2000). Yet, research on child personality lags far behind work in adults, partly due to unique measurement challenges that accompany the study of personality at a young age. Beyond purely practical concerns, these measurement challenges raise critical substantive questions about the perception and accuracy of child personality ratings and environmental/situational influences on child personality expression.

These interconnected issues—the lack of both *measurement* and *theoretical* knowledge regarding child personality—highlight the fundamental need for construct validation efforts (Cronbach & Meehl, 1955). Efforts to advance sound measurement should thus be guided by existing theory and inform theoretical knowledge, such that measurement and theory are inextricably linked. These represent the primary goals of the current study, which aimed to explore thin-slice (TS) perceptions of child personality traits as a substantive (i.e., theoretically grounded and relevant) and practical (i.e., psychometrically valid and reliable) tool for examining child personality.

Measuring Personality in Childhood

Two major approaches to measuring childhood individual differences have dominated the literature: questionnaires, which are typically completed by parents or teachers, and structured laboratory observational paradigms (Rothbart & Bates, 2006). We summarize some of the advantages and limitations of these approaches, although they have been fairly extensively reviewed and discussed elsewhere (Durbin, 2010; Rothbart & Bates, 2006).

Questionnaires are the standard approach to personality assessment, including research with adults. Personality questionnaires are generally accepted as valid and reliable self-report methods and are typically regarded as the standard in adult work (Vazire, 2006). Self-reports are more challenging to collect from children, however, and when they are collected from older children, they tend to suffer from psychometric limitations (Soto, John, Gosling, & Potter, 2008) and frequently require more intensive data collection efforts (e.g., assistance in comprehending items, potentially using an interview-based format; e.g., Brown, Mangelsdorf, Aga-

This article was published Online First May 4, 2015.

Jennifer L. Tackett and Kathrin Herzhoff, Department of Psychology, University of Houston; Shauna C. Kushner and Nicholas Rule, Department of Psychology, University of Toronto.

Support for this research was provided by funding from the Social Sciences and Humanities Research Council (SSHRC) of Canada awarded to Dr. Jennifer L. Tackett.

Correspondence concerning this article should be addressed to Jennifer L. Tackett, Department of Psychology, University of Houston, 126 Heyne Building, Houston, TX 77204. E-mail: jlackett@uh.edu or jenniferlackett@gmail.com

then, & Ho, 2008; Eder, 1990; Markey, Markey, Tinsley, & Eriksen, 2002; Measelle, John, Ablow, Cowan, & Cowan, 2005). Thus, personality questionnaires for assessing children are most often completed by parents (Rothbart & Bates, 2006; Tackett, 2011).

Although indirect reports from parents carry with them limitations, parents do possess characteristics of “good judges” such that they are likely motivated to provide thoughtful responses about their child and are highly familiar with their child, both of which should increase rating accuracy (Funder, 1995; Tackett, 2011). In addition, mothers and fathers show moderately high agreement regarding their child’s personality traits (Mullineaux, Deater-Deckard, Petrill, Thompson, & DeThorne, 2009; Tackett, 2011) and show good concurrent and predictive validity for a variety of constructs (Brown et al., 2008; De Pauw, Mervielde, & Van Leeuwen, 2009; Rothbart & Bates, 2006; Tackett, Krueger, Iacono, & McGue, 2008). Although questionnaire-based assessment of child personality has numerous advantages (highly knowledgeable informants when using parents, inexpensive and quick administration, and the ability to assess multiple traits simultaneously), there are drawbacks as well. Parents are most likely biased in rating their child (Rothbart & Bates, 2006; Tackett, 2011) and their ratings are likely to be a function of their own characteristics (Forman et al., 2003; Kagan, 1998).

An alternative approach to assessing early dispositions is the use of standardized laboratory situations to elicit specific traits in children, which are typically video-recorded and later scored for the constructs of interest (Durbin, 2010; Goldsmith, Reilly, Lemery, Longley, & Prescott, 1993; Kochanska, Barry, Jimenez, Hollatz, & Woodard, 2009; Kochanska, Murray, & Harlan, 2000; Rothbart & Bates, 2006). Such paradigms allow the researcher a high degree of control in designing the context and in attempts to specify trait elicitation. Moreover, they counteract some of the limitations of parent-based questionnaires, such as informant bias. There are limits to this approach as well, perhaps most notably the extensive resources often involved in collecting and coding these data (Prime, Perlman, Tackett, & Jenkins, 2014). Often, situations are coded primarily for the trait of interest, requiring a variety of tasks to elicit multiple traits. Other concerns include difficulty in rating low frequency characteristics, questions about accuracy and bias—particularly when using a single coder—and the impact of the novelty of the laboratory setting (Rothbart & Bates, 2006). Thus, both parent-based questionnaires and laboratory measures of child personality have advantages and disadvantages spanning theoretical and practical domains. Given these limitations, scholars have stressed that neither is appropriate for use as the sole measure of early personality (Rothbart & Bates, 2006); thus, the field of child personality remains in need of better measurement alternatives and more nuanced understanding of these commonly used approaches.

Another relevant aspect of child personality measurement is interinformant and intersource consensus and discrepancy. Associations across different informants and sources have been of interest to both adult-based and child-based personality researchers for some time (Rothbart & Bates, 2006; Tackett, 2011; Vazire, 2010). Informant consensus and discrepancy may offer important theoretical insight to the processes underlying person perception (Borkenau & Liebler, 1993; Kenny, Albright, Malloy, & Kashy, 1994), highlight meaningful characteristics of both targets and

judges (Funder, 1995; Letzring, Wells, & Funder, 2006), and facilitate substantive explanations for difficult-to-rate or highly variable individuals (De Los Reyes, Alfano, & Beidel, 2010). In addition to enhancing substantive understanding, intersource agreement is also an important aspect of establishing the validity and reliability of any assessment tool for child personality (Forman et al., 2003). For example, data from observational lab tasks showed that consensus with parental informants was higher when ratings were aggregated across multiple tasks (i.e., composite ratings), rather than relying on ratings for a single task (i.e., task-specific ratings; Forman et al., 2003). Thus, the current research sought to inform both theoretical and applied issues in examining indices of intersource agreement.

Thin Slices of Behavior

Over the past several decades, a rich literature has emerged on the inherent power of “snap judgments”—or, the impressive information contained in one’s initial impressions of others. For example, perceptions based on “thin slices” of another’s behavior predict many consequential constructs at greater than chance levels, including end-of-term teaching evaluations (Ambady & Rosenthal, 1993), sexual orientation (Rule & Ambady, 2008; Rule, Ambady, & Hallett, 2009), income (Christopher & Schlenker, 2000), religious beliefs (Rule, Garrett, & Ambady, 2010), and adjustment following divorce (Mason, Sbarra, & Mehl, 2010). We also know, from pioneering work conducted by Borkenau and Liebler (1992), that others’ perceptions of adult personality show impressive convergence with self-reports (see also Borkenau, Mauer, Riemann, Spinath, & Angleitner, 2004). TS impressions of others’ personalities have revealed that narcissists tend to be popular on first impression (Back, Schmukle, & Egloff, 2010) and that extraversion (E) can be reliably judged from a 50-ms glimpse of a person’s photo (Borkenau, Brecke, Mötting, & Paelecke, 2009). In contrast, people with personality pathology have been identified as less likable and less extraverted after only 30 s (Oltmanns, Friedman, Fiedler, & Turkheimer, 2004). These studies demonstrate how quickly individuals can process information and form accurate impressions of personality. With respect to differential accuracy across traits, there is substantial evidence that self-other agreement on adult personality is highest for highly observable traits (i.e., E) and lowest for traits that are harder to observe (i.e., neuroticism [N]; Borkenau et al., 2009, 2004; Hall, Andrzejewski, Murphy, Mast, & Feinstein, 2008; Hirschmüller, Egloff, Schmukle, Nestler, & Back, 2014; Vazire, 2006, 2010; Vazire & Gosling, 2004; Vazire, Naumann, Rentfrow, & Gosling, 2008). TS ratings thus appear to capture meaningful variation in characterological traits on the basis of limited (even entirely nonverbal) information.

How do we understand the role of unacquainted informants, as used in TS approaches, relative to other types of informants for personality? Considerable research has shown that self-other agreement for adult personality tends to be weaker for unacquainted observers than for close acquaintances (Borkenau et al., 2009; Connolly, Kavanagh, & Viswesvaran, 2007; de Vries, Lee, & Ashton, 2008; Funder & Colvin, 1988; Funder, Kolar, & Blackman, 1995; Hall et al., 2008; Vazire, 2010; Vazire & Gosling, 2004). Adult personality research also shows that self-other agreement for personality appears to increase over time (Kurtz &

Sherker, 2003; Paulhus & Bruce, 1992). Among youth, informant ratings tend to be more highly correlated for individuals who have more opportunities to observe children in the same contexts (e.g., mothers and fathers) than informants largely exposed to different contexts (e.g., parents and teachers). For example, mother–father agreement tends to be higher than parent–teacher agreement for ratings on children’s social skills (Fagan & Fantuzzo, 1999), psychopathology (Garrison & Earls, 1985; Lindholm & Touliatos, 1982; Stanger & Lewis, 1993), and personality traits (Laidra, Allik, Harro, Merenäkk, & Harro, 2006). Moreover, interparent agreement for child social skills tends to be higher among parents who share child rearing duties than for parent dyads where the mother is primarily responsible for child rearing (Fagan & Fantuzzo, 1999). Disagreement among informants is also important, as informant discrepancies may reflect psychologically meaningful constructs like disrupted communication and family conflict (De Los Reyes, Goodman, Kliewer, & Reid-Quiñones, 2008; De Los Reyes et al., 2012; Gaylord, Kitzmann, & Lockwood, 2003; Grills & Ollendick, 2002; Guion, Mrug, & Windle, 2009). Taken together, these findings suggest that—in addition to well-acquainted raters such as parents and teachers—unacquainted observers exposed to small slices of an individual’s behavior may represent an independent and valid source of information about a person’s personality characteristics.

The complexity of trait structures based on TS ratings might offer valuable clues to further understand potential sources of interjudge agreement and examine the validity of TS ratings of child personality. Personality traits in both adulthood and childhood are significantly intercorrelated and hierarchically related to one another (Markon, Krueger, & Watson, 2005; Tackett et al., 2012). Research on zero-acquaintance ratings has shown that these ratings usually give rise to a simpler trait structure (i.e., higher trait intercorrelations) than self-ratings do (Beer & Watson, 2008a, 2008b; Rauthmann, 2013; Rauthmann & Kolar, 2010; Wood, Harms, & Vazire, 2010). This more simplified trait structure may reflect the extent to which people use implicit personality theories when rating strangers’ personalities, which might also serve to increase their interjudge agreement. An exception to this general pattern of higher trait intercorrelations in zero-acquaintance (compared to self-ratings) is trait E, which may more readily provide explicit cues to guide trait ratings (Beer & Watson, 2008a, 2008b). Based on this research, we would expect TS ratings to show higher intercorrelations among traits than parent ratings, with E as a potential exception to this general pattern.

Although TS assessments clearly reflect a wealth of information about personality, researchers have not yet harnessed this information as an opportunity to assess personality in situations where no standard exists, such as personality in childhood. This represents the goal of the current study.

Cross-Situational Consistency

A relevant concern for observation-based assessments of personality traits, such as the TS approach, is the extent to which behaviors are consistent across situations (i.e., cross-situational consistency). That is, how much of an observed behavior can be attributed to underlying personality traits that remain consistent across situations? Prior research has shown that different raters’ agreement on children’s observable behaviors may depend on the

degree of match between the observational settings and the system used to code behaviors (Charlebois, Tremblay, Gagnon, Larivée, & Laurent, 1989). With few exceptions (e.g., Roberts & DelVecchio, 2000; Zupančič, Sočan, & Kavčič, 2009), there is a dearth of empirical literature on the consistency of child personality ratings across time, situations, or informants. Accordingly, the consistency of ratings of child personality across unique situations is not well understood. Among adults, prior tests have estimated consistent cross-situational impressions of personality at $r = .43$ to $.46$ (Borkenau et al., 2004; Funder & Colvin, 1991). On a larger macroscale, personality consistency estimated over longer periods of time (e.g., months and years) increases across childhood and into adulthood (Roberts & DelVecchio, 2000; Zupančič et al., 2009). Presently, research is missing from the broader literature that directly examines the consistency of child personality across situations on a microscale within a circumscribed period of time.

