NIH Public Access
Author Manuscript
Psychol Assess. Author manuscript; available in PMC 2012 June 1.

Published in final edited form as:
Psychol Assess. 2011 June ; 23(2): 337–353. doi:10.1037/a0021746.

Deriving Childhood Temperament Measures from Emotion-eliciting Behavioral Episodes: Scale Construction and Initial Validation

Jeffrey R. Gagne[1], **Carol A. Van Hulle**[1], **Nazan Aksan**[2], **Marilyn J. Essex**[1], and **H. Hill Goldsmith**[1]

[1] University of Wisconsin-Madison

[2] Koc University

Abstract

The authors describe the development and initial validation of a home-based version of the Laboratory Temperament Assessment Battery (Lab-TAB), which was designed to assess childhood temperament using a comprehensive series of emotion-eliciting behavioral episodes. This paper provides researchers with general guidelines for assessing specific behaviors using the Lab-TAB and for forming behavioral composites that correspond to commonly researched temperament dimensions. We used mother ratings and independent post-visit observer ratings to provide validity evidence in a community sample of 4.5 year-old children. 12 Lab-TAB behavioral episodes were employed, yielding 24 within-episode temperament components that collapsed into 9 higher-level composites (Anger, Sadness, Fear, Shyness, Positive Expression, Approach, Active Engagement, Persistence, and Inhibitory Control). These dimensions of temperament are similar to those found in questionnaire-based assessments. Correlations among the 9 composites were low to moderate, suggesting relative independence. As expected, agreement between Lab-TAB measures and post-visit observer ratings was stronger than agreement between the Lab-TAB and mother questionnaire. However, for Active Engagement and Shyness, mother ratings did predict child behavior in the Lab-TAB quite well. Findings demonstrate the feasibility of emotion-eliciting temperament assessment methodologies, suggest appropriate methods for data aggregation into trait-level constructs, and set some expectations for associations between Lab-TAB dimensions and the degree of cross-method convergence between the Lab-TAB and other commonly used temperament assessments.

Keywords

Temperament; Children; Behavioral Assessment; Scale Construction; Laboratory Temperament Assessment Battery (Lab-TAB)

Correspondence concerning this article should be addressed to Jeffrey R. Gagne, Department of Psychology, University of Wisconsin, 1202 West Johnson Street, Madison, WI 53706-1611., jgagne@wisc.edu.
Jeffrey R. Gagne, Department of Psychology, University of Wisconsin; Carol A. Van Hulle, Department of Psychology, University of Wisconsin; Nazan Aksan, Department of Psychology, Koc University; Marilyn J. Essex, Department of Psychiatry, University of Wisconsin; H. Hill Goldsmith, Department of Psychology, University of Wisconsin. Nazan Aksan is now at Department of Psychology, University of Iowa.

Publisher's Disclaimer: The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at www.apa.org/pubs/journals/HEA

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Temperament traits are conceptualized as behavioral and emotional dimensions that develop early in childhood and collectively form the basis for later personality (Goldsmith et al., 1987). Childhood temperament is often assumed--and to a degree has been demonstrated--to reflect biological individuality, to remain relatively stable across development, to have less cognitive, affective, and social complexity than personality traits, and to include self-regulatory as well as reactive components (Buss & Plomin, 1975; 1984; Goldsmith et al., 1987; Rothbart, 1989; Rothbart & Bates, 1998; Thomas & Chess, 1977). Research on childhood temperament intersects with several subdisciplines in psychology, including personality, developmental, clinical, and behavioral neuroscience (Gagne, Vendlinski & Goldsmith, 2009). The connections between early temperament traits and later personality and psychopathology have been intensely studied over the last decade (Caspi, Roberts & Shiner, 2005; Goldsmith, Lemery & Essex, 2004). However, unlike personality research, which typically employs interview and questionnaire methods, temperament research has historically been rooted in observational and laboratory-based assessment techniques.

Although even prehistoric humans probably invoked the concept of temperament in dealing with one another, the Greek physician Galen (129–199 CE) is generally credited with creating the first influential, systematic account of how aspects of temperament corresponded with specific personality profiles and diseases. At the turn of the 20th century, the Russian physiologist Pavlov studied “transmarginal inhibition” in canines. Pavlov and his followers observed reactive and regulatory aspects of the canine nervous system, phenomena that overlap with current conceptualizations of temperament. In the early 20th century, the German psychiatrist Kretschmer believed that body type was related to temperament and that the extremes of thinness and obesity were differentially related to distinct psychiatric disorders. In the late 1940s and 1950s, Escalona employed detailed and intensive observational methods to examine individual differences in personality development in a sample of healthy infants (Escalona & Leitch, 1952). Although she focused on psychoanalytic conceptions of personality and did not explicitly use the term temperament in her work, Escalona’s observations of early emerging patterns of emotions and behavioral repertoires laid some of the groundwork for contemporary temperament research. Although conceptions and theories changed over time, these early theorists all emphasized an observational quality to the study of temperament.

Thomas and Chess began the classic New York Longitudinal Study (NYLS; Thomas, Chess & Birch, 1968) of infant temperament in the early 1950s; they used ratings of nine temperament dimensions based on parent interviews and, later, questionnaires. Children with high or low ratings on these dimensions were considered at risk for poor psychological adjustment. This approach to conceptualizing temperament as a dimensional construct had a significant impact on temperament theory and research in the latter half of the 20th century. The original Thomas and Chess temperament dimensions were activity, regularity, initial reaction, adaptability, intensity, mood, distractibility, persistence/attention span, and sensitivity. Currently, the most commonly examined temperament dimensions are activity level, anger/frustration, behavioral inhibition/fear, effortful control, and positive affect (Gagne et al., 2009). Although these dimensions are not representative of any single theoretical framework regarding the structure of temperament, researchers have reached some consensus about the importance of these widely studied traits.

In addition to dimensional traits, temperament can also be conceptualized categorically or typologically. Thomas and Chess included “easy,” “difficult,” and “slow-to-warm-up” categories in their theory of temperament, and other behavioral scientists have also conceptualized temperament categorically. Meehl (1992) and Kagan (1994), among others, emphasized that the issue of dimensions versus types is conceptually complex, and that types remain a viable yet understudied alternative to temperamental dimensions. Although

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

our research group has described typological approaches to temperament using configural frequency analysis (Aksan et al., 1999), this paper adopts a dimensional conceptualization.

Several temperament researchers have designed theoretically comprehensive questionnaires covering multiple temperament dimensions; these efforts are similar to approaches in the personality domain that emphasize multiple traits. In some cases, temperament researchers have adapted the popular Five-Factor Model from adult personality research for use with children (Kohnstamm, Halverson, Mervielde, Havill, 1998). In another approach, the 15 scales of the Children's Behavior Questionnaire have been factor analyzed to yield three factors (Putnam & Rothbart, 2006). Typically, temperament measures are derived in a top-down fashion and questionnaire items reflect the concerns of developmental psychologists (e.g., Buss & Plomin, 1977; 1984; Gartstein & Rothbart, 2003; Goldsmith, 1996; Putnam, Gartstein & Rothbart, 2006) or clinicians (Thomas & Chess, 1977). Many temperament taxonomies and factor structures of temperament dimensions have originated from these questionnaire-based methods. Even when factor-analytic methods are used to determine dimensionality, the resulting structures reflected the temperament theory that informed item-generation (Kohnstamm et al., 1998).

Most studies use parental rating scales as the primary means of assessing temperament, typically without direct observation as a check on the validity of the questionnaire measures. Unfortunately, using parent ratings as the sole basis for assessment of child temperament introduces the potential for several biases. Specifically, parent ratings of temperament in studies with siblings often show contrast effects whereby the parent exaggerates differences between siblings or assimilation effects whereby the parent exaggerates similarities between siblings. Parent-rated temperament often evinces higher age-to-age stability than lab-based ratings, suggesting that parent expectations about child behavior may contribute to this stability (Saudino, 2003a). However, it is also possible that lab ratings are not as reliable or valid as parent ratings, or that the two types of assessment tap different aspects of behavior. Another criticism of parent ratings include the possibility of "halo effects" or "cries for help," in which case the parents' affective relationship with the child biases reports of temperament. Yet another criticism is that parents lack knowledge of the typical level and range of behavior of large, representative comparison groups of same-age children and are thus unable to accurately scale their own children's behavior (see Mangelsdorf, Schoppe, & Buur, 2000, for further discussion of this issue).

A special issue in *Infant Behavior and Development* (2003) focused on this important topic of parents as informants about the temperament of their own children (Goldsmith & Hewitt, 2003; Hwang & Rothbart, 2003; Saudino, 2003a; Saudino, 2003b; Seifer, 2003). These researchers also indicated that some parent assessments of temperament were less susceptible to rater biases than others, particularly rating systems that do not rely on global judgments of behavior and focus on specific, concrete behaviors. For instance, certain temperament questionnaires (e.g., the Child Behavior Questionnaire) employ specific content and time frames in items that allow parents to access more specific memories of their child's behavior (e.g., "has difficulty sitting still at dinner"). Measures that use global items such as "my child is highly active" appear to leave more room for biases to intrude. Although we need more studies that empirically compare different types of parent ratings, the issue of contrast effects in twin and family research clearly occurs most often in global questionnaire ratings (Goldsmith & Hewitt, 2003; Hwang & Rothbart, 2003; Saudino, 2003a).

Although most temperament studies have relied on global parent reports for behavioral assessment, several researchers have advocated the use of laboratory methods and some suggest that a multi-method perspective provides the best evidence for the significance of

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

temperament to important developmental outcomes (Hwang & Rothbart, 2003). However, the literature on early temperament clearly shows that parental report and observational measures of the putatively "same" temperament trait often lack substantial agreement (Goldsmith, Rieser-Danner & Briggs, 1991; Mangelsdorf et al., 2000; Saudino & Cherny, 2001; Saudino, Wertz, Gagne & Chawla, 2004; Seifer, Sameroff, Barrett & Krafchuck, 1994). Conceptually, this lack of agreement might result from flaws in one or both of the assessment approaches (which seems to be the default assumption in the literature) or from fundamental differences in the features of temperament captured by the two approaches. That is, the nature of, say, anger proneness assessed by independent observers may differ from the nature of anger proneness that parents report. The plausibility of this latter explanation must be adjudicated empirically. For instance, Saudino (2009) showed that parent-rated activity level tapped different genetic and environmental factors than lab-based and mechanical ratings of activity level in toddlers. Other studies have shown that parent and lab ratings have differential correlates to outcomes such as maternal depression (Gartstein & Marmion, 2008; Hayden, Klein, & Durbin, 2005). Whether weak convergence of parental report questionnaires and observational measures is due to different sources of systematic error or to these two assessment approaches tapping partially different temperament traits, or both, the advantage of employing carefully constructed observational ratings in addition to parent report is apparent.

Those who include lab-based assessments of temperament in their programs of research often focus on one specific dimension. For example, Kagan's research on reactively fearful children primarily used objective laboratory fear and inhibition paradigms (Kagan, 1994). Saudino and Eaton have produced a series of studies examining activity level using a combination of parent, laboratory and mechanical actometer measures (Saudino & Eaton, 1991; 1995). Kochanska's research on effortful control and inhibitory control also combined parent and lab-based assessments of temperament (Kochanska, Murray, & Harlan, 2000; Kochanska, Murray, Jacques, Koenig, & Vandegeest, 1996). Measuring a single dimension at a time allows researchers to be efficient. However, there is a natural trade-off between depth and breadth of measurement in observational research. The narrower approach needs to be balanced by research that takes a broader and more integrative view or many important aspects of temperament will not be properly understood. Such research would be more likely to be conducted if there were a set of easily available and well-validated observational assessment tools and protocols that can be used to measure a broad range of temperament-related variables.

Some researchers have assessed temperament more broadly by examining a wider range of dimensions in the lab. Most of these assessments were relatively free flowing in administration and were often conducted in conjunction with cognitive and motor testing in early childhood. In the early stages of the Louisville Twin Study, Matheny (1980, 1983) used observational measures of child temperament based on the items from Bayley's Infant Behavior Record (IBR; Bayley, 1969). The IBR is used to assess a range of temperament-like infant behaviors that are observed in the testing situation of the Bayley Scales of Infant Development (Bayley, 1969). A factor analysis of the IBR suggests that it measures task orientation, affect-extraversion, activity, auditory-visual awareness, and motor coordination. It is noteworthy that these assessments were not focused on eliciting specific temperament-related behavior. With the IBR and similar measures, child testers or observers provide global ratings based on impressions of the child's behavior during the testing situations that targeted other characteristics.

A few investigators have developed more fine grained methods for measuring and rating temperament in an observational setting. Rothbart (1986) assessed activity level, smiling and laughter, distress to limitations, fear, and vocal activity during bath, feeding and play

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

situations in a home visit with infants and their parents; the frequency and intensity of target behaviors, as well as contextual information regarding the situations and the behavior of the parent was obtained. Rothbart (1988) also assessed similar traits in structured tasks in the laboratory setting. This work of Rothbart's was one direct precursor of the approach reported in this paper (Goldsmith & Rothbart, 1991). In later Louisville Twin Study work than that described above, a series of behavioral vignettes (i.e., visible barrier with attractive toy, puppet, and mechanical toy games) were employed in the laboratory to derive scores of emotional tone, activity, attentiveness, and social orientation as judged by observers across all the episodes (Riese, 1998). Although the ratings and coding systems were more sophisticated than the IBR, these assessments were also not designed to elicit particular behaviors or to obtain a specific structure of temperament.

In addition to Rothbart's (1988) research mentioned above, a handful of studies have employed laboratory situations or episodes that evoke and assess multiple specific dimensions of temperament. For example, in another study of infant temperament and emotion that was a direct precursor to the methods reported in the current paper, Goldsmith & Campos (1990) assessed fearfulness and joy/pleasure using several laboratory vignettes. Fear was assessed in two visual cliff episodes and one stranger approach episode, and joy was assessed in a series of four game-like episodes. Subsequent studies from Goldsmith's group elaborated upon this approach (e.g., Pfeifer, Goldsmith, Davidson, & Rickman, 2002). In the MacArthur Longitudinal Twin Study (Robinson, McGrath & Corley, 2001), researchers included distinct episodes that elicit anger (restraint and toy removal), prohibition/inhibition (prohibition of touching an attractive toy), and behavioral inhibition (stranger approach) during home and laboratory visits with young children. Despite these examples, more comprehensive laboratory-based investigations of child temperament are the exception rather than the rule, and the assessments in the studies just reviewed were not usually designed to develop a taxonomic structure.

Developing valid and reliable objective temperament assessment methods will allow us to complement parent rating scales and return to the tradition exemplified by the early work of Pavlov and Escalona. As previously mentioned, objective assessment can take the form of free-flowing, unstructured observations or structured tasks and episodes that are designed to elicit specific behavioral responses and quantify individual differences in those responses. The purpose of this paper is to examine the psychometric properties of a comprehensive, in-home behavioral assessment of temperament in preschool-age children. The in-home assessment was modified from the Preschool version of the Laboratory Temperament Assessment Battery (Lab-TAB; Goldsmith, Reilly, Lemery, Longley & Prescott, 1993), a laboratory-based temperament measure that includes several behavioral episodes corresponding to the full range of temperament dimensions. We refer collectively to the episodes in this study as the Lab-TAB, although administration is in the home rather than the lab context. Although we expect there to be some variance in behavioral observation from place-to-place, most all aspects of administration and coding are consistent across the lab- and home-based versions of the Lab-TAB. The great majority of Lab-TAB episodes were originally intended to tap a single dimension of temperament; however, behavioral reactions in some episodes appear to be influenced by multiple traits. In the present study, we selected a community sample and used mother ratings and independent post-visit ratings to provide validity evidence for our home-based Lab-TAB assessment.

Much of what we know about child temperament assessment has been derived from questionnaire methodology, and the implications for temperament assessment using a lab-based methodology are not straightforward. For questionnaires, the basic unit of information is the item, and items are systematically organized into scales that reflect particular dimensions. Well known standards and practices have been developed to optimize item

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

features, how items should be combined to form scales, and what inter-relationships are to be expected among scales (Loevinger, 1957; Clark & Watson, 1995). Because observational assessments of temperament have rarely been organized into comprehensive protocols, little consensus exists on best practices and standards for assessing specific behaviors, how to form behavioral composites, and expectations for associations between dimensions. The same psychometric principles needed to ensure reliability and foster validity apply to both questionnaire and observational assessments, but practical assessment guidelines may be very different for the two modalities as we shall describe in the manuscript.

A central question is whether the dimensions of temperament that can be recovered from the Lab-TAB are similar to temperament questionnaire scales. If lab-based dimensions are comparable in nature to those derived from questionnaires, will correlations from "corresponding" dimensions show convergent validity across the methods? Based on previous findings, it is unlikely that the convergence between mother ratings and the Lab-TAB measures will be strong. However, we anticipate that agreement will be comparable to the typical levels of parent-laboratory agreement noted in the literature (Goldsmith et al., 1991; Mangelsdorf et al., 2000; Saudino & Cherny, 2001; Saudino et al., 2004; Seifer et al, 1994). Convergence between Lab-TAB measures and post-visit observer ratings will most likely be stronger than mother questionnaire by Lab-TAB convergence because both the post-visit observers and the Lab-TAB coders will be rating child behavior from the overlapping contexts. The purpose of this paper is not to promote the Lab-TAB assessment for general use, but to demonstrate the feasibility of explicitly behaviorally based temperament assessment methodologies, to suggest methods for data aggregation into trait-level constructs, and set some expectations for the degree of cross-method convergence that such assessments are likely to evince. We do not claim the superiority of objective laboratory ratings, but we offer an alternative or an adjunct to standard assessment methods.

Methods

Sample

The sample consisted of 408 children assessed at 4.5 years of age. Families resided in either the Milwaukee (78%) or Madison, Wisconsin, metropolitan areas (Hyde, Klein, Essex & Clark, 1995). This was a study of maternity leave and health outcomes, therefore all mothers were required to be employed or a full-time homemaker. Participants were selected during the second trimester of pregnancy with the target child, and only mothers who were cohabiting with the baby's biological father and older than 18 years of age were included. Of the 570 families that were initially recruited, 451 remained in the study at 4.5 years when the home-based assessment of temperament was conducted, although not all families participated in this aspect of the study. Inclusion criteria for this study focused on the mothers, and there were no exclusionary criteria for children with disabilities (e.g., Fragile X, Down's Syndrome). However, none of the children who participated in the 4.5 year assessment had any major disabilities. There were nearly equal numbers of females (228) and males (223), and 93% of the mothers identified themselves as Caucasian (2.4% African-American, 1.8% Hispanic, 2% Indian-Alaskan, and 0.7 % Asian). Approximately 44% of the children in the study participated in center-based day care at least 8 hours per week, 17% were in home-based day care, and the remaining children were not in day care (or data were missing). At the birth of the target child (1990–1991), the average age of mothers and fathers was 29 and 32 years, respectively. Thirty-eight percent of the children were first-born, 37% were second-born, 19% third-born and the remaining 6% were later born. At the time of recruitment, 36.8% of the participating mothers and 40.5% of the fathers had a high school or technical degree or less, and the remainder had a college degree or post-college education. At the time of pregnancy for the target child, 57% of participants had combined family incomes of less than $50,000 per year.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Laboratory Temperament Assessment Battery - Home Version (Lab-TAB)

During home visits, child temperament was assessed using the Lab-TAB–Home Version, a comprehensive home-based temperament assessment that includes behavioral episodes corresponding to dimensions of temperament. This administration of the Lab-TAB included 12 standardized behavioral episodes intended to elicit targeted affective and behavioral reactions. These episodes comprised a revised version of the Preschool Lab-TAB (Goldsmith et al., 1993). Preschool Lab-TAB data has previously shown convergent validity with temperament questionnaires (i.e., the Children's Behavior Questionnaire) as well as continuity across age in several studies (e.g., Pfeifer et al., 2002). The administration and coding of this home-based version was almost identical to the lab versions. The only significant differences were that the child was in the home, the camera was present as opposed to being in a control room, and the props that the experimenter used were in a bag in the room. Although there is always variance in behavioral observation from place-to-place (including from laboratory to laboratory), our goal was to make the administration as consistent as possible across home administrations. The Lab-TAB assessments typically lasted about 40 minutes and occurred at the midpoint of a two-hour home visit that included other activities such as maternal interviews and play sessions with a sibling. One child tester administered the battery after establishing appropriate rapport with the child. Eight individuals served as child testers for the 4.5 year visits, and each was highly trained and monitored to achieve consistency of administration. During administration of the Lab-TAB episodes, children's behavior was videotaped and later coded in the laboratory. Thirteen percent of the sample was rated by a second observer and every Lab-TAB episode had a mean Kappa of .90 or higher reflecting chance-corrected inter-rater agreement.