Moreover, evidence for cross-situational consistency of personality may depend upon the particular trait or behavior being observed. In early childhood, the consistency of inhibited behaviors appears to be positively associated with trait fearfulness (Rubin, Hastings, Stewart, Henderson, & Chen, 1997). This is consistent with research on young adults, whose cross-situational consistency positively associated with traits related to N (Compas, Forsythe, & Wagner, 1988), and negatively associated with reverse-coded facets of N (e.g., calm, relaxed; Sherman, Nave, & Funder, 2010). In research on adult TS ratings, cross-situational consistency was highest for E (.51), N (.46), and agreeableness ([A], .45), and lowest for conscientiousness ([C], .36) and openness ([O], .35; Borkenau et al., 2004). Although these combined findings provide reasonable grounds for assuming that consistency may vary across traits, cross-situational consistency for the full range of child personality traits has yet to be investigated. Thus, the extent to which such trait-specific variations generalize to children remains an open question.

One concern about the use of TS ratings for child personality is the potential bias or behavioral constraint that certain situations may impose on trait expression, thus influencing behavioral consistency (Rothbart & Bates, 2006). For example, it is important to examine the relevance of particular situational contexts to best measure given traits, as has often been the tradition in dispositional measurements of children (Durbin, 2010; McLarney-Vesotski, Bernieri, & Rempala, 2006), with similar ideas proposed in adult personality research (Hirschmüller et al., 2014). Furthermore, researchers have raised concerns over whether children behave in these settings as they would in natural, “real life” situations (Hayden, Klein, & Durbin, 2005). That is, unfamiliarity within the laboratory context may cause children to behave in an artificial manner, thereby not expressing their true personalities. These raise important theoretical and applied questions concerning the use of TS ratings of child personality, which we address in the current study. The present research will thus examine whether TS ratings become more similar to parent ratings as duration in the research protocol increased.

The Present Study

The present study sought to use TS personality ratings to examine theoretical questions regarding the observability, perceptual accuracy, and situational consistency of childhood personality

traits. In addition to theoretical contributions, this research aimed to investigate the characteristics of a TS approach to assessment of child personality as a novel and useful assessment tool that complements other approaches by addressing some of their pressing limitations, and that is easy, efficient, and cost-effective to implement in research, educational, and clinical settings. Specifically, the approach draws on a *construct validation* framework to pursue theoretical and measurement questions in tandem (Cronbach & Meehl, 1955). We do so using data from a sample of 326 children, exposed to 15 different situations, each of which was rated by three unacquainted observers.

We examined the following aspects of TS assessment of child personality:

1. Intersource agreement: How does TS assessment of child personality correspond with parents' ratings of their child's personality? Given that parents know their children well, have extremely high acquaintanceship with their child, are motivated to respond accurately, and share relational features with one another (as parents to the child; Tackett, 2011), we would expect interparent agreement to be substantially higher than parent-TS agreement.
2. Interjudge agreement: How well do unacquainted observers agree in their ratings of a child's personality? We examined evidence for agreement across any two judges, agreement across any three judges (or, the internal consistency of all three judges rating a single slice), and agreement across different judges viewing different tasks involving the same target. To further investigate aspects of interjudge agreement, we also compared trait intercorrelations for mother reports with those for TS ratings, hypothesizing that TS trait intercorrelations should be higher (Beer & Watson, 2008a, 2008b).
3. Cross-situational consistency: How consistent are ratings of a child's personality across 15 unique situations? We compared these estimates to those derived from adult participants in Borkenau, Mauer, Riemann, Spinath, and Angleitner's (2004) study to examine potential developmental differences in cross-situational consistency, with the expectation that cross-situational consistency in personality would be higher for adults than for children. Does behavioral expression of child personality become more accurate as they adjust to novel environments? We examined the extent to which TS ratings of traits grew more similar to mean-parent trait ratings later in the research protocol (i.e., as the child had been in the lab environment for a few hours).
4. Convergent/divergent validity: Are TS ratings of child personality associated with behavioral criteria in theoretically predicted ways? To further examine evidence for the construct validity of a TS approach, we also investigated associations with behavioral problems and competencies. General predictions were consistent with existing research on childhood personality-psychopathology asso-

ciations (Tackett, 2006): higher levels on N and lower levels on E would be associated with internalizing problems; higher levels on N and lower levels on A and C would be associated with externalizing problems. Less existing literature supports hypotheses for behavioral competencies, but we generally expected lower levels on N and higher levels on E, A, C, and O to be associated with positive adjustment.

Method

Participants

Participants were 326 youth (168 girls) and their parents. Youth primarily aged from 9 to 10 years ($M = 9.96$, $SD = 0.81$) and their parents were recruited using a community-based participant pool and flyers posted throughout the community. Inclusion criteria were English fluency in both the parent and child, and exclusion criteria were the presence of neurodevelopmental disorders, psychotic disorders, or intellectual disability in the child. We obtained informed consent from the parents and verbal assent from the youth. Parents reported the following ethnicity breakdown for the target children: 71.5% Caucasian, 7.7% Asian-Canadian, 3.1% African-Canadian, 0.6% Hispanic, 0.3% Pacific Islander, 16.6% other/multiracial, and 0.3% not reporting ethnicity. We coded parent occupation using the Hollingshead (1975) Occupational Scale, a 9-point scale wherein higher scores indicate a higher level of occupational attainment (here, $M = 7.04$, $SD = 1.48$; $range = 3-9$). We obtained additional demographic information for 75.2% of the sample during a follow-up wave of the study (approximately three years after intake). The following breakdown for marital status was reported by the primary caregiver: 87.3% married/living with a partner, 8.2% separated/divorced, 2.4% widowed, 1.2% never married, and 0.8% not reporting marital status.

Measures

Inventory of Children's Individual Differences—Short Form (ICID-S; Deal, Halverson, Martin, Victor, & Baker, 2007). We assessed child personality using informant (parent and TS) and youth self-report versions of the ICID-S, a 50-item questionnaire measuring the Big Five traits in children: neuroticism (N), extraversion (E), openness-to-experience (O), agreeableness (A), and conscientiousness (C). Items were rated on a 7-point scale (1 = *much less than the average child or not at all* and 7 = *much more than the average child*). For participating parents, the 50 items comprising ICID-S scores were drawn from the 144-item long-form version of the ICID (Halverson et al., 2003) to create identical scales across informants (only the short-form was administered to TS raters and youth). Facet allocation to higher order traits corresponded to the structure reported by Halverson et al. (2003) with one exception: the openness facet was used to compute the higher-order O score instead of the higher-order E score. Internal consistency ranged from Cronbach's alpha = .82 (A) to .89 (E and C) for mother report, .82 (N) to .88 (E and C) for father

report, .78 (O) and .89 (E) for youth report, and .86 (O) to .93 (C) for TS reports.

Child Behavior Checklist (CBCL, 6–18; Achenbach & Rescorla, 2001). Childhood psychopathology and competence was assessed using the CBCL, a 118-item questionnaire measuring the behavioral and emotional problems of youth aged 6–18 years. Items representing specific problems in the past 6 months were rated on a 3-point scale (0 = *not true as far as you know* and 2 = *very true or often true*). Additional items measure the amount and quality of children's involvement in extracurricular, social, and academic activities. The CBCL is scored to generate dimensional composite scores for internalizing problems (Cronbach's alpha = .83), externalizing problems (Cronbach's alpha = .87), total problems (Cronbach's alpha = .94), and total competencies, as well as specific syndrome and competency scales that have demonstrated good psychometric properties (Achenbach & Rescorla, 2001). The internalizing problems scale assesses problems with anxiety, depression, and somatic complaints, whereas the externalizing problems scale assesses problems with rule-breaking and aggression. In this investigation, we examined mean parent-reported CBCL scores.

Computerized Diagnostic Interview Schedule for Children (CDISC; Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000). We also assessed childhood psychopathology using the CDISC, a *DSM-IV*-based structured clinical interview. Symptoms were rated as absent (0) or present (1) over the last year. We scored the CDISC to generate dimensional composite scores for internalizing disorders and externalizing disorders so as to allow direct comparison with CBCL scales. The internalizing disorder score included symptoms of social phobia, separation anxiety, and major depressive disorder, whereas the externalizing disorder score included symptoms of attention deficit/hyperactivity disorder, oppositional defiant disorder, and conduct disorder. In this investigation, we examined both parent- and child-reported CDISC scores. Interrater reliability based on 21 interviews ranged from ICC = .91 (symptom counts during the last year for oppositional defiant disorder) to 1.00 (all other disorders).

Videotaped sequences. Children were videotaped performing 15 assigned tasks. We later partitioned these videos into 15 seg-

ments respective to each task and presented each segment to three judges who did not meet the target child, resulting in 45 independent TS ratings per child participant. We selected tasks based on the likeliness that they would elicit observable differences in child personality traits and in line with prior research (see Table 1; Borkenau et al., 2004; Kochanska et al., 2000).

Procedure

Data for the present investigation were collected at the intake phase of a longitudinal study examining the role of personality traits in predicting behavioral outcomes in the Personality Across Development Laboratory at the University of Toronto. Ethics approval for this investigation was obtained from the institutional review board. Packages including informed consent documentation, the ICID, and CBCL were mailed to two participating caregivers per child to be completed and returned upon arrival at their in-lab testing session. Adult caregivers (313 mothers and 215 fathers) were instructed to complete the questionnaires independently. Families received up to \$40 CAD and children received two small gifts for completing the full 2-hr protocol. At the lab visit, youth were administered the 15 videotaped tasks described in Table 1 to create TS videos and both youth and one parent completed the structured clinical interview. Data collection from a second caregiver (88% fathers) and inclusion of the CBCL was added to the study protocol partway after data collection had begun, resulting in some missing data for 124 children. A subsample of participants (76.7%; $n = 250$) completed a subsequent follow-up wave of the study approximately three years later, when participating youth were 12.65 ($SD = 0.86$) years old on average. During this wave, packages including informed consent documentation and parent and youth self-report versions of the ICID-S were mailed to participating youth and one parent, to be completed and returned by mail. For the current analyses, missing data were imputed using the expectation-maximization (EM) algorithm in SPSS 21.

TS videos were not further processed (i.e., judges saw the entire length of the clips with sound). In total, we obtained 14,043 TS observations (i.e., 326 participants \times 45 observations, minus miss-

Table 1
List of Thin-Slice Tasks

Task name	Description	Duration
1. Demographic questionnaire	Answer basic questions about oneself, including age, date of birth, year in school, favorite/least favorite subject, name of best friend, name of siblings.	1–2 min
2. Red-green sign	Perform a task while keeping 1–2 rules in mind (i.e., Go-No Go task).	2–3 min
3. Tower task	Take turns removing blocks from a tower without knocking it down.	~5 min
4. Digit span	Repeat digits of increasing length as presented and in reverse order.	2–4 min
5. Whisper task	Whisper the names of recognizable characters in pictures.	1–2 min
6. Introduce yourself	Pretend to introduce yourself to a child your age for the first time.	3–4 min
7. Stories	Tell a complete story based on pictures from Bellak and Abrams' (1997) Children's Apperception Test (CAT) picture cards.	~12 min
8. Best memory	Recall your best memory.	~3 min
9. Spelling	Oral spelling test.	1–2 min
10. Sad memory	Recall a sad memory.	~3 min
11. Definitions	Define neologisms aloud.	1–2 min
12. Math	Solve math problems aloud.	2–3 min
13. Favorite song	Sing your favorite song.	~3 min
14. Uses for paper	Describe multiple uses for a sheet of paper.	~3 min
15. Disappointing gift task	Receiving a desirable followed by a disappointing gift.	5–10 min

ing data¹). Individuals working in our lab made approximately 91.5% of the ratings most of whom were undergraduate volunteer research assistants (RAs, $n = 118$), and 8.5% of ratings were completed by psychology undergraduates ($n = 356$) who completed ratings for additional course credit (total judges $N = 474$). All judges completed a brief orientation to the TS approach and rated one practice video prior to completing ratings for their assigned slices. Our original protocol allowed judges to watch a maximum of three TS videos for each target child with a minimum of one month between each rating; however, we later revised this protocol to be more consistent with Borkenau et al.'s (2004) study by limiting judges to watch only one TS of each target child. Approximately 25% of the slices were rated by judges who had already watched a previous slice of the same child. The exact number of TS videos that each judge rated varied between 1 and 517, although psychology undergraduates receiving course credit generally watched four TS videos. The type of TS task (e.g., stories, spelling) rated by each judge was not restricted.

Results

Descriptive Statistics

Descriptive statistics for mother, father, and TS ratings of child personality traits are displayed in Table 2. We conducted paired-samples t tests to compare scores across source with each child as the unit of analysis and estimated effect sizes using Hedge's g . On average, fathers' ratings of A were lower than mothers' ratings, $t(325) = -2.01$, $p = .045$, $g = 0.11$, 95% CI [0.04, 0.19], and strangers' TS ratings, $t(325) = -3.18$, $p = .002$, $g = 0.22$, 95% CI [0.16, 0.29]. TS ratings for N were higher than both mothers', $t(325) = 2.04$, $p = .042$, $g = 0.13$, 95% CI [0.07, 0.19], and fathers', $t(325) = 2.68$, $p = .008$, $g = 0.19$, 95% CI [0.14, 0.24], ratings. In contrast, TS ratings of E and O were lower than mothers' ratings, $t(325) = -19.66$, $p < .001$, $g = -1.30$, 95% CI [-1.35, -1.24], and $t(325) = -23.20$, $p < .001$, $g = -1.43$, 95% CI [-1.49, -1.38], respectively; or fathers' ratings, $t(325) = -18.61$, $p < .001$, $g = -1.31$, 95% CI [-1.36, -1.26], and $t(325) = -21.48$, $p < .001$, $g = -1.48$, 95% CI [-1.53, -1.43], respectively.

We examined potential gender differences in a factorial repeated-measures ANOVA, specifying ratings of children's personality traits (mother-reported vs. father-reported vs. TS) as the within-subjects factor and child gender as the between-subjects factor, for each of the five higher-order personality traits. We examined both the main effect of gender and the interaction effect between gender and rater. For N, there was neither a significant main effect of gender, $F(1, 324) = 0.53$, $p = .467$, nor a significant interaction effect between gender and rater, $F(2, 648) = 0.23$, $p = .794$. For E, there was only a significant main effect of gender, $F(1, 324) = 4.16$, $p = .042$, whereby boys ($M = 4.86$, $SE = 0.04$) were rated significantly lower than girls ($M = 4.97$, $SE = 0.04$), but not a significant interaction effect between gender and rater, $F(2, 648) = 0.28$, $p = .758$. For O, there was neither a significant main effect of gender, $F(1, 324) = 0.48$, $p = .490$, nor a significant interaction effect between gender and rater, $F(2, 648) = 0.12$, $p = .884$. For A, there was no significant main effect of gender, $F(1, 324) = 3.66$, $p = .057$, but there was a significant interaction effect between gender and rater, $F(2, 648) = 7.82$, $p < .001$. Examination of the 95% confidence interval of the estimated

marginal means revealed that boys' and girls' confidence intervals did not overlap for TS ratings of A but did for mother and father reports; specifically, TS judges rated boys ($M = 4.50$, $SE = 0.05$) significantly lower on A than girls ($M = 4.88$, $SE = 0.04$). For C, there was only a significant main effect of gender, $F(1, 324) = 22.12$, $p < .001$, whereby boys ($M = 4.22$, $SE = 0.05$) were rated significantly lower than girls ($M = 4.56$, $SE = 0.05$) but not a significant interaction effect between gender and rater, $F(2, 648) = 2.41$, $p = .091$.

Intersource Agreement

Correlation coefficients between mothers', fathers', and judges' TS ratings of child personality are displayed in Table 3. Composite TS ratings reflect the mean of all 45 video-based ratings for a given trait (i.e., across three judges and 15 tasks). Task-specific TS ratings reflect the mean of ratings for a given trait within a single task (i.e., across three judges). We computed correlations between parent reports and task-specific TS ratings separately for each video slice; for brevity, we report only the average for each task-specific correlation in Table 3. We computed Steiger's z using an online Web utility for tests of the equality of dependent correlations (Lee & Preacher, 2013). Results from these comparisons revealed that composite TS ratings correlated more highly with mother- than father-reported child personality for N ($z = 4.04$, $p < .001$), E ($z = 2.15$, $p = .032$), and O ($z = 3.37$, $p < .001$), but not A ($z = 0.19$, $p = .847$) or C ($z = 1.18$, $p = .238$). Results from tests of equality for the task-specific TS domain score revealed that, on average, task-specific TS scores correlated more highly with mother- than father-reported child personality for N ($z = 2.87$, $p = .004$) and O ($z = 2.13$, $p = .033$), but not for E, C, or A (all $|z|$'s < 1.58 , all p 's $> .114$). Finally, mother reports of child personality correlated more with father reports than with the judges' overall composite TS ratings for N ($z = 2.33$, $p = .020$), E ($z = 2.22$, $p = .026$), A ($z = 5.37$, $p < .001$), and C ($z = 5.60$, $p < .001$), but not for O ($z = 0.34$, $p = .732$). Similarly, mother and father reports of child personality correlated more highly than the judges' average task-specific domain score TS ratings did with mother-reported child personality for all of N ($z = 3.70$, $p < .001$), E ($z = 3.33$, $p < .001$), O ($z = 2.64$, $p = .008$), A ($z = 6.07$, $p < .001$), and C ($z = 6.96$, $p < .001$).

Given that aggregated informant ratings have previously shown better predictive capacity over single informants (e.g., Kolar et al., 1996; Letzring, Wells, & Funder, 2006), we also examined correlations between composite TS domain scores and the mean of parents' reports of child personality; we then compared the correlation coefficients for the aggregated versus nonaggregated parent reports by examining whether the 95% confidence intervals overlapped (confidence intervals were not available for task-specific TS scores). Across all traits, the correlation coefficients for aggregated parent reports fell within the confidence intervals obtained for mother-TS correlations, suggesting that mean-parent ratings were comparable with mothers' reports. In contrast, the correlation coefficients for aggregated parent reports of N, E, and O fell above the confidence intervals obtained for father-TS correlations, sug-

¹ TS observations were missing for the Digit Span task, which was added to the protocol after data collection began, and due to task refusal by child participants.

Table 2
Descriptive Statistics for Mother, Father, and Thin-Slice Ratings of Child Personality

Variable	Range	<i>M</i>	<i>SD</i>	Skew	<i>SE</i>	Kurtosis	<i>SE</i>
Thin-slice personality ratings							
Neuroticism	1.99–5.05	3.37	0.55	0.38	0.14	−0.19	0.27
Extraversion	2.43–5.56	4.32	0.57	−0.38	0.14	−0.15	0.27
Openness	2.60–5.62	4.40	0.48	−0.34	0.14	0.24	0.27
Agreeableness	2.42–5.74	4.69	0.59	−0.96	0.14	1.21	0.27
Conscientiousness	2.54–5.88	4.39	0.63	−0.48	0.14	0.14	0.27
Mother-reported personality ratings							
Neuroticism	1.00–5.82	3.27	0.91	−0.17	0.14	−0.23	0.27
Extraversion	2.00–6.92	5.23	0.81	−0.40	0.14	0.51	0.27
Openness	2.00–7.00	5.40	0.86	−0.33	0.14	0.02	0.27
Agreeableness	1.29–7.00	4.62	0.95	−0.08	0.14	−0.02	0.27
Conscientiousness	1.30–7.00	4.38	1.01	−0.07	0.14	−0.23	0.27
Father-reported personality ratings							
Neuroticism	1.00–5.39	3.24	0.80	−0.18	0.14	−0.10	0.27
Extraversion	2.22–7.00	5.19	0.74	−0.23	0.14	0.62	0.27
Openness	3.00–7.00	5.36	0.78	−0.10	0.14	−0.46	0.27
Agreeableness	1.71–7.00	4.51	0.97	−0.09	0.14	0.35	0.27
Conscientiousness	1.60–6.90	4.41	0.87	0.02	0.14	0.33	0.27
Mean parent-reported child behavior checklist syndromes							
Internalizing	0.00–36.00	5.05	4.48	2.20	0.14	8.95	0.27
Externalizing	0.00–33.00	5.15	4.95	1.64	0.14	4.36	0.27
Total	0.00–109.00	19.30	14.76	1.54	0.14	4.29	0.27
Mean parent-reported child behavior checklist competencies							
Activities	5.00–14.75	10.99	1.84	−0.63	0.14	0.13	0.27
Social	1.50–14.00	9.02	2.29	−0.37	0.14	−0.05	0.27
Academic	2.50–6.00	5.12	0.81	−1.18	0.14	1.15	0.27
Total	13.19–33.00	25.08	3.67	−0.45	0.14	0.25	0.27

Note. All personality ratings were made on a 7-point scale. $N = 326$.

gesting that mean-parent ratings for these traits correlated more with TS ratings than fathers' reports alone.

Our design allowed a preliminary test of whether judges perform better after having seen multiple slices from different targets (i.e., an initial comparison of "experts" to "nonexperts"). To explore this question, we examined the subset of 70 children who were rated by at least 10 "nonexperts" (i.e., students receiving course credit who watched a maximum of four TS videos in a single session), and compared their ratings with those made by RAs who watched and rated multiple slices over time. We computed Steiger's z to test for differences in how much the "expert" (i.e., RAs) versus "nonexpert" judgments correlated with mother-

and father-reported ratings for these children. In other words, we correlated the "nonexpert" ratings for the target child with parent reports, and compared this with the "expert" rating correlations with parent reports for the same child. Expert ratings, $r = .22$, $p = .071$ correlated stronger with mother-reported ratings than nonexpert ratings, $r = .02$, $p = .880$ did for A ($z = -2.43$, $p = .015$). Expert and nonexpert ratings did not differ in the strength of their correlation with the remaining mother-reported ratings (N: $z = -1.37$, $p = .171$; E: $z = -0.14$, $p = .887$; O: $z = -1.60$, $p = .111$; C: $z = -0.34$, $p = .731$) or any of the father-reported traits (N: $z = -0.24$, $p = .810$; E: $z = -0.61$, $p = .541$; O: $z = -0.43$, $p = .671$; A: $z = -1.72$, $p = .086$; C: $z = -0.53$, $p = .594$). Thus,

Table 3
Correlation Coefficients Between Mother, Father, and Thin-Slice Ratings of Child Personality

Traits	Mother:Father	Mother:Thin-slice		Father:Thin-slice		Mean-parent:Thin-slice	
		Composite	Task-specific	Composite	Task-specific	Composite	Task-specific
Neuroticism	.48*** [.39, .56]	.33*** [.23, .42]	.23	.11 [.01, .20]	.07	.26*** [.16, .36]	.18
Extraversion	.45*** [.34, .55]	.31*** [.21, .41]	.23	.19** [.08, .30]	.14	.30*** [.20, .40]	.22
Openness	.45*** [.36, .54]	.43*** [.33, .53]	.28	.25*** [.15, .35]	.16	.40*** [.31, .50]	.26
Agreeableness	.54*** [.44, .64]	.21*** [.09, .33]	.15	.20*** [.08, .32]	.14	.23*** [.11, .35]	.16
Conscientiousness	.66*** [.60, .73]	.38*** [.27, .48]	.28	.33*** [.23, .43]	.24	.39*** [.29, .49]	.29
Mean	.52	.33	.23	.22	.15	.32	.22

Note. The composite domain score is the mean of all 45 ratings across judges and tasks (3 judges \times 15 tasks), whereas task-specific scores are the mean of ratings across three judges for a single task. Correlations for task-specific ratings were computed separately for each video slice; the results displayed in this table reflect the average correlation obtained between parent reports and each task-specific rating. Values in brackets denote 95% confidence intervals.