Table 1 lists the 12 Lab-TAB episodes, beginning with a descriptive title that will be used throughout the paper (Bookmark, Box Empty, Dinky Toys, End of the Line, Perpetual Motion, Popping Bubbles, Pop-up Snakes, Snack Delay, Spider, Stranger Approach, Transparent Box, Workbench). Brief descriptions of the episodes as well as the broad domains of temperamental reactions that each episode was intended to elicit are also noted. Within each episode epoch or trial, a number of child responses are coded. Lab-TAB coding involves multiple domains of responses including facial, vocalic, motoric, behavioral and postural modalities (e.g., smiling, reaching, crying, touching, or changes in facial expression). Sometimes the presence or absence of a response is simply noted; however, more often parameters of the response, such as latency, duration, and intensity, are timed or rated. Expressive (e.g., facial and vocal) measures and instrumental or motoric measures often fall into different clusters and can be classified as different episode component scores. For example, the Box Empty episode yields anger, sadness, and approach scores. If episode component scores are intercorrelated, an overall episode summary score is often justified. However, we used many episode-level component scores in the present analyses. The actual process of scoring and scale construction for the Lab-TAB is a key element of our results and will be described in the next section.

Post-visit Observer Temperament Ratings

In addition to the Lab-TAB assessments, the child tester and the person who videotaped the child during the entire visit completed post-visit ratings of child temperament. These ratings were conducted independently by the two observers after reviewing the videotape immediately upon return to the project offices after the home visit. These post-visit observer ratings included 28 items rated on a 1–5 scale. The items were intended to be unipolar, with "1" describing the absence of the quality being rated (e.g., positive affect, energy, cooperation), and "5" describing the intense, consistent, and/or extreme reaction. For example, for the impulsivity rating a score of "1" indicated "no sign of impulsivity, ever", "2" indicated "only slight or ambiguous signs of impulsivity; but restrained quickly", "3"

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

indicated “unambiguous tendency towards impulsivity; often shows some restraint”, “4” indicated “typically impulsive; may show signs of restraint in 1 or 2 situations”, and 5 indicated “consistently impulsive, shows little, if any, inhibition.” Sixteen of the 28 post-visit rating items overlapped with similar items on the Behavior Rating Scales (BRS) from the Bayley Scales of Infant Development-II (Bayley, 1993), although slight modifications were made for age appropriateness (see Goldsmith, 1978, for more detail). Use of Bayley Scale items to assess temperament yields interrater reliability estimates of 87% (Goldsmith & Gottesman, 1981; Matheny, 1980). The other items included four ratings of child’s behavior with the mother and eight items constructed to test for convergent validity for behaviors scored during the Lab-TAB episodes. The post-visit rating variables that we used in our analyses were Anger Proneness, Sadness, Fear, Shyness, Positive Affect, Exuberance, Hyperactivity, Persistence and Impulsivity. For all items, a κ of .85 or higher was obtained.

Children’s Behavior Questionnaire (CBQ)

An abridged 80-item version of the Children’s Behavior Questionnaire (CBQ; Rothbart, Ahadi & Hershey, 1994) was completed before the home visit by the mother. The CBQ requires parents to judge their children’s reactions to a variety of situations over the last six months (e.g., “Can lower his/her voice when asked to do so”) and is appropriate for children from 3 to 7 years of age (Rothbart, Ahadi, Hershey & Fisher, 2001). Each item is rated on a 1–7 scale with 1 indicating the reaction is “extremely untrue” of the child and 7 indicating that the reaction is “extremely true.” CBQ scores have shown high internal consistency, parental agreement and convergent validity with socialization-relevant traits (Rothbart et al., 2001) and have been used in numerous studies with a wide range of empirical correlates. At the time the CBQ was completed, the mothers had no knowledge of how their children would behave during the Lab-TAB assessment. The eight CBQ scales that we used were selected for overlap with temperament dimensions assessed in the Lab-TAB, and each CBQ scale had 10 items. Estimates of internal consistency for each CBQ scale were as follows: Anger (α = .78), Fear (α = .73), Shyness (α = .92), Sadness (α = .63), Approach (α = .74), Activity Level (α = .73), Attentional Focusing (α = .78), and Inhibitory Control (α = .82).

Data Analysis

Approach—This article reports on the initial steps taken in deriving scores from the Lab-TAB–Home Version as well as evidence for convergent validity. An average score approach to data reduction and composite construction was employed, beginning with raw scores from each Lab-TAB episode. We used average as opposed to differentially weighted scores because average scores tend to replicate better and there is typically very little practical difference when the two types of scores are related to external variables. Although differentially weighted scores are more precise than average scores, the differential weights are usually sample-specific. We based our validity evidence on convergence with the post-visit observer ratings and the CBQ scales.

Imputation—Prior to analyses, we used imputation to avoid biases due to missing data (Graham, 2009). A total of 408 children completed at least most of the Lab-TAB episodes at age 4.5 years. Out of this group of 408, 382 children (94%) had complete Lab-TAB, post-visit observer rating, and CBQ data. Eleven children were missing the CBQ, 6 were missing the Stranger Approach episode of the Lab-TAB, and 9 were missing scattered single variables. We used Little’s Missing Completely at Random Test (MCAR) to determine whether these data were missing at random. Little’s MCAR is a chi-square test, whereby if the *p* value is not significant, then the data may be assumed to be MCAR. The resulting MCAR score for our data (χ^2 = 187.25, *df* = 195, *p* = .64) indicates that we cannot reject the hypothesis that the data is MCAR. Based on this finding and the relatively small amount of missing data, we had sufficient justification to impute the missing values in our dataset using

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

the SPSS Missing Value Analysis expectation maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977). After this single imputation procedure, neither means nor standard deviations of any study variables differed in the 1st or 2nd decimal place from their values before imputation. Therefore, all data analyses used the complete set of data, with imputed values where necessary, for all 408 participants.

Results

Construction of the Lab-TAB Temperamental Composite Measures

Each of the 12 Lab-TAB episodes used in the home administration at 4.5 years yields substantial raw data. Multiple response categories (ranging from 2 to perhaps 8–10) are scored in each episode. Typically, we scored the latency of the first response, the occurrence/non-occurrence of the response in each scoring interval of 5–10 seconds (or, in some cases, each discrete trial), and, in most cases, the magnitude or intensity of each response. These responses are depicted in the top level of Figure 1 (this figure should be read and understood in conjunction with Table 2). From these raw data, we derived the latency or speed, the mean level of response, and the peak intensity of the response. Latency, mean level, and peak are referred to as parameters of response. Sometimes, the simple occurrence of a behavior was noted. The raw data points can range from 12–197 per episode.

We computed descriptive statistics for each directly coded variable (i.e., parameters of response) to detect variables with low variance or markedly non-normal distributions. Then, the lowest-order composite variables were computed at this stage (i.e., means within the four parameters of a given response: occurrence, intensity, peak, and latency); this level is referred as the “response parameter” level in Figure 1. In cases where distributions were skewed, appropriate transformations were made. Most transformations involved either a reciprocal square root function for latency measures to transform them to speed measures or a square root function for other positively skewed variables. Because later steps in the process would involve combining measures that were scored in different metrics, we also applied z-transformations to each variable.

The next step in data reduction was to combine correlated parameters of the same response within an episode (e.g., latency, mean level, and peak angry facial expression). These response parameter level composites were then examined for covariation with other response parameter level composites, still within the episode (e.g. postural anger and facial anger). Often, principal component analysis, common factor analysis, or simply examination of the correlation matrices suggested that a higher-level, within-episode component was justified. For example, “facial anger,” “postural anger,” and “protest” lower-level composites could be combined into an “anger” component within the End-of-the-Line episode. These episode level components (also depicted in Figure 1) were always calculated by taking the mean of the constituent measures, with means still being calculated in the few cases where one or more of the constituent measure was missing. Not every measure in the raw data was used in a component. The subsets of items used in each episode level component, the components themselves, and the internal consistency estimates for the components are presented for each episode in Table 2.

As can be seen from Table 2, this process yielded a total of 24 within-episode components for the 12 episodes. For example, this process yielded anger and sadness components within Box Empty, End-of-the-Line, and Transparent Box episodes. Fear (object) was derived from the Jumping Spider episode and Shyness (social fear) was derived from the Stranger Approach. The Positive Expression components were derived from Bookmark, Popping Bubbles and Pop-up Snakes episodes; Approach (anticipatory positive affect) components

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

were derived from Box Empty, Perpetual Motion, Popping Bubbles, and Pop-up Snakes episodes; and the Active Engagement components were derived from both Workbench and Bookmark episodes. Finally, Persistence components were derived from Perpetual Motion and Transparent Box episodes, while Inhibitory Control components were derived from Snack Delay and Dinky Toy episodes. The internal consistency or alpha estimates for these 24 episode specific components are presented in Table 2. Twenty of these 24 components had internal consistencies in the range of .70 to .90. Furthermore, components from the same episode such as anger and sadness components from the Box Empty episodes often showed low correlations indicating that the components were tapping largely independent sources of variance.

Because our interest was mainly in behavioral tendencies that transcend any one episode or situation, in the third step we calculated higher-level "temperamental" composites, which are shown at the bottom level of Figure 1. In other words, we were more confident in designating a measure as "temperamental" when it exhibited cross-situational consistency. The empirical starting point for forming these higher-level composites was the correlation matrix of the 24 within-episode components (see Appendix). The conceptual starting point was our theoretical view of temperament as involving emotional and regulatory dimensions of behavior. Some expectations were also set by earlier preschool Lab-TAB results (Pfeifer et al., 2002). Intercorrelations among the component measures from each episode indicated that the degree of overall shared variance from one episode to the next ranged from low to moderate. Thus, in forming these higher-level temperamental composites, similar responses from different episodes were unit weighted and averaged. For example, sadness components from all three negative affectivity episodes were averaged to form the Sadness composite.

Descriptive Statistics

Table 3 lists the means and standard deviations of all temperament dimensions in the study for both males and females. For Lab-TAB dimensions, males had significantly higher levels of anger, approach, active engagement, and persistence than females, and females were higher on inhibitory control. Females had higher sadness and shyness scores on the post-visit ratings and males had higher exuberance, hyperactivity and impulsivity scores. On the CBQ, females had higher sadness, attentional focusing and inhibitory control, and males had higher activity level. Gender differences are fairly consistent across the three types of ratings and follow patterns that are typically found in the literature.

Interrelations among the Lab-TAB Temperamental Dimension Composites

The intercorrelations between the Lab-TAB temperamental composites are displayed in Table 4. In general, correlations between the composites were low to moderate. The patterns of covariance shown are typical of child temperament dimensions in the literature. For example, correlations in both the positive affectivity domain (Positive Expression, Approach and Active Engagement) and the regulation domain (Inhibitory Control and Persistence) were moderate, ranging from .20–.56. On the other hand, correlations in the negative affectivity domain (Anger, Sadness, Fear, and Shyness) ranged from −.12 to .10, suggesting that negative emotions are more distinct from one another in early childhood than are positive emotions[1]. Anger was positively related to the positive affectivity dimensions, presumably due to common approach motivation, and Anger was negatively associated with Shyness and the regulatory dimensions. Inhibitory Control showed negative associations with Positive Affect and a positive correlation with Shyness.

We used Bartlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy to determine the factorability of the intercorrelation matrix shown in Table 4. Bartlett's test of sphericity calculates the determinant of the matrix of the sums of

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

products and cross-products (S) from which the intercorrelation matrix is derived. The determinant of the matrix S is converted to a chi-square statistic and tested for significance. The null hypothesis is that the intercorrelation matrix comes from a population in which the variables are non-collinear (i.e., an identity matrix). If two variables share a common factor with other variables, their partial correlation will be small. If the variables in the matrix are measuring a common factor, the KMO value should be close to 1.0. If the variables are not measuring a common factor, the KMO will be closer to 0.0. The results of the Bartlett's test ($\chi^2 = 454.06$, $df = 36$, $p < .0001$) indicates that the matrix departs significantly from an identity matrix. However, the KMO of .715 suggests that the degree of common variance among the measures in Table 4 is marginal (a threshold of .70 is often considered the minimal level needed to justify factor analysis of the matrix). In summary, the results of these two tests are somewhat equivocal as to whether factor analyzing the matrix of Lab-TAB variables is fully justified; we interpret this result to suggest that the primary trait level of analysis is preferable for interpretation.

Correlations between the Lab-TAB Composites and Other Temperament Measures

Post-visit Observer Temperament Ratings—Table 5 shows the intercorrelations between the Lab-TAB composite measures and the conceptually matching cross-situational variables derived from the post-visit observer ratings. All of these correlations were modest to moderate in magnitude, and many provide a degree of convergent validation for scores on the composite Lab-TAB measures, which were derived from microanalytic coding of specific situations. The sadness, persistence, and active engagement by hyperactivity correlations were all under .30, indicating somewhat lower convergence. The negative association between Lab-TAB Inhibitory Control and post-visit ratings of Impulsivity shows that children with poor inhibitory behavior have increased levels of impulsivity.

CBQ Temperament Dimensions—To further examine the convergent validity of these composite Lab-TAB scores, we used maternal reports on eight scales from the CBQ. The correlations between cross-situational composites from the Lab-TAB and CBQ are presented in Table 6. Eight of the cross-situational Lab-TAB dimensional composites have a closely corresponding CBQ scale, and four of these Lab-TAB composites have correlations of small magnitude with the corresponding CBQ scale (Fear, Shyness, Inhibitory Control, and Active Engagement). In contrast, Sadness, Anger, Approach, and Persistence did not correlate significantly with the corresponding CBQ scales. The pattern of correlations

[1]Given that the pattern of correlations suggests that negative emotions (anger, sadness, fear, and shyness) are more distinct from one another in early childhood than are positive emotions (positive expression, approach, and active engagement), we sought to formally test the idea that coherence within positive affect may be greater than coherence within negative affect dimensions. Because our N was large, any formal test of equality in subsets of correlations speaking to coherence within and across those two domains is overpowered. Hence, we decided to test two sets of pattern hypotheses against a null pattern that assumed uniformity in coherence within and across positive and negative affect dimensions. In this baseline null model, we assumed that the coherence within positive affect and coherence within negative affect dimensions would be similar to each other, $\chi^2(8) = 119.35$, p = .00 (RMSEA 90% CI = .15–.21; CFI = .64; AIC= 154.16), i.e. correlations among the 3 positive affect variables and correlations among the 4 negative affect variables were set equal to each other. We tested this model against 2 alternative models.
In model A, the correlations were set equal only within the positive affect domain and all other correlations were freely estimated, assuming coherence within the positive affect domain only. The fit of this model was inadequate, χ^2 (2) = 43.41, p = .00 (RMSEA 90% CI= .18–.29; CFI = .87; AIC= 98.09), but it represented a significant improvement in fit over the baseline model with equality in coherence within and across positive and negative affective domains, $\Delta \chi 2$ (6) = 75.89, p = .00. In model B, the correlations were set equal only within the negative affect domain and all other correlations were freely estimated, assuming coherence within the negative affect domain only. The fit of model B was also inadequate, χ^2 (5) = 13.84, p = .02 (RMSEA 90% CI= .12–.11; CFI = .97; AIC= 56.69). However, Model B represented a significant improvement in fit over the baseline model with equality in coherence within and across positive and negative affective domains, $\Delta \chi^2$ (3) = 105.46, p = .00. However, Model B is not nested within model A, which incorporates equality in coherence only in positive affect. The chi-square to the df ratio, the RMSEA interval, and CFI indicate an adequate fit, and the AIC indicates that Model B fits better than Model A. This sequence of tests indicates that coherence within and across positive and negative affect dimensions is not equal. These model fits, in conjunction with the observation that correlations among the negative affect dimensions cluster around zero (Table 3) collectively support the notion that coherence within negative affect is weaker than coherence within positive affect.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

between the CBQ and post-visit ratings (Table 7) was similar to the pattern of Lab-TAB by CBQ correlations. The only small to moderate post-visit rating correlations were also with CBQ Fear, Shyness, Inhibitory Control, and Active Engagement. Three of these correlations were greater in magnitude than the Lab-TAB by CBQ associations. In general, associations between both objective ratings (i.e., Lab-TAB temperament composites and post-visit temperament dimensions) and the CBQ scales were much weaker than those between Lab-TAB and the post-visit observer ratings.

Regression Analyses: Lab-TAB Composites and CBQ Temperament Dimensions

The moderate or even non-existent associations between primary level Lab-TAB composites and "corresponding" CBQ maternal report scales are open to various interpretations, including the observation we have already offered that the content of the questionnaire and laboratory measures does not align precisely--or not at all in some cases--between measures that on the surface appear to correspond. However, the broader question of whether maternal report can predict children's actual behavior under standardized conditions can be addressed by examining the power of the full set of eight CBQ scales that we used for predicting each primary level Lab-TAB composite. In conducting this analysis, we initially considered using only plausible CBQ scales for each regression equation. For instance, it might be plausible that Lab-TAB Inhibitory Control could be predicted by CBQ Inhibitory Control, CBQ Attentional Focusing, CBQ Approach (negatively), and perhaps CBQ Fear and Shyness (positively). However, such considerations are somewhat subjective. Instead, we decided on a more exploratory approach wherein all CBQ scales were used in each regression equation, as Table 7 shows. We expected several of the CBQ scales to have no predictive power for each Lab-TAB composite.

As the second line of Table 8 shows, the overall multiple regression was significant at $p<.05$ for seven of the nine Lab-TAB composites, and the significance level for predicting Persistence was $p=.05$. In contrast, Lab-TAB Sadness was clearly not predicted by the set of CBQ scales. For each of the CBQ subscale predictors, different patterns emerged. CBQ Shyness positively predicted Lab-TAB Shyness and Inhibitory Control, and negatively predicted Lab-TAB Anger, Positive Expression and Approach. The CBQ Activity Level scale predicted Active Engagement and Anger in the standardized assessment with positive partial regression coefficients and Shyness with a negative coefficient. The CBQ Approach, Attentional Focusing, Fear, and Inhibitory Control scales also evinced significant predictive power of Lab-TAB temperament composites. The CBQ Approach scale did not predict the corresponding Lab-TAB composite, and the Sadness and Anger scales did not predict any Lab-TAB composites. The patterns of significant findings indicate that mother reports of child temperament can indeed predict child behavior in a standardized situation, but that these predictions do not necessarily correspond to single objectively-assessed dimensions of affect and/or behavior. One can compare the R^2 values in the Table 8 regression analyses with the r^2 values in the last column of Table 5 to estimate the incremental predictive power of multiple CBQ scales over a single "corresponding" CBQ scale in predicting a Lab-TAB dimension. In only some of these cases does the adjusted R^2 suggest incremental prediction. Interestingly, mother-rated Shyness and Activity Level show the most significant associations with children's Lab-TAB ratings.

Discussion

In this paper, we examined psychometric properties of a home-based Preschool Lab-TAB assessment with a community sample of preschool-age children; we used maternal report questionnaires and observers' post-visit ratings to provide validity evidence. Because the bulk of what we know about child temperament assessment has been derived from questionnaire methodology, and observational assessments of temperament have rarely been

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

organized into comprehensive protocols, little consensus exists on the best practices for assessing specific temperament-related behaviors or for forming composites of these behaviors to reflect temperamental dispositions. Our research questions focused on whether temperament dimensions that can be recovered from the Lab-TAB are similar to temperament questionnaire scales and whether correlations from "corresponding" dimensions show convergent validity across data derived from Lab-TAB, parental questionnaire, and post-visit observation rating methods. Previous articles that offered "how-to" advice on scale derivation and that proposed general principles for data reduction or latent trait estimation (e.g., Clark & Watson, 1995, in this Journal) have not generally included assessment methods that elicit actual behavior as one of the data types considered.

The home-based administration of the Lab-TAB included 12 episodes designed to elicit targeted behavioral reactions; most of these episodes generated dozens of raw data points. From these raw data, we calculated higher-level temperament composites. Our method of transforming short videotaped periods of elicited behavior into numerical scores that can be used to study temperament is not meant to be prescriptive; rather, the method is meant to illustrate a strategy that can be modified by other investigators. This process reflects an implicit model of temperament that includes the concept that the behavioral level of temperament is organized around affective/motivational constructs, such as anger proneness. The implicit model does *not* include the concept that the behavioral level of temperament is organized around parameters of response, such as latency or duration of response. Thus, an early stage in the transformation of raw behavioral into temperament measures involves averaging across response parameters. Likewise, the modality of response (e.g., facial, gestural, vocal) does not define temperament, and thus we also averaged across modality in deriving affective/motivational composites within episodes (where the episode is the task or trial that is defined by standardized affective incentives). Once scores defined by a common affective core are derived for each episode, our implicit model of temperament calls for averaging (e.g., by taking a principal component) the common scores across episodes. This process was illustrated in Table 2. The reason for averaging across episodes is that theories of temperament (as well as theories of personality more generally) posit cross-situational consistency as a defining feature of traits. Indeed, Allport (1937) defined "trait," in part, as "…the capacity to render many stimuli functionally equivalent…"

The decision to aggregate across episodes does not mean that situations are unimportant as organizers of behavior. Some researchers will find that a specific Lab-TAB episode captures a context that is crucial for a given study, and thus would not pursue aggregation. The power of situations (in this case, Lab-TAB episodes) to organize behavior is also one of the key reasons that the most basic elements of behavior (e.g., intensity of an angry facial expression in the End of the Line episode or vigor of approach in the Popping Bubbles episode) across all the episodes cannot be submitted to an exploratory factor analysis to derive temperament trait measures. Basic elements of elicited behavior are *not* comparable to the initial questionnaire items in a large pool that might be used to derive a set of questionnaire-based trait dimensions. When answering a questionnaire, the respondent is not constrained in answering item #2 by his or her response to item #1. This independence does not apply to actual behavior; the child who has withdrawn across a room in one 10-second interval of the Dinky Toys episode cannot touch the toys during the same interval. Similarly, once a child sadly resigns from trying to open the Transparent Box with faulty keys, he or she is highly unlikely to shift to anger, frustration or persistence within the next few seconds.