** $p < .01$. *** $p < .001$.

we found limited evidence for an effect of “expert” rating superiority in this preliminary examination.

Interjudge Agreement

Intraclass correlation coefficients (ICCs) among judges’ ratings are displayed in Table 4. We first examined the average agreement between individual judges who observed the same task (ICC [1,1]) and the average agreement of the composite score across tasks of the three judges who observed the same task (ICC[1,3]). Next, we examined the average cross-task/cross-judge correlation between ratings (ICC[2,1]; mean rater, 15 tasks; $n = 326$). Agreement between individual judges who observed the same task (i.e., the ICC using single ratings for a single task) ranged from .32 to .53 with a mean of .43, whereas agreement of the composite score of the three judges who observed the same task (i.e., the ICC using the averaged ratings for a single task, which is the equivalent of Cronbach’s alpha for three raters assessing the same TS) ranged from .58 to .77 with a mean of .69, and agreement for composite ratings across tasks (i.e., the ICC using the averaged ratings across tasks) ranged from .27 to .54, with a mean of .43.²

Intercorrelations among personality traits in TS and mother-reported ratings are displayed in Table 5. Similar to above, we computed Steiger’s z to test for differences in the magnitude of the different correlations (e.g., intercorrelations between N and E for TS vs. mother-reported ratings; Lee & Preacher, 2013). TS ratings were more highly intercorrelated than mother-reported ratings for the following trait pairs: N–E ($z = 11.84, p < .001$), N–O ($z = 10.92, p < .001$), E–O ($z = 11.10, p < .001$), O–A ($z = 2.65, p = .008$), and A–C ($z = 9.22, p < .001$). Mother ratings were more highly intercorrelated than TS ratings only for the relationship between N and A ($z = 4.72, p < .001$). TS and mother ratings did not differ in the strength of their intercorrelations for N–C ($z = 1.27, p = .205$), E–A ($z = 0.38, p = .706$), E–C ($z = 0.88, p = .378$), or O–C ($z = 0.58, p = .563$). Furthermore, principal factor analyses of item-level data conducted separately on TS and mother-ratings revealed that the first factor explained more variance in TS ratings (eigenvalue = 25.43, 50.86% of variance) than in mother-ratings (eigenvalue = 12.99, 25.98% of variance), suggesting greater overlap in strangers’ TS ratings of different personality traits versus mothers’ ratings of their own children’s different traits.

Cross-Situational Consistency

We estimated the cross-situational consistency of children’s TS trait ratings (i.e., cross-rater/cross-task ICC divided by average rater/single task ICC). Cross-situational consistency estimates ranged from .41 to .71, with a mean of .63 (see Table 4). These estimates were notably higher than those derived in adult participants in Borkenau et al.’s (2004) work, suggesting that there are potential developmental differences in cross-situational consistency (i.e., children appear to be more consistent than adults in the personality traits that they exhibited across different situations). Estimates in Borkenau et al.’s (2004) study ranged from .29 to .61 with a mean of .43. Similar to Borkenau et al.’s (2004) research, however, cross-situational consistency estimates were highest for E and lowest for O in our child sample.

To test for potential familiarity effects, we examined whether the absolute magnitude of discrepancies between TS ratings and an

external criterion (i.e., caregiver ratings of personality) decreased as time in the study protocol and familiarity with the laboratory context increased. Unconditional latent growth curve (LGC) analyses were estimated in Mplus version 6.1 (Muthén & Muthén, 2010) to examine whether children were more likely to express their true personalities as they adjusted to the novel laboratory environment (see Figure 1). We computed absolute difference scores (ADS) by subtracting the mean-parent rating for each trait from scores obtained for each TS situation (e.g., TS A minus mean-parent A) and converting these differences to their corresponding absolute values (i.e., signs are ignored). We estimated separate models for each trait. The results of the LGC models showed that the slopes of discrepancy scores for N ($M = 0.00, SE = 0.00, p = .224$), A ($M = 0.00, SE = 0.00, p = .708$), and C ($M = 0.00, SE = 0.00, p = .372$) were not significant, suggesting that parent–TS discrepancies remained relatively stable across the study protocol. In contrast, LGC models showed that the slopes for E ($M = 0.01, SE = 0.00, p = .016$) and O ($M = 0.01, SE = 0.00, p = .010$) were significant, suggesting that discrepancies for these traits increased across the study protocol. With the exception of N, the slope variances were also significant, suggesting individual variability in discrepancy trajectories across the study protocol.

Convergent, Divergent, and Predictive Validity

Correlation coefficients between TS ratings of child personality traits and parent- and youth-reported behavior are displayed in Table 6. Given strong theoretical rationale for specific hypotheses for psychopathology-relevant outcomes (Tackett, 2006), only hypothesized associations are presented. Associations with competency outcomes were treated as exploratory given less existing evidence; thus, all associations are presented.

Of the six predicted associations with TS N, none were significant. Of the three predicted associations between TS E and internalizing behaviors, one was significant (parent-reported internalizing based on the clinical interviews). Of the six predicted associations between TS A and C and externalizing behaviors, all were significant—for both parent and youth report, and for both questionnaire and clinical interview methods. All CBCL competency scores negatively correlated with N and positively associated with E, O, A, and C, with the exception of activities competency, which did not significantly correlate with A.

Given that mother report overlapped most substantially with TS ratings, we further tested whether TS ratings (entered in Step 2 of a hierarchical regression) predicted significant variance in these outcomes, above and beyond mother ratings (entered in Step 1). The addition of TS ratings incrementally predicted variance above and beyond mother reports for the following outcomes: parent-reported CBCL externalizing behaviors (R^2 change = .02, $p = .044$), parent-reported CDISC externalizing behaviors (R^2 change = .02, $p = .030$), youth-reported CDISC externalizing

² In order to directly test the robustness of the psychometric estimates, we randomly split the total sample into two parts (Group 1: $n = 158$; Group 2: $n = 168$) and examined interjudge agreement and cross-situational consistency within each subsample. The results of these analyses yielded coefficients that were virtually identical across subsamples, demonstrating the robustness of the psychometric properties estimated for the TS ratings. Results are available upon request.

Table 4
Intraclass Correlation Coefficients Among Judges of Thin-Slice Videotaped Behavior Sequences who Observed Targets Performing the Same or a Different Task

Traits and facets	Intraclass correlation coefficients (ICC)			Cross-situational consistency
	Single rater/ single task (<i>n</i> = 4,681)	Average rater/ single task (<i>n</i> = 4,681)	Cross-rater/ cross-task (<i>n</i> = 326)	
Neuroticism	.45	.71	.47	.66
Fearful	.40	.67	.45	.67
Negative affect	.33	.59	.38	.64
Shy	.50	.75	.53	.71
Extraversion	.53	.77	.51	.66
Activity level	.51	.76	.54	.71
Considerate	.42	.69	.42	.61
Positive emotions	.44	.70	.46	.66
Sociability	.53	.77	.54	.70
Openness	.43	.69	.37	.54
Intellect	.39	.66	.27	.41
Openness	.41	.68	.39	.57
Agreeableness	.39	.65	.41	.63
Antagonism	.39	.66	.44	.67
Strong willed	.32	.58	.33	.57
Conscientiousness	.49	.75	.48	.64
Achievement oriented	.45	.71	.39	.55
Distractible	.40	.67	.39	.58
Organized	.43	.70	.49	.70

Note. The two to four subcomponents underneath each Big Five trait are its lower-order facets. Column 1: Average agreement between two individual judges of the same behavior episode [ICC (1,1)]. Column 2: Average reliability of the composite scores of three judges of same behavior episode [ICC (1,3)]. Column 3: Average cross-task/cross-judge correlations between the composite scores of 3 judges [ICC(2,1)]. Column 4: Cross-situational consistency = Column 3/Column 2.

behaviors (R^2 change = .03, $p = .047$), activities competencies (R^2 change = .05, $p = .008$), social competencies (R^2 change = .06, $p < .001$), and school competencies (R^2 change = .07, $p < .001$); but not for parent-reported CBCL internalizing behaviors, parent-reported CDISC internalizing behaviors, youth-reported CDISC internalizing behaviors. This incremental prediction represents a particularly stringent test, given the overlapping method variance for most of these associations (i.e., most analyses relied on mother report and questionnaire methods for both Step 1 IVs and DVs).

We also examined predictive validity across both the parent- and youth -reported ICID traits approximately 3 years after the initial TS ratings. TS ratings in middle childhood predicted the analogous trait ratings for N ($r = .09$, $p = .098$ for parent reports, $r = .32$, $p < .001$ for youth self-reports), for E ($r = .38$,

$p < .001$ for parent reports, $r = .36$, $p < .001$ for youth self-reports), for O ($r = .30$, $p < .001$ for parent reports, $r = .39$, $p < .001$ for youth self-reports), for A ($r = .17$, $p = .002$ for parent reports, $r = .24$, $p < .001$ for youth self-reports), and for C ($r = .28$, $p < .001$ for parent reports, $r = .42$, $p < .001$ for youth self-reports). We further examined whether TS ratings of each trait at baseline would incrementally predict self-reported traits 3 years later controlling for the initial mother-reported traits. The results showed incremental validity for all trait ratings above and beyond mothers' first reports: N (R^2 change = .01, $p = .010$), E (R^2 change = .02, $p < .001$), O (R^2 change = .02, $p = .002$), A (R^2 change = .01, $p = .015$), and C (R^2 change = .03, $p < .001$).

Discussion

These initial findings support the use of TS as a novel and incremental method for assessing child personality. Specifically, TS ratings of child personality showed convergence with parents' ratings of their child's personality. TS ratings demonstrated encouraging psychometric properties, including within-task inter-judge agreement, internal consistency, and cross-judge/cross-task agreement. TS ratings were highly consistent across situations, surprisingly more than has been previously demonstrated for adults. In addition, TS ratings showed associations with behavioral problems and competencies in theoretically predicted ways. Taken together, these results provide encouraging support for a TS approach to assessing child personality that is easy, cheap, efficient to implement, and overcomes limitations of other dominant assess-

Table 5
Intercorrelations Among Personality Traits in Thin-Slice and Mother-Reported Ratings

Traits	1	2	3	4	5
1. Neuroticism		-.92***	-.83***	-.17**	-.23***
2. Extraversion	-.62***		.84***	.40***	.34***
3. Openness	-.37***	.39***		.29***	.48***
4. Agreeableness	-.49***	.42***	.10		.79***
5. Conscientiousness	-.31***	.29***	.44***	.36***	

Note. $N = 326$. Thin-slice ratings are presented above the diagonal and mother-reported ratings below the diagonal.

** $p < .01$. *** $p < .001$.

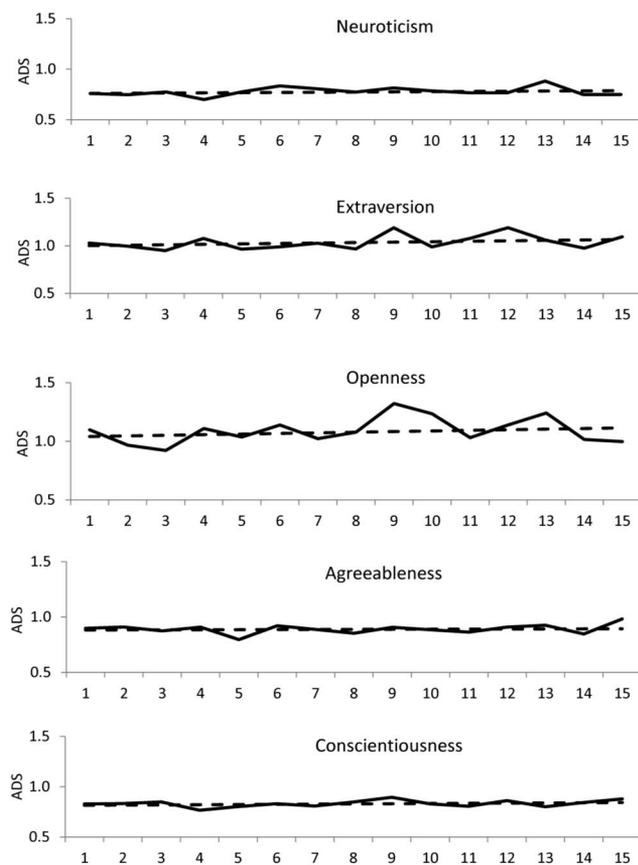


Figure 1. Slopes from latent growth curve models of absolute difference scores (ADS) computed by subtracting the mean-parent report from thin-slice ratings across 15 video segments. Solid lines denote sample means, whereas broken lines denote estimated means.

ment methods. Furthermore, the present study highlights the utility of a *construct validation* framework, such that enhanced measurement draws from previous substantive contributions, and similarly informs them in return.