Once formed, the Lab-TAB temperamental composites showed low to moderate intercorrelations, as anticipated by prior studies (e.g., Gagne & Goldsmith, 2010; Pfeifer et al., 2002). Covariances between dimensions within the positive affectivity domain and within the behavioral control-regulation domain were significant. In contrast, the negative

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

affectivity dimensions showed little association with one another, indicating that negative emotions as assessed by these procedures are more distinct from one another in early childhood than positive emotions. However, Anger was positively correlated with the positive affectivity dimensions and negatively correlated with Shyness and the regulatory dimensions. We interpret this set of relations as indicative of a common approach motivation, whereby angry children are more likely to be generally expressive and active, and less likely to inhibit their behavior. Similarly, Inhibitory Control was negatively correlated with Positive Affect and positively correlated with Shyness. These findings are generally consistent with both contemporary theoretical and empirical perspectives on temperament.

The hypothesized associations between the Lab-TAB composites and the corresponding post-visit observer rating variables were all moderate (correlations ranged from .21 to .76). This pattern of correlations provides convergent validity for the Preschool Lab-TAB assessment. Lab-TAB temperament dimensions were derived from microanalytic coding of specific situations, whereas the post-visit ratings were based on rater's global impressions of child behavior across the entire home visit, including transitions between situations. For example, in the Lab-TAB Persistence episodes ("Perpetual Motion" and "Transparent Box") temperament is assessed in a very specific manner--the child is presented with stimuli designed to evoke a persistent response. However, some children may not be sufficiently engaged with the eliciting stimuli or their emotional reactions may prevent a persistent strategy. The coding of persistent behavior in these episodes is also very specific and may be qualitatively different from observer impressions across all of the Lab-TAB episodes. Differences in the magnitude of post-visit rating by Lab-TAB correlations might be affected by these differences, and dimensions that reflect lower agreement, such as persistence, may be a more subtle quality to observers than, for instance, overt fearful reactions.

As expected, the correlations between Lab-TAB dimensions and the most comparable CBQ questionnaire scales were much lower than those between Lab-TAB and the post-visit observer ratings. Poor agreement between lab and parent ratings of temperament is consistent with the larger cross-informant agreement in the child development and child psychopathology literatures. Lower covariance could be due to limited content overlap between the measures, despite similarities in how the dimensions are named. As described earlier, the post-visit observer ratings were conducted after the Lab-TAB assessments, which were witnessed by the observers as a part of the visit. Therefore, the overlapping behavioral "content" being tapped by both the Lab-TAB and post-visit observer ratings probably contributed to their higher covariance as opposed to the CBQ scales, which were based on mother's impressions of child behavior in the home. In addition, CBQ ratings reflect child behavior across many contexts over time whereas Lab-TAB and post-visit ratings reflect behaviors present only during the Lab-TAB situations and the home visit.

We also used an exploratory approach wherein all CBQ scales were entered in a regression equation to predict each Lab-TAB temperament composite. Our goal was to investigate whether the "non-corresponding" CBQ scales held predictive power for each Lab-TAB composite and whether each CBQ scale showed independent predictive power. Of course, we expected that several of the CBQ scales would hold no predictive power. The overall multiple regression was significant for seven of the nine Lab-TAB composites, and the regression for Lab-TAB Persistence was borderline significant (Sadness was the Lab-TAB dimension clearly not predicted by the CBQ scales). These findings suggest that maternal reports of child temperament can indeed predict child behavior in a standardized assessment, but that single dimensions of mother report do not typically correspond to single dimensions of standardized assessments. Mother-assessed Shyness and Activity Level showed the strongest associations with children's Lab-TAB scores, a finding that mirrored our

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

correlational results. Perhaps mothers report Shyness and Activity Level more accurately than other temperamental dimensions. However, although the internal consistency of CBQ Shyness assessment was higher than most other CBQ dimensions (α = .92), the internal consistency of CBQ Activity Level was fairly typical (α = .73). Alternately, Shyness and Activity Level reactions may have more salience for mothers on a day-to-day basis, resulting in better prediction of Lab-TAB temperament. For instance, the r^2 for prediction of Lab-TAB Shyness by CBQ Shyness was as high (10%) as the adjusted R^2 for prediction of Lab-TAB Shyness by all CBQ scales (also 10%). Perhaps mothers observe more instances of salient activity level and shyness in their children than they do of fear, anger, sadness, etc., and this larger parental "database" of activity and shyness observations leads to questionnaire responses that correspond better with Lab-TAB behavior. In any case, the results suggest that not all domains of parental report possess equal predictive power.

Unlike the Lab-TAB coders, mothers and observers made 'summary' judgments and attended to more globally meaningful units of child behavior. At this level of analysis, the ratings extracted a macro level view of temperament as opposed to the Lab-TAB composites, which focused on emotional reactions at a very discrete, or micro level. This contrast between micro vs. macro levels of analysis likely contributes to instances of low covariance between the Lab-TAB and maternal ratings in this study. Although a few of the associations between Lab-TAB and observer post-visit ratings were somewhat modest (e.g., Sadness, Active Engagement, and Persistence), most were quite strong. Relations between the Lab-TAB and CBQ were weaker, with half of the temperament dimensions showing no covariance (Anger, Sadness, Approach, and Persistence). The same four dimensions were not significantly correlated between the CBQ and post-visit ratings. However, the correlations between CBQ and post-visit observer ratings were slightly higher than the CBQ and Lab-TAB, particularly for the more "salient" dimensions of Shyness and Activity. We suggest that the shared macro level of analysis most likely contributed to the slightly higher associations. Yet, if this level of analysis issue were solely driving the lack of covariance between the Lab-TAB and the CBQ, we would have expected weaker correlations between Lab-TAB and the post-visit observer ratings than we actually observed.

Although the issue was not addressed empirically in our study, we suggest that another element in interpreting our results is that mothers probably view their children's behavior much more pragmatically than do researchers who are trained to assess the subtleties of emotional reactions and changes. Mothers' concerns may be with the relative "success" of the child's behavior; that is, mothers may focus on the child's ability to negotiate a task or situation regardless of the affective quality of the child's engagement in the task. For example, if a child tolerates the presence of strangers with little intervention and these situations usually "work out" without disrupting others, mothers might be less inclined to rate the child as being fearful of strangers, regardless of fearful emotional expressions that the child might show during interactions with strangers. Analogous lab-based ratings (e.g., in the Lab-TAB Stranger Approach episode) focus on the specific emotional reactions of the child, and do not emphasize the resolution of the episode.

A clear advantage of objective assessment of temperament is its flexibility. Different data reduction schemes allow temperament to be characterized in terms of specific reactions to specific eliciting stimuli, as well as in terms of more functional units of behavior such as emotion-attention or emotion-action pairings that occur simultaneously. For example, in the Transparent Box episode, the child is assessed for a range of anger and sadness reactions (e.g., facial, bodily, vocal), persistence ("stops"), and attention (gaze aversion). The meaning of attending to this task and expressing anger simultaneously is qualitatively different by simply being rated "high" on questionnaire scales tapping attention and anger; i.e., the Lab-TAB reaction is a functional, contextualized attention-action pairing. The Lab-TAB allows

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

us to encapsulate situational expressions of emotion and attention/action across a multitude of episodes with various target and secondary behaviors assessed. Depending on the coding and data reduction scheme, such episodes and coding systems can lend greater richness and flexibility to temperament assessment than standardized questionnaire or global observer ratings. This molecular view can contribute to the work of researchers interested in the specificity of emotional expression in the context of individual tasks, multiple expressions of emotion within a situation (e.g., anger and sadness), and the stability of emotion and temperament across eliciting contexts. It is important to note that Lab-TAB episodes can also be coded from a global perspective (much like some parent and observer ratings), wherein coders view a range of videotaped episodes and rate their overall impression of the child's temperament. Thus, depending on the goals of a given study, objective assessment of temperament can yield different types of data.

A few researchers have approached Lab-TAB coding from both micro and macro levels of analysis in the same investigation. In one study examining laboratory based temperament assessment of preschoolers for associations with later childhood disorders, micro level and global coding schemes were developed for 12 Lab-TAB episodes (Hayden et al., 2005). The molecular coding followed the system of affective coding devised by Goldsmith and used in this paper. The global coding scheme involved the raters watching an entire episode and making a rating based on all behaviors relevant to each dimension of temperament. Findings indicated that laboratory-based positive emotionality predicted depression risk, and that both micro and macro level coding schemes were fairly congruent in predictive power. These global coding schemes were considered successful and used in subsequent analyses (Durbin, Hayden, Klein & Olino, 2007).

The use of the Lab-TAB procedures in these studies also demonstrates the stability of temperament assessment from one developmental period to another. Lab-TAB ratings of positive and negative emotionality at age 3 showed moderate to high levels of stability with matched laboratory assessments at 5–6 and 9 years (Durbin et al., 2007). These results were particularly robust in that stability was high even though the tasks were different at the three ages. Stability was higher in emotional aspects of temperament compared with other dimensions. One of the disadvantages of parent report is the potential overestimation of stability due to parents' wish to present a consistent picture of their child's behavior (Durbin, in press; Kagan, 1998). Laboratory temperament batteries that employ multiple layered tasks not only provide a more nuanced measure of individual differences in emotional reactivity, but also may offer more accurate estimates of change and stability (Durbin, in press).