Intersource Agreement

TS ratings showed significant convergence with both mothers' and fathers' ratings of their child's personality traits, despite the enormous differences between parents and strangers in acquaintanceship and the motivation to perform as a valid judge (Tackett, 2011). TS ratings also overlapped significantly more with mothers' ratings than with fathers' ratings, at least partly supporting the predominant use of mothers as informants in most child research and applied settings (De Los Reyes & Kazdin, 2005; Duhig, Renk, Epstein, & Phares, 2000). Consistent with other findings for observational lab tasks, consensus with parental informants increased when aggregating across multiple TS situations, rather than relying on a single situation (e.g., Forman et al., 2003). Given the relative ease of assessing multiple TS videos (five diverse situations could easily be assessed in 15 min), the validity pay-off for using multiple TS is likely substantial.

We also conducted preliminary analyses to examine whether "nonexperts" showed lower convergence with parent ratings than "experts" (raters who observed and rated substantially more children in the protocol). Previous research showed that judges who watched longer slices of a target (i.e., were more of an expert on the target) were more accurate in their personality ratings of the target, potentially suggesting an expert advantage for a specific target (Carney, Colvin, & Hall, 2007). In our study, we extended this work to test whether an expert advantage for general TS ratings would emerge by comparing judges who had watched many slices of different targets to those who had watched only a limited number of slices. Our analyses showed an expert advantage only for trait A, but not for N, E, O, or C. This study, however, was not designed to examine this issue directly; thus, more rigorous examinations are needed to shed light on this important issue. For example, tracking expert TS ratings consecutively would allow an examination of their potentially enhanced validity as the same rater is exposed to more children, and an expanded scope of potential behavior in these situations. This remains an interesting question for future research.

We directly compared intersource agreement between TS ratings and mother–father agreement. Mothers and fathers are gen-

Table 6
Correlations Between Thin Slice Ratings of Personality and Child Psychopathologies and Competencies

Behavioral outcomes	Neuroticism	Extraversion	Openness	Agreeableness	Conscientiousness
CBCL syndrome scales					
P-R internalizing problems	.08	-.07			
P-R externalizing problems	.05			-.23***	-.23***
CDISC symptoms					
P-R internalizing	.11	-.12*			
Y-R internalizing	.07	-.02			
P-R externalizing	.02			-.29**	-.33**
Y-R externalizing	.01			-.21**	-.14**
Competence scales					
P-R activities competence	-.20***	.19**	.21***	.07	.19**
P-R social competence	-.31***	.29***	.29***	.23***	.34***
P-R academic competence	-.17**	.16**	.27***	.26***	.37***

Note. $N = 326$. CBCL = Child Behavior Checklist; CDISC = Computerized Diagnostic Interview Schedule for Children; P-R = parent report; Y-R = youth report. Only hypothesized associations are presented for behavioral problem scales, given clear theoretical context for such associations in the literature (Tackett, 2006).

* $p < .05$. ** $p < .01$. *** $p < .001$.

erally expected to overlap considerably in their ratings of their children's personality traits, given the largely overlapping biases in motivation and available information that they possess (Tackett, 2011). Indeed, mother–father ratings correlated more highly than mother–TS ratings in the current study, which may reflect parents' greater opportunities to observe their children in similar contexts (i.e., Laidra et al., 2006; Stanger & Lewis, 1993). Nonetheless, the magnitude of mother–TS ratings might be considered especially impressive against the highly stringent comparison of mother–father ratings. One notable exception emerged for composite TS ratings of O, which showed comparable agreement for mother–father and mother–TS ratings. Tasks in the videotaped protocol may have elicited aspects of child intellect, which tends to be weighted heavily in measures of child O, in addition to capturing aspects of creativity and curiosity (De Fruyt, Mervielde, & Van Leeuwen, 2002; Herzhoff & Tackett, 2012). TS methodology may therefore be advantageous for lab-based assessment of child O, a personality dimension that is less understood than the other Big Five traits, particularly in children (Connelly, Ones, & Chernyshenko, 2014; Mervielde, De Fruyt, & Jarmuz, 1998).

In contrast to prior research on adults (e.g., Kolar et al., 1996; Letzring et al., 2006), child personality ratings aggregated across multiple informants (i.e., mean-parent report) did not correlate more strongly with TS ratings than parents' nonaggregated reports. Rather, the relationships between parents' aggregated reports and TS ratings were comparable in size with the relationships between mothers' nonaggregated reports and TS ratings. In contrast, the relationships between parents' aggregated reports and TS ratings were stronger than the relationships between fathers' nonaggregated ratings and TS ratings for N, E, and O, whereas the relationships between parents' aggregated reports and TS ratings were comparable in size to the relationships between fathers' nonaggregated ratings and TS ratings for A and C. These results may provide insights into which informants show the greatest utility in capturing child personality. In the current sample, mother reports showed an advantage in reporting on observable aspects of child emotionality (N and E) and intellect/imagination (O), whereas mothers and fathers provided comparable reports of child self-regulatory capacity (A and C). This does not imply that fathers' unique perspectives are not valuable sources of information. Indeed, prior research has shown that discrepancies between mothers' and fathers' reports of child personality account for incremental variance in children's internalizing problems, such as depression and anxiety (Tackett, 2011). Nevertheless, the current results may be informative to researchers who, for practical reasons (e.g., time and monetary constraints), are limited to single-informant reports of child personality.

The overall pattern of findings for TS–parent agreement was somewhat deviant from TS research in adults, and also divergent across mother and father reports. Child TS ratings showed high agreement with both mothers' and fathers' ratings of O and C, in contrast to robust findings that adult TS ratings show superior validity for E (Borkenau et al., 2009, 2004; Hall et al., 2008; Vazire, 2006; Vazire & Gosling, 2004; Vazire et al., 2008) and, in some cases, O (Borkenau et al., 2004). This may partly reflect a tendency for O to covary more highly with C among youth (Tackett et al., 2012), whereas O tends to covary more highly with E among adults (DeYoung, 2006; Digman, 1997). In addition, particular traits may be more or less observable across develop-

ment. That is, laboratory protocols may elicit more variance in youth-specific aspects of C (e.g., low distractibility, self-control) than aspects of C that are more apparent among adults (e.g., competence, dutifulness, deliberation). In contrast, youth may be less likely to express aspects of E that may be more visible to their parents (e.g., sociability) due to unfamiliarity with the laboratory context.

Results from LGC models also showed that TS ratings of E and O became increasingly discrepant from parent ratings over the study protocol, suggesting that increased familiarity with the laboratory setting did not facilitate elicitation of those aspects of E and O reported by parents. These results should be considered in the context of overall higher ratings for children's E and O from parent reports, compared with TS ratings (see Table 2). Together, these findings suggest that children's presentation early in the lab visit is more similar to their parent's ratings than later in the lab visit, such that TS raters see the children as lower in both E and O after they have been engaged in the protocol for some time. One interesting hypothesis may be that these traits are more strongly affected by impression management (and thus, correspond more with potential positivity biases of their parents), which dissipates over time.

The poorest agreement differed by parental informant: TS ratings showed poorest overlap with A when compared with mothers' reports, but showed poorest overlap with N when compared with fathers' reports. Again, this difficulty in conceptualizing these dimensions may partly reflect greater overlap between N and A among youth than in adults (Tackett et al., 2012). Consistent with findings on adults, informants may be generally poor at assessing traits that are low in observability, like N (Borkenau et al., 2009, 2004; Hall et al., 2008; Vazire & Gosling, 2004; Vazire et al., 2008). Gender differences have emerged within some of the adult TS literature, such that women tend to be more accurate at rating N (Lippa & Dietz, 2000). Accordingly, mothers may be more sensitively attuned to the inner emotional states of their children. In contrast, mothers' ratings of A may be biased by the evaluative nature of this personality dimension. Among adults, the evaluativeness of traits is believed to be problematic and to increase the likelihood that self-ratings will be biased (Vazire, 2006). Childhood A reflects aspects of obedience versus willful behavior, which mothers may perceive as reflecting the quality of their child rearing abilities; as a result, ego-protective biases may distort mothers' ratings of this trait. Such findings highlight potentially powerful information embedded in TS ratings, which may allow critical triangulation on constructs of interest (Kraemer et al., 2003). Indeed, the use of such ratings may prove highly informative in attempts to disentangle various forms of informant/source bias by providing an external assessment of traits, and offering information on informant-specific strengths for different types of dispositions.

Given gender differences in child personality/temperament traits (e.g., Else-Quest, Hyde, Goldsmith, & Van Hulle, 2006), another substantive question is whether gender differences might differ between sources (i.e., TS vs. mother ratings vs. father ratings). We found a significant interaction between gender and rater for trait A only, such that TS judges rated girls significantly higher than boys whereas mothers and fathers did not. We also found main effects for gender, but only for E and C, with boys scoring lower than girls on these traits across all raters. This pattern of results is consistent

with the general literature on gender differences in personality traits, which shows that women generally score higher than men on E, A, and C (e.g., Schmitt, Realo, Voracek, & Allik, 2008), whereas gender differences in N may increase only in adolescence (Soto, John, Gosling, & Potter, 2011). Taken together, we found little evidence for gender differences in our study that were robust across informant. Future research on gender differences in temperament and personality may benefit from more focused attention to how specific informant biases confound or elucidate such differences across development.

Interjudge Agreement

Interjudge agreement across TS raters was substantial. Two unacquainted observers exposed to the same TS of a child showed strong convergence (average ICC = .43), rivaling that of mother–father agreement for most traits. We also examined internal consistency across all three raters (i.e., the equivalent of Cronbach’s alpha for three raters assessing the same slice), and found that the reliability of averaged ratings was good across all traits (average ICC = .69), comparable with those found for the domain composites in Borkenau et al.’s (2004) research. Although we performed analyses consistent with Borkenau et al.’s (2004) report, specifically comparing one to three raters and finding substantial improvement with three raters, it is likely that once idiosyncratic biases and noise are aggregated out of TS judgments, more raters do not necessarily offer significantly greater information (Beer, 2013). Furthermore, agreement of different judges rating different tasks also evidenced strong consensus (average $r = .43$), suggesting that independent TS ratings show the type of generalizability that would be expected if the ratings were tapping into valid underlying traits.