Although in this paper we opted for simple data analytic approaches, the data structure produced by Lab-TAB affords structural equation modeling and multilevel modeling (MLM) approaches. Kiel and Buss (2006) applied MLM to Lab-TAB data from toddlers tested in our laboratory, and Durbin (in press) discussed the advantages of MLM approaches. In brief, MLM can be employed to model the temperamental reactivity for a particular emotion as a function of the "potency" of the task or situation. The variable of interest (i.e., a specific emotion or temperament dimension) is identified as the dependent variable and the tasks are assigned a potency value related to ability to elicit that trait. For example, the Stranger Approach task would have a high potency value for fear, but a low value for other aspects of temperament (Durbin, in press). These analyses provide both overall emotion/temperament scores (the intercept) and the slope of the emotion variable across increasingly potent lab episodes. MLM can improve the estimation of cross-contextual consistency and inconsistency and suggest better strategies for relating emotion variables to outcomes of interest based on the contexts that are most salient to elicitation.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

It is not our ultimate goal to replace parent ratings of temperament with laboratory assessments. Although significant methodological concerns attend the use of parental report, questionnaires are inexpensive and easy to administer. In addition, parental perspectives on their children's behavior are valuable and impossible to capture using laboratory measures. Rather than downplaying the importance of parental questionnaire assessment, our intention is to offer a psychometrically sound, objective measure of temperament as an option for a temperament assessment strategy. As previously mentioned, reliance on any one approach to temperament assessment carries with it risks. The use of multiple sources of information about participants' behavior in a quantitative analysis allows for firmer conclusions about the behavior being investigated (Saudino, 2005) and relations to important developmental outcomes.

We view the main limitations of the Lab-TAB approach as being the following: (1) the range of contexts and incentive events that are sampled is limited; (2) the procedures are relatively time-consuming and expensive, at least as compared with questionnaires (but not when compared with common assessment methods in cognitive, social, and neuroscience approaches); (3) exact replication of procedures across laboratories or different field settings is difficult, given factors ranging from room dimensions to subtle differences in examiner behavior; and (4) some theories of temperament posit typologies and Lab-TAB is designed to produce dimensional outcomes (although profiles or types can be derived statistically). Also, in this study, our sample was relatively homogeneous in racial/ethnic composition, which limits generalizability. A more general problem with lab- and home-based objective behavioral assessment of temperament is the question of how best to derive measures of temperament traits, and of course, the purpose of this paper is to suggest a fruitful strategy.

Acknowledgments

This research was supported in part by grants R01-MH044340 and P50-MH052354 from the National Institute of Mental Health and the Health Emotions Research Institute, University of Wisconsin-Madison. The project described was also supported by Award Number T32-MH018931 from the National Institute of Mental Health (Program Director: R. J. Davidson). We also acknowledge support from R37-MH050560, P30-HD03352 and P50-MH084051.

References

Aksan N, Goldsmith HH, Smider N, Essex M, Clark R, Hyde J, Klein M, Vandell D. Derivation and prediction of temperamental types among preschoolers. Developmental Psychology. 1999; 35:958–971. [PubMed: 10442865]

Allport, GW. Personality-A psychological interpretation. New York: Henry Holt & Company; 1937.

Bayley, N. Bayley Scales of Infant Development. New York, NY: The Psychological Corporation; 1969.

Bayley, N. Bayley Scales of Infant Development. 2. San Antonio, TX: The Psychological Corporation; 1993.

Buss, AH.; Plomin, R. A temperament theory of personality development. Oxford, U.K: Wiley; 1975.

Buss, AH.; Plomin, R. Temperament: Early developing personality traits. Hillsdale, NJ: Erlbaum; 1984.

Caspi A, Roberts BW, Shiner RL. Personality development: Stability and change. Annual Review of Psychology. 2005; 56:453–484.

Clark LA, Watson D. Constructing validity: Basic issues in scale development. Psychological Assessment. 1995; 7:309–319.

Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society Series. 1977; 39:1–38.

Durbin CE. Modeling temperamental risk for depression using developmentally sensitive laboratory paradigms. Child Development Perspectives. (in press).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Durbin CE, Hayden EP, Klein DN, Olino TM. Stability of laboratory assessed temperament traits from ages 3 to 7. Emotion. 2007; 7:388–399. [PubMed: 17516816]

Escalona S, Leitch M. Early phases of personality development: A non-normative study of infant behavior. Monographs of the Society for Research in Child Development. 1952; 17(vi):72.

Gagne JR, Goldsmith HH. A longitudinal analysis of anger and inhibitory control in twins from 12 to 36 months of age. Developmental Science. 2010 published online July ‘10.

Gagne, JR.; Vendlinski, MK.; Goldsmith, HH. The genetics of childhood temperament. In: Kim, Y-K., editor. Handbook of Behavioral Genetics. New York: Springer; 2009.

Gartstein MA, Marmion J. Fear and positive affectivity in infancy: Convergence/discrepancy between parent-report and laboratory-based indicators. Infant Behavior and Development. 2008; 31:227–238. [PubMed: 18082892]

Gartstein MA, Rothbart MK. Studying infant temperament via the revised infant behavior questionnaire. Infant Behavior & Development. 2003; 26:64–86.

Goldsmith, HH. Doctoral dissertation. University of Minnesota; 1978. Behavior genetic analyses of early personality (temperament): Developmental perspectives from the longitudinal study of twins during infancy and early childhood.

Goldsmith HH. Studying temperament via construction of the Toddler Behavior Assessment Questionnaire. Child Development. 1996; 67:218–235. [PubMed: 8605830]

Goldsmith HH, Buss AH, Plomin R, Rothbart MK, Thomas A, Chess S, Hinde RA, McCall RB. Roundtable: What is temperament? Four Approaches. Child Development. 1987; 58:505–529. [PubMed: 3829791]

Goldsmith HH, Campos JJ. The structure of temperamental fear and pleasure in infants: A psychometric perspective. Child Development. 1990; 61:1944–1964. [PubMed: 2083507]

Goldsmith HH, Gottesman II. Origins of variation in behavioral style: A longitudinal study of temperament in young twins. Child Development. 1981; 52:91–103. [PubMed: 7195330]

Goldsmith HH, Hewitt EC. Validity of parental report of temperament: Distinctions and needed research. Infant Behavior & Development. 2003; 26:108–111.

Goldsmith, HH.; Lemery, KS.; Essex, MJ. Temperament as a liability factor for childhood behavioral disorders: The concept of liability. In: DiLalla, LF., editor. Behavior genetics principles: Perspectives in development, personality, and psychopathology. Washington, DC: American Psychological Association; 2004. p. 19-39.

Goldsmith, HH.; Reilly, J.; Lemery, KS.; Longley, S.; Prescott, A. Technical Report. Department of Psychology, University of Wisconsin; Madison: 1993. Preliminary manual for the Preschool Laboratory Temperament Assessment Battery (version 1.0).

Goldsmith HH, Rieser-Danner LA, Briggs S. Evaluating convergent and discriminant validity of temperament questionnaires for preschoolers, toddlers, and infants. Developmental Psychology. 1991; 27:566–579.

Goldsmith, HH.; Rothbart, MK. Contemporary instruments for assessing early temperament by questionnaire and in the laboratory. In: Strelau, J.; Angleitner, A., editors. Explorations in Temperament. New York: Plenum Press; 1991. p. 249-272.

Graham JW. Missing data analysis: Making it work in the real world. Annual Review of Psychology. 2009; 60:549–576.

Hayden EP, Klein DN, Durbin CE. Parent reports and laboratory assessments of child temperament: A comparison of their associations with risk for depression and externalizing disorders. Journal of Psychopathology and Behavior Assessment. 2005; 27:89–100.

Hwang J, Rothbart MK. Behavior genetics studies of infant temperament: Findings vary across parent-report instruments. Infant Behavior & Development. 2003; 26:112–114.

Hyde JS, Klein MH, Essex MJ, Clark R. Maternity leave and women’s mental health. Psychology of Women Quarterly. 1995; 19:257–285.

Kagan, J. Galen’s Prophecy: Temperament in Human Nature. New York: Basic Books; 1994.

Kagan, J. Biology and the child. In: Eisenberg, N., editor. Handbook of child psychology: Vol 3. Social, emotional, and personality development. 5. New York: Wiley; 1998. p. 177-235.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Kiel EJ, Buss KA. Maternal accuracy in predicting toddlers' behaviors and associations with toddlers' fearful temperament. Child Development. 2006; 77:355–370. [PubMed: 16611177]

Kochanska G, Murray KT, Harlan ET. Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. Developmental Psychology. 2000; 36:220–232. [PubMed: 10749079]

Kochanska G, Murray K, Jacques TY, Koenig AL, Vandegeest KA. IC in young children and its role in emerging internalization. Child Development. 1996; 67:490–507. [PubMed: 8625724]

Kohnstamm, GA.; Halverson, CF.; Mervielde, I.; Havill, VL. Analyzing parental free descriptions of child personality. In: Kohnstamm, GA.; Halverson, CF.; Mervielde, I.; VL, editors. Parental Descriptions of Child Personality: Developmental Antecedents of the Big Five?. Mahwah, NJ: Erlbaum; 1998. p. 1-19.

Loevinger J. Objective tests as instruments of psychological theory. Psychological Reports. 1957; 3:635–694.

Mangelsdorf, SC.; Schoppe, SJ.; Buur, H. The meaning of parental reports: A contextual approach to the study of temperament and behavior problems in childhood. In: Molfese, VJ.; Molfese, DL., editors. Temperament and personality development across the life span. Mahwah, NJ: Lawrence Erlbaum Associates; 2000. p. 121-140.

Matheny AP Jr. Bayley's Infant Behavior Record: Behavioral components and twin analyses. Child Development. 1980; 51:1157–1167. [PubMed: 7193557]

Matheny AP Jr. A longitudinal twin study of stability of components from Bayley's Infant Behavior Record. Child Development. 1983; 54:356–360. [PubMed: 6683619]

Meehl PE. Factors and taxa, traits and types, differences of degree and differences in kind. Journal of Personality. 1992; 60:117–174.

Pfeifer M, Goldsmith HH, Davidson RJ, Rickman M. Continuity and change in inhibited and uninhibited children. Child Development. 2002; 73:1474–1485. [PubMed: 12361313]

Putnam SP, Gartstein MA, Rothbart MK. Measurement of fine-grained aspects of toddler temperament: The Early Childhood Behavior Questionnaire. Infant Behavior and Development. 2006; 29:386–401. [PubMed: 17138293]

Putnam SP, Rothbart MK. Development of short and very short forms of the Children's Behavioral Questionnaire. Journal of Personality Assessment. 2006; 87:102–112. [PubMed: 16856791]

Riese ML. Predicting infant temperament from neonatal reactivity for AGA/SGA twin pairs. Twin Research. 1998; 1:65–70. [PubMed: 10051347]

Robinson, JL.; McGrath, J.; Corley, RP. The conduct of the study: Sample and procedures. In: Emde, RN.; Hewitt, JK., editors. Infancy to Early Childhood: Genetic and Environmental Influences on Developmental Change. New York, NY: Oxford University Press; 2001. p. 23-41.

Rothbart MK. Longitudinal observation of infant temperament. Developmental Psychology. 1986; 22:356–365.

Rothbart MK. Temperament and the development of inhibited approach. Child Development. 1988; 59:1241–1250. [PubMed: 3168640]

Rothbart, MK. Temperament in childhood: A framework. In: Kohnstamm, GA.; Bates, JE.; Rothbart, MK., editors. Temperament in Childhood. Chicester, U.K: Wiley; 1989. p. 59-73.

Rothbart MK, Ahadi SA, Hershey KL. Temperament and social behavior in childhood. Merrill-Palmer Quarterly. 1994; 40:21–39.

Rothbart MK, Ahadi SA, Hershey KL, Fisher P. Investigations of temperament at 3–7 years: The Children's Behavior Questionnaire. Child Development. 2001; 72:1394–1408. [PubMed: 11699677]

Rothbart, MK.; Bates, JE. Temperament. In: Eisenberg, N., editor. Handbook of Child Psychology: Social, Emotional, & Personality Development. New York: Wiley; 1998. p. 105-176.

Saudino KJ. Parent ratings of infant temperament lessons from twin studies. Infant Behavior & Development. 2003a; 26:100–107.

Saudino KJ. The need to consider contrast effects in parent-rated temperament. Infant Behavior & Development. 2003b; 26:118–120.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Saudino, KJ. Multiple informants. In: Everitt, B.; Howell, D., editors. Encyclopedia of Statistics in Behavioral Science. Vol. 4. Chichester, UK: John Wiley & Sons; 2005. p. 1332-1333.

Saudino KJ. Do different measures tap the same genetic influences? A multi-method study of activity level in young twins. Developmental Science. 2009; 12:626–633. [PubMed: 19635088]

Saudino, KJ.; Cherny, SS. Parental ratings of temperament in twins. In: Emde, RN.; Hewitt, JK., editors. Infancy and early childhood: Genetic and environmental influences on developmental change. New York, NY: Oxford University Press; 2001. p. 73-88.

Saudino KJ, Eaton WO. Infant temperament and genetics: An objective twin study of motor activity level. Child Development. 1991; 62:1167–1174. [PubMed: 1756660]

Saudino KJ, Eaton WO. Continuity and change in objectively assessed temperament: A longitudinal twin study of activity level. British Journal of Developmental Psychology. 1995; 13:81–95.

Saudino KJ, Wertz AE, Gagne JR, Chawla S. Night and day: are siblings as different in temperament as parents say they are? Journal of Personality and Social Psychology. 2004; 87:698–706. [PubMed: 15535780]

Seifer R. Twin studies, biases of parents, and biases of researchers. Infant Behavior & Development. 2003; 26:115–117.

Seifer R, Sameroff AJ, Barrett LC, Krafchuck E. Infant temperament measured by multiple observations and mother report. Child Development. 1994; 65:1478–1490. [PubMed: 7982363]

Thomas, A.; Chess, S. Temperament and Development. New York: Brunner/Mazel; 1977.

Thomas, A.; Chess, S.; Birch, HG. Temperament and Behavior Disorders in Children. New York: New York University Press; 1968.

Figure 1.
Depiction of the Levels of Analysis in Deriving Primary Level Temperament Composites from Lab-TAB

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Lab-TAB Episodes, Descriptions, and Target Responses.

Lab-TAB Episodes	Description	Target Responses
Negative Affectivity Domain		
Box Empty	Failed expectations due to a wrapped box (“present”) being empty	Sadness, Anger, Negative Affect
End of the Line	Unreasonable prohibition (a novel toy is retrieved by the parent after a demonstration)	Anger, Sadness, Negative Affect
Stranger Approach	Social interaction with unfamiliar adult wearing hat and sunglasses	Shyness, Person Fear
Spider	Physical contact with unknown, hidden object (a “jumping spider” toy)	Object Fear, Startle
Transparent Box	Desirable object kept out of reach (locked inside a transparent box)	Anger, Sadness, Frustration, Persistence (in trying to open the box)
Positive Affectivity Domain		
Bookmark	Paper, stamps and markers are used to make a bookmark with experimenter	Seated Activity Level, Engagement, Contentment
Perpetual Motion	Seated activity with “space wheel” toy, child left alone with toy for 3 minutes	Engagement, Persistence
Popping Bubbles	Blowing (low intensity) and popping (high intensity) bubbles using a bubble gun	High & low intensity pleasure
Pop-up Snakes	Tester surprises child with pop-up snake in a can, child and tester then surprise parent	Pleasure, contentment, anticipatory positivity
Workbench	Child manipulates various objects on a toy workbench, fine motor activity	Seated activity level, engagement
Behavioral Control-Regulation Domain		
Dinky Toys	Child forced to make a toy choice among many alternatives, 2 trials	Inhibitory Control
Snack Delay	Child must wait for a signal before eating a snack (M&M’s or Goldfish crackers), 6 trials	Inhibitory Control

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 2

Lab-TAB Content of Temperament Composites.

Temperament dimensions	Constituents of the episode level components	Internal consistency (α)	Mean corrected Item-total *r*
Anger	Box Empty episode: Anger component of mean, peak, speed of facial anger, postural anger, speed of frustration (9 items)	.93	.73
	End of the Line episode: Anger component of mean, peak, speed of facial anger, postural anger, speed of frustration (9 items)	.87	.61
	Transparent Box episode: Anger component of mean, peak, speed of facial anger, postural anger, speed of frustration (9 items)	.84	.55
Sadness	Box Empty episode: Sadness component of mean, peak, speed of facial sadness, postural sadness (6 items)	.87	.68
	End of the Line episode: Sadness component of mean, peak, speed of facial sadness, postural sadness (6 items)	.80	.55
	Transparent Box episode: Sadness component of mean, peak, speed of facial sadness, postural sadness (6 items)	.85	.64
Fear	Spider episode: Approach-related fear component of initial touch and peak wariness of approach (2 items)	(*r* = .77)	n/a
	Spider episode: Post-approach fear expression component of mean facial fear, bodily fear, vocal distress and withdrawal (4 items)	.87	.71
Shyness	Stranger Approach episode: Shyness component of approach and shyness ratings across 2 raters (4 items)	.90	.78
Positive Expression	Bookmark episode: Smiling component of mean, peak and speed of smiling (3 items)	.73	.56
	Popping Bubbles episode, low intensity trials: Positive affect expression component of mean, peak and speed of smiling, % intervals laughter (6 items)	.78	.57
	Popping Bubbles episode, high intensity trials: Positive affect expression component of mean, peak and speed of smiling, % intervals laughter (4 items)	.84	.65
	Pop-up Snakes episode: Positive affect expression component of mean, peak and speed of smiling, % intervals laughter (14 items)	.88	.54
Approach	Box Empty episode: Anticipation component of mean, peak, speed of anticipatory behavior (3 items)	.94	.94
	Perpetual Motion episode: Approach component of mean and peak of active approach, mean frequency of touches, and speed of anticipatory behavior (4 items)	.86	.71
	Popping Bubbles episode, low intensity trials: Approach component of mean, peak, speed of vigor of approach (3 items)	.76	.61
	Popping Bubbles episode, high intensity trials: Approach component of mean, peak, speed of vigor of approach (2 items)	(*r* = .83)	n/a
	Pop-up Snakes episode: Approach component of mean, peak, speed of vigor of approach (6 items)	.84	.63
Active Engagement	Bookmark episode: Active engagement component of mean, peak of active approach (2 items)	(*r* = .62)	n/a
	Workbench episode: Activity level component of mean, peak of play (2 items)	(*r* = .62)	n/a
Persistence	Perpetual Motion episode: Persistence component of % time on task and latency to off-task behavior (2 items)	(*r* = .52)	n/a
	Transparent Box episode: Persistence component of % time on task and latency to off-task behavior (2 items)	(*r* = .83)	n/a
Inhibitory Control	Dinky Toys episode: Inhibitory control component of mean and speed of impulsivity across trials (4 items)	.50	.30

Temperament dimensions	Constituents of the episode level components	Internal consistency (α)	Mean corrected Item-total *r*
	Snack Delay episode: Inhibitory control component of global inhibitory control across trials (4 items)	.75	.43

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

Means, Standard Deviations, t-Tests: Lab-TAB, Post-visit, and CBQ Temperament Dimensions.

Temperament Dimension	Mean (Standard Deviation)		t
Lab-TAB	**Males**	**Females**	
Anger	.18 (1.04)	−.16 (.93)	3.47**
Sadness	−.03 (1.02)	.03 (.98)	−.60
Fear	−.05 (1.03)	.05 (.97)	−.99
Shyness	.02 (1.04)	−.02 (.97)	.36
Positive Expressiveness	.05 (.94)	−.05 (1.05)	.99
Approach	.32 (.89)	−.30 (1.0)	6.62**
Active Engagement	.11 (.97)	−.10 (1.01)	2.17*
Persistence	.11 (1.01)	−.10 (.98)	2.10*
Inhibitory Control	−.20 (1.03)	.18 (.93)	−3.93**
Post-visit Ratings	**Males**	**Females**	
Anger Proneness	.10 (1.0)	−.09 (1.0)	1.85
Sadness	−.12 (1.04)	.11 (.95)	−2.31*
Fear	−.01 (1.01)	.01 (.99)	−.29
Shyness	−.15 (.93)	.13 (1.04)	−2.82*
Positive Affect	.10 (1.04)	−.09 (.95)	1.96
Exuberance	.23 (.97)	−.21 (.99)	4.55**
Hyperactivity	.34 (1.05)	−.31 (.85)	6.88**
Persistence	−.05 (1.0)	.05 (1.0)	−1.04
Impulsivity	.20 (.96)	−.18 (1.0)	3.92**
CBQ	**Males**	**Females**	
Anger	.15 (1.0)	−.04 (.97)	1.91
Sadness	−.11 (.98)	.18 (.95)	−3.04**
Fear	.02 (1.02)	.03 (1.0)	−.09
Shyness	−.08 (1.0)	.09 (1.0)	−1.69
Approach	.04 (1.08)	.02 (.98)	.10
Activity Level	.24 (1.06)	−.17 (.90)	4.07**
Attentional Focusing	−.22 (1.02)	.11 (.98)	−3.26**
Inhibitory Control	−.22 (1.05)	.16 (.95)	−3.78**

Note.

* $p < .05$,

** $p < .01$ (2-tailed), N=408 (196 males and 212 females).

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 4

Correlations: Lab-TAB Temperament Dimension Composites.

	Sadness	Fear	Shyness	Positive. Express	Approach	Active Engage.	Persistence	Inhibitory Control
Anger	.08	.06	−.12*	.25**	.42**	.18**	−.27**	−.30**
Sadness		.10*	.05	−.02	−.05	−.08	−.12*	.00
Fear			−.02	.00	−.06	−.02	−.12*	−.09
Shyness				−.16**	−.18**	−.09	.00	.19**
Positive Expression					.56**	.20**	−.20**	−.30**
Approach						.28**	−.18**	−.44**
Active Engagement							.01	−.12*
Persistence								.24**

Note.