In addition, we tested the hypothesis that trait intercorrelations for TS ratings would be higher than those for mother–ratings, based on robust evidence from TS ratings of adults (Beer & Watson, 2008a, 2008b; Rauthmann, 2013; Rauthmann & Kolar, 2010). Overall, the pattern of intercorrelations was more complicated than has been observed in adult samples. Specifically, most trait correlations were higher for TS ratings than for mother report, as expected. However, mother ratings showed higher correlations for N and A than did TS ratings, and several intercorrelations were no higher in TS ratings than in mother report. The higher correlations for N and A in mother ratings than TS ratings is particularly surprising given that these traits have shown consistently greater covariation in other-ratings (i.e., spouses, dating partners, friends, and strangers) compared with self-ratings among adults (Beer & Watson, 2008a). It is interesting to note that hierarchical structure analyses of caregiver-reported children’s personality also show that N and A are linked more strongly in early ages than in adulthood, suggesting that their overlapping features might be more salient when parents describe their children than when adults describe others (e.g., Tackett et al., 2012). Consistent with findings in adults (Beer & Watson, 2008a, 2008b; Rauthmann, 2013; Rauthmann & Kolar, 2010; Wood et al., 2010, but see Srivastava, Guglielmo, & Beer, 2010, for a more differentiated trait structure), we found that the first extracted component score explained a greater amount of variance in TS ratings than in mother–ratings. Taken together, these findings suggest that TS ratings do not demonstrate an overall simplified structure when compared with

mother ratings, although some trait differentiation may be better achieved by better-acquainted informants, such as mothers. Indeed, some of the TS rating intercorrelations (e.g., between N, E, and O) were quite high, evidencing poorer discrimination. This poor differentiation may partly reflect general difficulty in accurately observing and rating child N (from acquainted or unacquainted informants). However, unacquainted observers may differentiate other traits—including N and A—better than parents do. This pattern of findings underscores potential advantages of examining other sources of information, such as TS reports, alongside standard informants such as mothers.

Cross-Situational Consistency

Estimates of trait cross-situational consistency across the 15 TS situations was remarkably high, ranging from .54 (O) to .66 (N and E). These findings represent a departure from Borkenau et al.’s (2004) examination in adults, such that cross-situational consistency estimates for the children in this study were substantially higher (average estimate = .43), despite the same number of TS situations. Nevertheless, the pattern was similar in showing lowest cross-situational consistency for O in both our childhood sample and in Borkenau et al.’s adult sample. It is interesting to consider why these differences emerged in distinct age groups. The stability of personality traits is thought to be lower in childhood than adulthood (e.g., Roberts & DelVecchio, 2000). These findings raise the provocative suggestion that short-term stability (i.e., within-episode) may be independent of, or even negatively correlated with, longer-term stability (i.e., over months or years) across child–adult development.

The extent to which children are more behaviorally constrained by situational features, such as the laboratory setting, may exaggerate behavioral consistency across the types of TS examined here. It may also be, however, that children are *less* constrained by such situational features, allowing for true personality consistency to emerge more so than for adults. These are exciting ideas to examine in future research and have important implications for competing hypotheses of child personality as reflecting stable individual differences versus child personality as more radically influenced by situational features in potentially erroneous or noisy ways. Our findings suggest that child personality is observable and measurable across a variety of situations, and that such individual differences may be even less susceptible to momentary and situation-based change than adult behavior in similar circumstances.

We did examine whether the novelty of the lab situation may have a constraining effect on “true” behavioral expression (or, at least, the trait variance reported by parents), by comparing TS ratings in the first TS (i.e., at the very beginning of the protocol when the children have just arrived at the lab) to TS ratings obtained later in the protocol (i.e., as they adjusted to the lab environment). Specifically, we fit LGC models to discrepancies between TS ratings and parent ratings to examine whether TS ratings showed greater convergence with parent ratings over time. These findings suggested no change in TS–parent discrepancies for N, A, and C across the protocol, suggesting that TS impressions of these traits did not differ for the child’s behavior later versus earlier in the protocol. However, counter to expectations, TS–parent discrepancies did change for E and O, but in the direction

of *greater* discrepancies later in the protocol. That is, TS ratings for the child's E and O showed greater overlap with parental ratings when based on situations early in the protocol, rather than later. One possible interpretation of these findings is that initial impressions of children's E and O earlier in the protocol may reflect the children "on their best behavior," possibly tapping into motivated biases in parental report for these traits. Future research will be helpful to better understand this effect.

Convergent, Divergent, and Predictive Validity

We explored evidence for the convergent and divergent validity of TS ratings by examining associations across major domains of child behavior problems (i.e., internalizing and externalizing problems) and major domains of child competencies (i.e., extracurricular activities, social functioning, and academic performance). Overall, patterns of association further supported the validity of TS ratings of child personality. Externalizing problems showed anticipated associations with low A and C (Tackett, 2006). However, internalizing problems only associated with low E based on parent-reported (and not youth self-reported) internalizing symptoms assessed via clinical interview. This is consistent with evidence for attenuated associations with child internalizing problems, given the greater difficulty in accessing internalizing-relevant information (e.g., feelings of sadness and anxiety are less observable than disruptive behavior patterns; Achenbach, McConaughy, & Howell, 1987; De Los Reyes & Kazdin, 2005; Duhig et al., 2000). In addition, this pattern of findings was largely robust across both method (i.e., questionnaire and structured interview) and informant (i.e., parent- and youth self-report).

Associations emerged for competencies, as well. In general, an "adaptive" trait profile of lower N, and higher E, O, A, and C, associated with almost all competency domains, as would be expected. Associations did show some potential evidence of differentiation as well, however. For example, E was most highly correlated with social competencies, whereas C was most highly correlated with academic competencies. These findings provide further support in TS ratings of child personality as tapping into valid trait information with import across major domains of behavioral, social, and academic functioning.

Finally, analyses of the predictive validity of childhood TS ratings for later personality ratings from both parent and self were also robust and impressive. TS ratings of child personality traits significantly predicted analogous trait ratings from both parents and self several years later, with the exception of later parental ratings of N. These findings were also notable in that, with the exception of E, TS ratings showed larger associations with later youth self-report than they did with parent report. These results may indicate that TS ratings are more sensitive to potentially valid trait variance that even the parents cannot access. Indeed, the overall lack of associations with trait N for parents' reports—which is consistent with the broader literature regarding difficulty of assessing N via informant reports—did not persist in this domain. Rather, TS ratings of children's N at ages 9–10 significantly predicted youth's self-reported N several years later. Thus, the attenuated associations with N elsewhere belie the fact that TS ratings, even for N, are capturing meaningful trait variance in middle childhood.

Limitations and Future Directions

One limitation of our study is the exclusive use of situations that included an adult authority figure. It has been hypothesized that personality is less expressed in "strong" or structured situations (although empirical research seems scarce; Cooper & Withey, 2009). Being in the presence of an adult authority figure following structured lab tasks could be argued to model a strong or structured situation. The fact that we measured a sufficient degree of variance in TS ratings suggests that the presence of the adult experimenter was not imposing enough constraints on the child's personality expression to prevent meaningful individual differences from emerging. However, future research might include less-structured situations and situations that involve other interaction partners, such as mothers, siblings, or peers, to better identify situational impact on child trait expression.

Intersource disagreement may be informative, as informant discrepancies have been shown to predict important clinical outcomes (e.g., treatment response; De Los Reyes et al., 2010; Ferdinand, van der Ende, & Verhulst, 2004). However, the extent to which discrepancies for child personality traits relate to functioning has not been extensively studied. Indeed, intersource agreement tends to be higher for well-adjusted children and lower for maladjusted children, perhaps because well-adjusted children are more consistent across situations (Victor, Halverson, & Wampler, 1988). Interrater discrepancies for child personality traits across TS tasks can potentially be used to examine whether children who are difficult to rate experience higher levels of problems, and whether this varies across different types of psychopathology. In addition, the collection of TS data opens many new avenues for pursuing a variety of theoretical and applied questions. Collecting TS ratings from standardized judges, who rate all available tasks, would lend itself to interesting differentiation of target and perceiver effects. The data available here could be readily extended in this way by modifying the TS rater paradigm. This example highlights the longevity of data collected using a TS paradigm, as used in the present study, which easily lends itself to many different approaches and applications for a variety of substantive research questions.

Work on personality hierarchies has also noted differences in the structure of traits in childhood compared with adulthood (DeYoung, 2006; Digman, 1997; Tackett et al., 2012). The elucidation of structural changes across development, however, faces challenges posed by the necessity of relying on different informants as children transition through adolescence, when parental knowledge typically declines (Collins & Laursen, 2004). Using TS ratings within a longitudinal design may partially address this issue. A final notable future direction involves direct comparison of TS ratings with microcodings from existing structured laboratory batteries in order to determine whether microcoding retains advantages in the measurement of child personality above and beyond the more efficient TS approach. We have previously demonstrated the effectiveness of TS ratings over microcoding for other child constructs (e.g., sibling interactions; Prime et al., 2014) but such work should be extended to child personality as well.

Conclusions

Overall, these results suggest that TS ratings of child personality may offer a valid, reliable, and resource-efficient method of as-

assessment to increment common approaches of parent/teacher questionnaire measurement or intensive, structured lab-based paradigms. Such ratings appear to tap into meaningful personality variance based on brief exposure to the child in various situations. They do not appear to be particularly situation-dependent, and the validity of ratings can be achieved with little to no training of raters. Indeed, examining TS ratings here helped to inform several important theoretical issues in the extant literature, such as the stability and legibility of children's personality compared to adults. Thus, these data may promote new insights for the development of the character and expression of personality and for broader processes underlying person perception. As such, the TS methodology may substantially accelerate theoretical and empirical understanding of child personality, overcoming previous barriers to better understanding dispositional differences in early life.