* $p < .05$,

** $p < .01$ (2-tailed), N=408.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 5

Correlations: Lab-TAB Temperament Dimensions and Post-visit Temperament Dimensions.

Lab-TAB Composites	Corresponding Post-visit Ratings	Correlation
Temperament Dimensions		
Anger	Anger Proneness	.33
Sadness	Sadness	.25
Fear	Fear	.76
Shyness	Shyness	.46
Positive Expressiveness	Positive Affect	.48
Approach	Exuberance	.59
Active Engagement	Hyperactivity	.21
Persistence	Persistence	.24
Inhibitory Control	Impulsivity	−.51

Note. All correlations significant at the $p < .01$ level (2-tailed), N=408.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 6

Correlations: Lab-TAB Temperament Dimensions and CBQ Temperament Questionnaire Scales.

Lab-TAB Composites	Most comparable CBQ questionnaire scales	Correlation	r^2
Temperament Dimensions			
Anger	Anger	.03	.00
Sadness	Sadness	.00	.00
Fear	Fear	.15**	.02
Shyness	Shyness	.32**	.10
Approach	Approach	.09	.01
Active Engagement	Activity Level	.20**	.04
Persistence	Attentional Focusing	.09	.01
Inhibitory Control	Inhibitory Control	.19**	.04

Note.

* $p < .05$,

** $p < .01$ (2-tailed), N=408.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 7

Correlations: Post-visit Temperament Dimensions and CBQ Temperament Questionnaire Scales.

Post-visit Ratings	Most comparable CBQ questionnaire scales	Correlation
Temperament Dimensions		
Anger Proneness	Anger	.08
Sadness	Sadness	.09
Fear	Fear	.15**
Shyness	Shyness	.49**
Exuberance	Approach	.10
Hyperactivity	Activity Level	.29**
Persistence	Attentional Focusing	.04
Impulsivity	Inhibitory Control	−.22**

Note.

* $p < .05$,

** $p < .01$ (2-tailed), N=408.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 8

Regression Analyses: Predicting Each Primary Level Lab-TAB Composite from the Set of Eight CBQ Scales.

	Laboratory Temperament Assessment Battery: Primary Level Composite Variables								
	Shyness	**Fear**	**Anger**	**Sadness**	**Positive Expression**	**Active Engagement**	**Approach**	**Persistence**	**Inhibitory Control**
F, full model	6.80	2.00	3.59	0.17	5.66	4.17	8.27	1.93	4.63
p, full model	<.001	.046	<.001	.995	<.001	<.001	<.001	.054	<.001
Multiple *R*	.35	.20	.26	.06	.32	.28	.38	.19	.29
Adjusted R^2	.10	.02	.05	−.02	.08	.06	.13	.02	.07
Predictors: CBQ scales									
B-Activity Level	**−.14** (.07)	−.05 (.07)	**.16** (.07)	.01 (.07)	.04 (.07)	**.24** (.07)	.03 (.06)	.05 (.07)	.03 (.07)
B-Shyness	**.29** (.05)	−.08 (.06)	**−.16** (.06)	.04 (.06)	**−.22** (.06)	−.03 (.05)	**−.28** (.05)	.09 (.06)	**.16** (.06)
B-Approach	.04 (.06)	.09 (.06)	−.03 (.06)	−.02 (.06)	−.00 (.06)	**−.14** (.06)	.02 (.06)	**−.12** (.06)	−.08 (.06)
B-Attentional Focusing	−.03 (.06)	−.04 (.06)	.01 (.06)	.02 (.06)	**−.13** (.06)	**−.16** (.06)	−.03 (.06)	.10 (.06)	.01 (.06)
B-Fear	−.06 (.05)	**.16** (.05)	−.01 (.05)	.00 (.06)	−.08 (.05)	−.07 (.05)	.01 (.05)	.01 (.05)	.05 (.05)
B-Inhibitory Control	−.01 (.07)	−.04 (.07)	−.03 (.07)	.02 (.07)	−.06 (.07)	.11 (.07)	**−.18** (.07)	.01 (.07)	**.21** (.07)
B-Sadness	.04 (.06)	.03 (.06)	−.04 (.06)	.00 (.06)	−.03 (.06)	.11 (.06)	−.06 (.06)	.09 (.06)	.07 (.06)
B-Anger	−.04 (.07)	−.06 (.07)	−.01 (.07)	.01 (.07)	−.01 (.07)	.03 (.07)	−.08 (.06)	.05 (.07)	.04 (.07)

Note. Each regression had (8, 375) *df*. B, the unstandardized partial regression coefficient was equivalent to the standardized beta because all predictors and outcomes were standardized, standard errors of the estimate are in parentheses. Significant partial regression coefficients ($p < .05$) are printed in bold font. *N*=408.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Psychol Assess. Author manuscript; available in PMC 2012 June 1.

Appendix

Correlations_ Lab-TAB Episode-level Temperament Components.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1. Box Empty Anger	1	.34	.31	.18	−.12	.08	.00	.06	−.10	.13	.08	.08	.12	.18	.20	.13	.24	.05	.24	.04	−.06	.14	.08	−.20
2. End of the Line Anger	.34	1	.26	.04	−.11	.08	.07	.10	−.10	.16	.18	.20	.03	.23	.34	.26	.22	.06	.02	.07	−.13	.21	.10	−.34
3. Transparent Box Anger	.31	.26	1	.04	.01	.17	−.06	−.01	−.07	.14	.10	.11	.15	.17	.28	.13	.23	.14	.06	.13	−.02	.32	.15	−.13
4. Box Empty Sadness	.18	.04	.04	1	.11	.26	.08	.08	.00	−.03	.03	.00	−.02	−.01	−.04	.05	.10	−.13	.01	−.04	.01	.06	−.04	−.04
5. End of the Line Sadness	−.12	−.11	.01	.11	1	.07	.03	−.02	.00	−.13	−.06	−.07	−.09	−.03	−.13	.10	−.06	.03	−.07	.04	.00	−.07	−.04	.04
6. Transparent Box Sadness	.08	.08	.17	.26	.07	1	.03	.03	.09	.08	.07	.14	−.02	.09	−.02	−.02	.00	−.05	−.11	−.06	−.03	.34	.01	−.07
7. Jumping Spider Approach Fear	.00	.07	−.06	.08	.03	.03	1	.69	.01	−.07	.03	.03	−.16	.00	−.03	−.01	.01	−.26	.00	−.09	.11	.05	.09	−.01
8. Jumping Spider Post-Approach Fear	.06	.10	−.01	.08	−.02	.03	.69	1	−.05	.09	.13	.13	−.13	−.01	.03	.13	.15	−.27	.07	.05	−.08	.10	.17	−.02
9. Stranger Approach Shyness	−.10	−.10	−.07	.00	.00	.09	.01	−.05	1	−.08	−.14	−.10	−.13	−.07	−.05	−.16	−.12	−.16	−.08	−.06	.09	.07	−.13	.16
10. Bookmark Positive Affect	.13	.16	.14	−.03	−.13	.08	−.07	.09	−.08	1	.34	.33	.17	.19	.21	.27	.19	.06	.18	.06	−.10	.14	.17	−.18
11. Popping Bubbles Low Positive Affect	.08	.18	.10	.03	−.06	.07	.03	.13	−.14	.34	1	.50	.27	.24	.22	.51	.29	.13	.02	.09	−.12	.11	.14	−.20
12. Popping Bubbles High	.08	.20	.11	.00	−.07	.14	.03	.13	−.10	.33	.50	1	.27	.24	.20	.41	.39	.14	.09	.11	−.17	.15	.20	−.23

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Positive Affect																								
13. Pop-up Snakes Positive Affect	.12	.03	.15	−.02	−.09	−.02	−.16	−.13	−.13	.17	.27	.27	1	.16	.11	.20	.19	.50	.12	.14	−.02	−.01	.12	.00
14. Box Empty Approach	.18	.23	.17	−.01	−.03	.09	.00	−.01	−.07	.19	.24	.24	.16	1	.24	.25	.20	.17	.11	.08	−.16	.14	.17	−.26
15. Perpetual Motion Approach	.20	.34	.28	−.04	−.14	−.02	−.03	.03	−.05	.21	.22	.20	.11	.24	1	.34	.29	.14	.13	.15	.18	.25	.26	−.29
16. Popping Bubbles Low Approach	.13	.26	.13	.05	−.09	−.02	−.01	.13	−.16	.27	.51	.41	.20	.25	.34	1	.40	.16	.08	.18	−.15	.14	.15	−.27
17. Popping Bubbles High Approach	.24	.22	.23	.10	−.06	.00	.01	.15	−.12	.19	.29	.39	.19	.20	.29	.40	1	.07	.19	.20	−.06	.08	.21	−.24
18. Pop-up Snakes Approach	.05	.06	.14	−.13	.03	−.05	−.26	−.27	−.16	.06	.13	.14	.50	.17	.14	.16	.07	1	−.01	.13	.03	−.01	.10	−.07
19. Bookmark Active Engagement	.24	.02	.06	.01	−.06	−.11	.00	.07	−.08	.18	.02	.09	.12	.11	.13	.08	.19	−.01	1	.14	.01	−.03	.12	.00
20. Workbench Active Engagement	.04	.07	.13	−.04	.04	−.06	−.09	.05	−.06	.06	.09	.11	.14	.08	−.05	.18	.20	.13	.14	1	.00	.02	.10	−.03
21. Perpetual Motion Persistence	−.06	−.13	−.02	.01	.00	−.03	−.11	−.08	.09	−.10	−.12	−.17	−.02	−.16	.18	−.15	−.06	.03	.01	.00	1	−.10	−.09	.12
22. Transparent Box Persistence	.14	.21	.32	.06	−.07	.34	.05	.10	.07	.14	.11	.15	−.01	.14	.25	.14	.08	−.01	−.03	.02	−.10	1	.15	−.19
23. Dinky Toys Inhibitory Control	.08	.10	.15	−.04	−.04	.01	.09	.17	−.13	.17	.14	.20	.12	.17	.26	.15	.21	.10	.12	.10	−.09	.15	1	−.13
24. Snack Delay	.20	−.34	−.13	−.04	.04	−.07	−.01	−.02	.16	−.18	−.20	−.23	.00	−.26	−.29	−.27	−.24	−.07	.00	−.03	.12	−.19	−.13	1

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Inhibitory Control																								

Note. Correlations with an absolute value of .099 or greater and .129 or greater are significant at a $p < .05$ level, and $p < .01$ (two-tailed), respectively (values in the table are rounded to two decimal places). N=408.