References

- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, *101*, 213–232. <http://dx.doi.org/10.1037/0033-2909.101.2.213>
- Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*. Department of Psychiatry, University of Vermont, Burlington, VT.
- Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology*, *64*, 431–441. <http://dx.doi.org/10.1037/0022-3514.64.3.431>
- Back, M. D., Schmukle, S. C., & Egloff, B. (2010). Why are narcissists so charming at first sight? Decoding the narcissism-popularity link at zero acquaintance. *Journal of Personality and Social Psychology*, *98*, 132–145. <http://dx.doi.org/10.1037/a0016338>
- Beer, A. (2013). Group personality judgments at zero acquaintance: Communication among judges versus aggregation of independent evaluations. *Journal of Research in Personality*, *47*, 385–389. <http://dx.doi.org/10.1016/j.jrp.2013.03.008>
- Beer, A., & Watson, D. (2008a). Asymmetry in judgments of personality: Others are less differentiated than the self. *Journal of Personality*, *76*, 535–560. <http://dx.doi.org/10.1111/j.1467-6494.2008.00495.x>
- Beer, A., & Watson, D. (2008b). Personality judgment at zero acquaintance: Agreement, assumed similarity, and implicit simplicity. *Journal of Personality Assessment*, *90*, 250–260. <http://dx.doi.org/10.1080/00223890701884970>
- Bellak, L., & Abrams, D. M. (1997). *The thematic apperception test, the children's apperception test, and the senior apperception technique in clinical use* (6th ed.). Boston, MA: Allyn & Bacon.
- Borkenau, P., Brecke, S., Möttig, C., & Paelecke, M. (2009). Extraversion is accurately perceived after a 50-ms exposure to a face. *Journal of Research in Personality*, *43*, 703–706. <http://dx.doi.org/10.1016/j.jrp.2009.03.007>
- Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology*, *62*, 645–657. <http://dx.doi.org/10.1037/0022-3514.62.4.645>
- Borkenau, P., & Liebler, A. (1993). Consensus and self-other agreement for trait inferences from minimal information. *Journal of Personality*, *61*, 477–496. <http://dx.doi.org/10.1111/j.1467-6494.1993.tb00779.x>
- Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology*, *86*, 599–614. <http://dx.doi.org/10.1037/0022-3514.86.4.599>
- Brown, G. L., Mangelsdorf, S. C., Agathen, J. M., & Ho, M. H. (2008). Young children's psychological selves: Convergence with maternal reports of child personality. *Social Development*, *17*, 161–182. <http://dx.doi.org/10.1111/j.1467-9507.2007.00421.x>
- Carney, D. R., Colvin, C. R., & Hall, J. A. (2007). A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality*, *41*, 1054–1072. <http://dx.doi.org/10.1016/j.jrp.2007.01.004>
- Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality development: Stability and change. *Annual Review of Psychology*, *56*, 453–484. <http://dx.doi.org/10.1146/annurev.psych.55.090902.141913>
- Charlebois, P., Tremblay, R. E., Gagnon, C., Larivée, S., & Laurent, D. (1989). Situational consistency in behavioral patterns of aggressive boys: Methodological considerations on observational measures. *Journal of Psychopathology and Behavioral Assessment*, *11*, 15–27. <http://dx.doi.org/10.1007/BF00962696>
- Christopher, A. N., & Schlenker, B. R. (2000). The impact of perceived material wealth and perceiver personality on first impressions. *Journal of Economic Psychology*, *21*, 1–19. [http://dx.doi.org/10.1016/S0167-4870\(99\)00033-1](http://dx.doi.org/10.1016/S0167-4870(99)00033-1)
- Collins, W. A., & Laursen, B. (2004). Changing relationships, changing youth interpersonal contexts of adolescent development. *The Journal of Early Adolescence*, *24*, 55–62. <http://dx.doi.org/10.1177/0272431603260882>
- Compas, B. E., Forsythe, C. J., & Wagner, B. M. (1988). Consistency and variability in causal attributions and coping with stress. *Cognitive Therapy and Research*, *12*, 305–320. <http://dx.doi.org/10.1007/BF01176192>
- Connelly, B. S., Ones, D. S., & Chernyshenko, O. S. (2014). Introducing the special section on openness to experience: Review of openness taxonomies, measurement, and nomological net. *Journal of Personality Assessment*, *96*, 1–16. <http://dx.doi.org/10.1080/00223891.2013.830620>
- Connolly, J. J., Kavanagh, E. J., & Viswesvaran, C. (2007). The convergent validity between self and observer ratings of personality: A meta-analytic review. *International Journal of Selection and Assessment*, *15*, 110–117. <http://dx.doi.org/10.1111/j.1468-2389.2007.00371.x>
- Cooper, W. H., & Withey, M. J. (2009). The strong situation hypothesis. *Personality and Social Psychology Review*, *13*, 62–72. <http://dx.doi.org/10.1177/1088868308329378>
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302. <http://dx.doi.org/10.1037/h0040957>
- Deal, J. E., Halverson, C. F., Jr., Martin, R. P., Victor, J., & Baker, S. (2007). The inventory of children's individual differences: Development and validation of a short version. *Journal of Personality Assessment*, *89*, 162–166. <http://dx.doi.org/10.1080/00223890701468550>
- De Fruyt, F., Mervielde, I., & Van Leeuwen, K. (2002). The consistency of personality type classification across samples and five-factor measures. *European Journal of Personality*, *16*, S57–S72. <http://dx.doi.org/10.1002/per.444>
- De Los Reyes, A., Alfano, C. A., & Beidel, D. C. (2010). The relations among measurements of informant discrepancies within a multisite trial of treatments for childhood social phobia. *Journal of Abnormal Child Psychology*, *38*, 395–404. <http://dx.doi.org/10.1007/s10802-009-9373-6>
- De Los Reyes, A., Goodman, K. L., Kliewer, W., & Reid-Quifones, K. (2008). Whose depression relates to discrepancies? Testing relations between informant characteristics and informant discrepancies from both informants' perspectives. *Psychological Assessment*, *20*, 139–149. <http://dx.doi.org/10.1037/1040-3590.20.2.139>
- 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. <http://dx.doi.org/10.1037/0033-2909.131.4.483>
- De Los Reyes, A., Thomas, S. A., Swan, A. J., Ehrlich, K. B., Reynolds, E. K., Suarez, L., . . . Pabón, S. C. (2012). "It depends on what you mean by 'disagree'": Differences between parent and child perceptions of parent-child conflict. *Journal of Psychopathology and Behavioral Assessment*, *34*, 293–307. <http://dx.doi.org/10.1007/s10862-012-9288-3>

- De Pauw, S. S., Mervielde, I., & Van Leeuwen, K. G. (2009). How are traits related to problem behavior in preschoolers? Similarities and contrasts between temperament and personality. *Journal of Abnormal Child Psychology*, *37*, 309–325. <http://dx.doi.org/10.1007/s10802-008-9290-0>
- de Vries, R. E., Lee, K., & Ashton, M. C. (2008). The Dutch HEXACO Personality Inventory: Psychometric properties, self-other agreement, and relations with psychopathy among low and high acquaintanceship dyads. *Journal of Personality Assessment*, *90*, 142–151. <http://dx.doi.org/10.1080/00223890701845195>
- DeYoung, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample. *Journal of Personality and Social Psychology*, *91*, 1138–1151. <http://dx.doi.org/10.1037/0022-3514.91.6.1138>
- Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology*, *73*, 1246–1256. <http://dx.doi.org/10.1037/0022-3514.73.6.1246>
- Duhig, A. M., Renk, K., Epstein, M. K., & Phares, V. (2000). Interparental agreement on internalizing, externalizing, and total behavior problems: A meta-analysis. *Clinical Psychology: Science and Practice*, *7*, 435–453. <http://dx.doi.org/10.1093/clipsy.7.4.435>
- Durbin, C. E. (2010). Modeling temperamental risk for depression using developmentally sensitive laboratory paradigms. *Child Development Perspectives*, *4*, 168–173. <http://dx.doi.org/10.1111/j.1750-8606.2010.00137.x>
- Eder, R. A. (1990). Uncovering young children's psychological selves: Individual and developmental differences. *Child Development*, *61*, 849–863. <http://dx.doi.org/10.2307/1130969>
- Else-Quest, N. M., Hyde, J. S., Goldsmith, H. H., & Van Hulle, C. A. (2006). Gender differences in temperament: A meta-analysis. *Psychological Bulletin*, *132*, 33–72. <http://dx.doi.org/10.1037/0033-2909.132.1.33>
- Fagan, J., & Fantuzzo, J. W. (1999). Multirater congruence on the social skills rating system: Mother, father, and teacher assessments of urban Head Start children's social competencies. *Early Childhood Research Quarterly*, *14*, 229–242. [http://dx.doi.org/10.1016/S0885-2006\(99\)00010-1](http://dx.doi.org/10.1016/S0885-2006(99)00010-1)
- Ferdinand, R. F., van der Ende, J., & Verhulst, F. C. (2004). Parent-adolescent disagreement regarding psychopathology in adolescents from the general population as a risk factor for adverse outcome. *Journal of Abnormal Psychology*, *113*, 198–206. <http://dx.doi.org/10.1037/0021-843X.113.2.198>
- Forman, D. R., O'Hara, M. W., Larsen, K., Coy, K. C., Gorman, L. L., & Stuart, S. (2003). Infant emotionality: Observational methods and the validity of maternal reports. *Infancy*, *4*, 541–565. http://dx.doi.org/10.1207/S15327078IN0404_08
- Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, *102*, 652–670. <http://dx.doi.org/10.1037/0033-295X.102.4.652>
- Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology*, *55*, 149–158. <http://dx.doi.org/10.1037/0022-3514.55.1.149>
- Funder, D. C., & Colvin, C. R. (1991). Explorations in behavioral consistency: Properties of persons, situations, and behaviors. *Journal of Personality and Social Psychology*, *60*, 773.
- Funder, D. C., Kolar, D. C., & Blackman, M. C. (1995). Agreement among judges of personality: Interpersonal relations, similarity, and acquaintanceship. *Journal of Personality and Social Psychology*, *69*, 656–672. <http://dx.doi.org/10.1037/0022-3514.69.4.656>
- Garrison, W. T., & Earls, F. (1985). The child behavior checklist as a screening instrument for young children. *Journal of the American Academy of Child Psychiatry*, *24*, 76–80. [http://dx.doi.org/10.1016/S0002-7138\(09\)60413-1](http://dx.doi.org/10.1016/S0002-7138(09)60413-1)
- Gaylord, N. K., Kitzmann, K. M., & Lockwood, R. L. (2003). Child characteristics as moderators of the association between family stress and children's internalizing, externalizing, and peer rejection. *Journal of Child and Family Studies*, *12*, 201–213. <http://dx.doi.org/10.1023/A:1022862816506>
- Goldsmith, H. H., Reilly, J., Lemery, K. S., Longley, S., & Prescott, A. (1993). *Preliminary manual for the preschool laboratory temperament assessment battery*. Department of Psychology, University of Wisconsin-Madison, Madison, WI.
- Grills, A. E., & Ollendick, T. H. (2002). Issues in parent-child agreement: The case of structured diagnostic interviews. *Clinical Child and Family Psychology Review*, *5*, 57–83. <http://dx.doi.org/10.1023/A:1014573708569>
- Guion, K., Mrug, S., & Windle, M. (2009). Predictive value of informant discrepancies in reports of parenting: Relations to early adolescents' adjustment. *Journal of Abnormal Child Psychology*, *37*, 17–30. <http://dx.doi.org/10.1007/s10802-008-9253-5>
- Hall, J. A., Andrzejewski, S. A., Murphy, N. A., Mast, M. S., & Feinstein, B. A. (2008). Accuracy of judging others' traits and states: Comparing mean levels across tests. *Journal of Research in Personality*, *42*, 1476–1489. <http://dx.doi.org/10.1016/j.jrp.2008.06.013>
- Halverson, C. F., Havill, V. L., Deal, J., Baker, S. R., Victor, J. B., Pavlopoulos, V., . . . Wen, L. (2003). Personality structure as derived from parental ratings of free descriptions of children: The inventory of child individual differences. *Journal of Personality*, *71*, 995–1026. <http://dx.doi.org/10.1111/1467-6494.7106005>
- Hampson, S. E. (2008). Mechanisms by which childhood personality traits influence adult well-being. *Current Directions in Psychological Science*, *17*, 264–268. <http://dx.doi.org/10.1111/j.1467-8721.2008.00587.x>
- Hayden, E. P., Klein, D. N., & Durbin, C. E. (2005). Parent reports and laboratory assessments of child temperament: A comparison of their associations with risk for depression and externalizing disorders. *Journal of Psychopathology and Behavioral Assessment*, *27*, 89–100. <http://dx.doi.org/10.1007/s10862-005-5383-z>
- Herzhoff, K., & Tackett, J. L. (2012). Establishing construct validity for openness to experience in middle childhood: Contributions from personality and temperament. *Journal of Research in Personality*, *46*, 286–294. <http://dx.doi.org/10.1016/j.jrp.2012.02.007>
- Hirschmüller, S., Egloff, B., Schmukle, S. C., Nestler, S., & Back, M. D. (2014). Accurate judgments of neuroticism at zero acquaintance: A question of relevance. *Journal of Personality*. Advance online publication.
- Hollingshead, A. B. (1975). *Four factor index of social status*. Unpublished manuscript, Yale University, New Haven, CT.
- Kagan, J. (1998). Biology and the child. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (5th ed., pp. 177–235). New York, NY: Wiley.
- Kenny, D. A., Albright, L., Malloy, T. E., & Kashy, D. A. (1994). Consensus in interpersonal perception: Acquaintance and the big five. *Psychological Bulletin*, *116*, 245–258. <http://dx.doi.org/10.1037/0033-2909.116.2.245>
- Kochanska, G., Barry, R. A., Jimenez, N. B., Hollatz, A. L., & Woodard, J. (2009). Guilt and effortful control: Two mechanisms that prevent disruptive developmental trajectories. *Journal of Personality and Social Psychology*, *97*, 322–333. <http://dx.doi.org/10.1037/a0015471>
- Kochanska, G., Murray, K. T., & Harlan, E. T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology*, *36*, 220–232. <http://dx.doi.org/10.1037/0012-1649.36.2.220>
- Kolar, D. W., Funder, D. C., & Colvin, C. R. (1996). Comparing the accuracy of personality judgements by the self and knowledgeable others. *Journal of Personality*, *64*, 311–337. <http://dx.doi.org/10.1111/j.1467-6494.1996.tb00513.x>

- Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *The American Journal of Psychiatry*, *160*, 1566–1577. <http://dx.doi.org/10.1176/appi.ajp.160.9.1566>
- Kurtz, J. E., & Sherker, J. L. (2003). Relationship quality, trait similarity, and self-other agreement on personality ratings in college roommates. *Journal of Personality*, *71*, 21–48. <http://dx.doi.org/10.1111/1467-6494.t01-1-00005>
- Laidra, K., Allik, J., Harro, M., Merenäkk, L., & Harro, J. (2006). Agreement among adolescents, parents, and teachers on adolescent personality. *Assessment*, *13*, 187–196. <http://dx.doi.org/10.1177/1073191106287125>
- Lee, I. A., & Preacher, K. J. (2013). *Calculation for the test of the difference between two dependent correlations with one variable in common* [Computer software]. Retrieved from <http://quantpsy.org>
- Letzring, T. D., Wells, S. M., & Funder, D. C. (2006). Information quantity and quality affect the realistic accuracy of personality judgment. *Journal of Personality and Social Psychology*, *91*, 111–123. <http://dx.doi.org/10.1037/0022-3514.91.1.111>
- Lindholm, B. W., & Toulaitos, J. (1982). Checklist agreement among observers of children. *Psychology in the Schools*, *19*, 548–551. [http://dx.doi.org/10.1002/1520-6807\(198210\)19:4<548::AID-PITS2310190424>3.0.CO;2-M](http://dx.doi.org/10.1002/1520-6807(198210)19:4<548::AID-PITS2310190424>3.0.CO;2-M)
- Lippa, R. A., & Dietz, J. K. (2000). The relation of gender, personality, and intelligence to judges' accuracy in judging strangers' personality from brief video segments. *Journal of Nonverbal Behavior*, *24*, 25–43. <http://dx.doi.org/10.1023/A:1006610805385>
- Markey, P. M., Markey, C. N., Tinsley, B. J., & Ericksen, A. J. (2002). A preliminary validation of preadolescents' self-reports using the Five-Factor Model of personality. *Journal of Research in Personality*, *36*, 173–181. <http://dx.doi.org/10.1006/jrpe.2001.2341>
- Markon, K. E., Krueger, R. F., & Watson, D. (2005). Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. *Journal of Personality and Social Psychology*, *88*, 139–157. <http://dx.doi.org/10.1037/0022-3514.88.1.139>
- Mason, A. E., Sbarra, D. A., & Mehl, M. R. (2010). Thin-slicing divorce: Thirty seconds of information predict changes in psychological adjustment over 90 days. *Psychological Science*, *21*, 1420–1422. <http://dx.doi.org/10.1177/0956797610381507>
- McLarney-Vesotski, A. R., Bernieri, F., & Rempala, D. (2006). Personality perception: A developmental study. *Journal of Research in Personality*, *40*, 652–674. <http://dx.doi.org/10.1016/j.jrp.2005.07.001>
- Measelle, J. R., John, O. P., Ablow, J. C., Cowan, P. A., & Cowan, C. P. (2005). Can children provide coherent, stable, and valid self-reports on the big five dimensions? A longitudinal study from ages 5 to 7. *Journal of Personality and Social Psychology*, *89*, 90–106. <http://dx.doi.org/10.1037/0022-3514.89.1.90>
- Mervielde, I., De Fruyt, F., & Jarmuz, S. (1998). Linking openness and intellect in childhood and adulthood. In G. A. Kohnstamm, C. F. Halverson, I. Mervielde, & V. Havill (Eds.), *Parental descriptions of child personality: Developmental antecedents of the Big Five?* (pp. 105–126). Mahwah, NJ: Erlbaum.
- Mullineaux, P. Y., Deater-Deckard, K., Petrill, S. A., Thompson, L. A., & Dethorne, L. S. (2009). Temperament in middle childhood: A behavioral genetic analysis of fathers' and mothers' reports. *Journal of Research in Personality*, *43*, 737–746. <http://dx.doi.org/10.1016/j.jrp.2009.04.008>
- Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.
- Oltmanns, T. F., Friedman, J. N., Fiedler, E. R., & Turkheimer, E. (2004). Perceptions of people with personality disorders based on thin slices of behavior. *Journal of Research in Personality*, *38*, 216–229. [http://dx.doi.org/10.1016/S0092-6566\(03\)00066-7](http://dx.doi.org/10.1016/S0092-6566(03)00066-7)
- Ozer, D. J., & Benet-Martínez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology*, *57*, 401–421. <http://dx.doi.org/10.1146/annurev.psych.57.102904.190127>
- Paulhus, D. L., & Bruce, M. N. (1992). The effect of acquaintanceship on the validity of personality impressions: A longitudinal study. *Journal of Personality and Social Psychology*, *63*, 816–824. <http://dx.doi.org/10.1037/0022-3514.63.5.816>
- Prime, H., Perlman, M., Tackett, J. L., & Jenkins, J. M. (2014). Cognitive sensitivity in sibling interactions: Development of the construct and comparison of two coding methodologies. *Early Education and Development*, *25*, 240–258. <http://dx.doi.org/10.1080/10409289.2013.821313>
- Rauthmann, J. F. (2013). Perceiving others is so simple: Implicit simplicity of personality perceptions in a dyadic design. *Personality and Individual Differences*, *54*, 634–639. <http://dx.doi.org/10.1016/j.paid.2012.11.024>
- Rauthmann, J. F., & Kolar, G. P. (2010). Implicit simplicity at low acquaintanceship: Evidence for a g-factor of personality in personality judgments. *Personality and Individual Differences*, *48*, 522–526. <http://dx.doi.org/10.1016/j.paid.2009.11.020>
- Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, *126*, 3–25. <http://dx.doi.org/10.1037/0033-2909.126.1.3>
- Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, *2*, 313–345. <http://dx.doi.org/10.1111/j.1745-6916.2007.00047.x>
- Rothbart, M. K., & Bates, J. E. (2006). Temperament. In N. Eisenberg, & W. Damon, R. M. Lerner (Eds.), *Handbook of child psychology, 6th ed., Vol. 3: Social, emotional, and personality development* (pp. 99–166). Hoboken, NJ: Wiley.
- Rubin, K. H., Hastings, P. D., Stewart, S. L., Henderson, H. A., & Chen, X. (1997). The consistency and concomitants of inhibition: Some of the children, all of the time. *Child Development*, *68*, 467–483. <http://dx.doi.org/10.2307/1131672>
- Rule, N. O., & Ambady, N. (2008). Brief exposures: Male sexual orientation is accurately perceived at 50 ms. *Journal of Experimental Social Psychology*, *44*, 1100–1105. <http://dx.doi.org/10.1016/j.jesp.2007.12.001>
- Rule, N. O., Ambady, N., & Hallett, K. C. (2009). Female sexual orientation is perceived accurately, rapidly, and automatically from the face and its features. *Journal of Experimental Social Psychology*, *45*, 1245–1251. <http://dx.doi.org/10.1016/j.jesp.2009.07.010>
- Rule, N. O., Garrett, J. V., & Ambady, N. (2010). On the perception of religious group membership from faces. *PLoS ONE*, *5*, e14241. <http://dx.doi.org/10.1371/journal.pone.0014241>
- Schmitt, D. P., Realo, A., Voracek, M., & Allik, J. (2008). Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures. *Journal of Personality and Social Psychology*, *94*, 168–182. <http://dx.doi.org/10.1037/0022-3514.94.1.168>
- Shaffer, D., Fisher, P., Lucas, C. P., Dulcan, M. K., & Schwab-Stone, M. E. (2000). NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): Description, differences from previous versions, and reliability of some common diagnoses. *Journal of the American Academy of Child and Adolescent Psychiatry*, *39*, 28–38. <http://dx.doi.org/10.1097/00004583-200001000-00014>
- Sherman, R. A., Nave, C. S., & Funder, D. C. (2010). Situational similarity and personality predict behavioral consistency. *Journal of Personality and Social Psychology*, *99*, 330–343. <http://dx.doi.org/10.1037/a0019796>
- Shiner, R. L. (2000). Linking childhood personality with adaptation: Evidence for continuity and change across time into late adolescence. *Journal of Personality and Social Psychology*, *78*, 310–325. <http://dx.doi.org/10.1037/0022-3514.78.2.310>

- Shiner, R. L., Masten, A. S., & Roberts, J. M. (2003). Childhood personality foreshadows adult personality and life outcomes two decades later. *Journal of Personality, 71*, 1145–1170. <http://dx.doi.org/10.1111/1467-6494.7106010>
- Slobodskaya, H. R. (2011). Two superordinate personality factors in childhood. *European Journal of Personality, 25*, 453–464. <http://dx.doi.org/10.1002/per.810>
- Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of big five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology, 94*, 718–737. <http://dx.doi.org/10.1037/0022-3514.94.4.718>
- Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. *Journal of Personality and Social Psychology, 100*, 330–348. <http://dx.doi.org/10.1037/a0021717>
- Srivastava, S., Guglielmo, S., & Beer, J. S. (2010). Perceiving others' personalities: Examining the dimensionality, assumed similarity to the self, and stability of perceiver effects. *Journal of Personality and Social Psychology, 98*, 520–534. <http://dx.doi.org/10.1037/a0017057>
- Stanger, C., & Lewis, M. (1993). Agreement among parents, teachers, and children on internalizing and externalizing behavior problems. *Journal of Clinical Child Psychology, 22*, 107–116. http://dx.doi.org/10.1207/s15374424jccp2201_11
- Tackett, J. L. (2006). Evaluating models of the personality-psychopathology relationship in children and adolescents. *Clinical Psychology Review, 26*, 584–599. <http://dx.doi.org/10.1016/j.cpr.2006.04.003>
- Tackett, J. L. (2011). Parent informants for child personality: Agreement, discrepancies, and clinical utility. *Journal of Personality Assessment, 93*, 539–544. <http://dx.doi.org/10.1080/00223891.2011.608763>
- Tackett, J. L., Krueger, R. F., Iacono, W. G., & McGue, M. (2008). Personality in middle childhood: A hierarchical structure and longitudinal connections. *Journal of Research in Personality, 42*, 1456–1462. <http://dx.doi.org/10.1016/j.jrp.2008.06.005>
- Tackett, J. L., Slobodskaya, H. R., Mar, R. A., Deal, J., Halverson, C. F., Jr., Baker, S. R., . . . Besevegis, E. (2012). The hierarchical structure of childhood personality in five countries: Continuity from early childhood to early adolescence. *Journal of Personality, 80*, 847–879. <http://dx.doi.org/10.1111/j.1467-6494.2011.00748.x>
- Van den Akker, A. L., Deković, M., Asscher, J., & Prinzie, P. (2014). Mean-level personality development across childhood and adolescence: A temporary defiance of the maturity principle and bidirectional associations with parenting. *Journal of Personality and Social Psychology, 107*, 736–750. <http://dx.doi.org/10.1037/a0037248>
- Vazire, S. (2006). Informant reports: A cheap, fast, and easy method for personality assessment. *Journal of Research in Personality, 40*, 472–481. <http://dx.doi.org/10.1016/j.jrp.2005.03.003>
- Vazire, S. (2010). Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology, 98*, 281–300. <http://dx.doi.org/10.1037/a0017908>
- Vazire, S., & Gosling, S. D. (2004). e-Perceptions: Personality impressions based on personal websites. *Journal of Personality and Social Psychology, 87*, 123–132. <http://dx.doi.org/10.1037/0022-3514.87.1.123>
- Vazire, S., Naumann, L. P., Rentfrow, P. J., & Gosling, S. D. (2008). Portrait of a narcissist: Manifestations of narcissism in physical appearance. *Journal of Research in Personality, 42*, 1439–1447. <http://dx.doi.org/10.1016/j.jrp.2008.06.007>
- Victor, J. B., Halverson, C. F., Jr., & Wampler, K. S. (1988). Family-school context: Parent and teacher agreement on child temperament. *Journal of Consulting and Clinical Psychology, 56*, 573–577. <http://dx.doi.org/10.1037/0022-006X.56.4.573>
- Wood, D., Harms, P., & Vazire, S. (2010). Perceiver effects as projective tests: What your perceptions of others say about you. *Journal of Personality and Social Psychology, 99*, 174–190. <http://dx.doi.org/10.1037/a0019390>
- Zupančič, M., Sočan, G., & Kavčič, T. (2009). Consistency in adult reports on child personality over the pre-school years. *European Journal of Developmental Psychology, 6*, 455–480. <http://dx.doi.org/10.1080/17405620701439887>

Received December 1, 2014

Revision received March 6, 2015

Accepted March 20, 2015 ■